2020-2021
Master 1 Archives

Records in Contexts : évolution de la description archivistique vers le web sémantique.

De la (méta)donnée à la connaissance.

Cepparo Estelle

Sous la direction de Madame Bénédicte Grailles

Membres du jury
Bénédicte GRAILLES | Maîtresse de conférences en archivistique (Université d'Angers)
Magalie MOYSAN | Maîtresse de conférences en archivistique (Université d'Angers)

FORMATION ARCHIVES ANGERS

Soutenu le :
10 juin 2021

FACULTÉ DES LETTRES, LANGUES ET SCIENCES HUMAINES
UNIVERSITÉ D'ANGERS

2020-2021
Master 1 Archives

Records in Contexts : évolution de la description archivistique vers le web sémantique.

De la (méta)donnée à la connaissance.

Cepparo Estelle

Sous la direction de Madame Bénédicte Grailles

Membres du jury
Bénédicte GRAILLES | Maîtresse de conférences en archivistique (Université d'Angers)
Magalie MOYSAN | Maîtresse de conférences en archivistique (Université d'Angers)

FORMATION ARCHIVES ANGERS

FACULTÉ DES LETTRES, LANGUES ET SCIENCES HUMAINES
UNIVERSITÉ D'ANGERS

Soutenu le :
10 juin 2021

L'auteur du présent document vous autorise à le partager, reproduire, distribuer et communiquer selon les conditions suivantes :

- Vous devez le citer en l'attribuant de la manière indiquée par l'auteur (mais pas d'une manière qui suggérerait qu'il approuve votre utilisation de l'œuvre).
- Vous n'avez pas le droit d'utiliser ce document à des fins commerciales.
- Vous n'avez pas le droit de le modifier, de le transformer ou de l'adapter.

Consulter la licence creative commons complète en français : http://creativecommons.org/licences/by-nc-nd/2.0/fr/

Ces conditions d'utilisation (attribution, pas d'utilisation commerciale, pas de modification) sont symbolisées par les icônes positionnées en pied de page.

REMERCIEMENTS

L'aboutissement de ce travail de recherche a été possible grâce au concours de nombreuses personnes à qui je souhaite témoigner toute ma gratitude.

Mes remerciements vont tout d'abord à ma directrice de mémoire, Madame Bénédicte Grailles, maîtresse de conférences en archivistique à l'Université d'Angers, pour son encadrement de qualité et ses conseils avisés qui ont nourri ma réflexion au fil des mois.

J'adresse toute ma reconnaissance à Monsieur Daniel Pitti pour le temps accordé et les informations apportées. En tant que président de l'EGAD, le partage de votre expérience et de votre connaissance du sujet a été indispensable à mon mémoire.

J'ai une pensée de gratitude particulière envers Madame Florence Clavaud, membre exécutif de l'EGAD à la tête de l'équipe de développement de RiC-O et responsable de la mission référentiels aux Archives nationales, pour les réponses fournies et l'entretien accordé ; ce mémoire n'aurait pu se faire sans vous.

Mes remerciements vont aussi à Madame Claire Sibille de Grimoüard pour le partage de son expertise en termes de normalisation de la description archivistique ; vos paroles ont guidé ma réflexion sur les enjeux inhérents aux processus de normalisation.

J'adresse aussi mes remerciements à Monsieur Thomas Francart, expert en modélisation des connaissances et en gestion des ontologies, qui m'a beaucoup appris sur les défis rencontrés lors de l'élaboration de RiC-O Converter et de Sparnatural.

Je souhaite également remercier Monsieur Jean-Séverin Lair ainsi que Monsieur Edouard Vasseur pour le partage de leur point de vue au sujet de RiC, ainsi que pour leur retour d'expérience du tandem archivistes/informaticiens mis en place lors du développement du programme VITAM ; le seul à ce jour.

Je tiens à exprimer toute ma gratitude envers Monsieur David Chavalarias, expert en graphe de connaissance et en sciences cognitives, qui a pris le temps de discuter de répondre à mes nombreuses questions et dont les réponses ont grandement fait avancer mon analyse.

Enfin, que soit aussi remercié mon entourage pour son soutien indéfectible.

Table des sigles et abréviations utilisés

A2A : Access to Archives

AACR : Anglo-American Cataloging Rules (anglais américain) ou Anglo-American Cataloguing Rules (anglais britannique)

AAF : Association des archivistes français

AAQ : Association des archivistes du Québec

ABES : Agence bibliographique de l'enseignement supérieur

ACA : Association of Canadian Archivists

AGRkMS : Australian Government Recordkeeping Metadata Standard

ALEGORIA : Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge

APPM : Archives, Personal Papers, and Manuscripts

AtoM : Access to Memory

BCA : Bureau of Canadian Archivists

BFAP : Berkeley Finding Aid Project

BNB : British National Bibliography

BNBMARC : British National Bibliography Machine Readable Catalogue

BnF : Bibliothèque nationale de France

CAIE : Committee on Archival Information Exchange

CBPS : Committee on Best Practices and Standards

CDS : Committee on Descriptive Standards

CIDOC-CRM : International Committee for Documentation — Conceptual Reference Model

CITRA : Conférence internationale de la Table ronde des archives

CNEDA : Comisión de Normas Españolas de Descripción Archivística

CPA : Commission on Preservation and Access

CRIHN : Centre de recherche interuniversitaire en humanités numériques

CSS : Cascading Style Sheets

DDS : Ad Hoc Commission on Descriptive Standards

EAC : Encoded Archival Context

EAD : Encoded Archival Description

EADWG : EAD Working Group

EAG : Encoded Archival Guide

EGAD : Expert Group on Archival Description

FAIR : Findable, Accessible, Interoperable, Reusable

FNE : Fichier national d'entités

FRBR : Functional Requirements for Bibliographic Records

FRBRoo : Functional Requirements for Bibliographic Records object oriented

HTML : HyperText Markup Language

ICA : International Council on Archives

IFLA : International Federation of Library Associations

IGN : Institut national de l'information géographique et forestière
INSEE : Institut national de la statistique et des études économiques
IRI : Internationalized Resource Identifier
ISAAR/CPF : International Standard Archival Authority Record for Corporate Bodies, Persons and Families
ISAD/G : International Standard Archival Description – General
ISBD : General International Standard Bibliographic Description
ISBD(G) : General International Standard Bibliographic Description
ISBD(M) : International Standard Bibliographic Description for single volume and multi-volume Monographs
ISDF : International Standard for Describing Functions
ISDIAH : International Standard for Describing Institutions with Archival Holdings
ISO : International Organization for Standardization
JSCAACR : Joint Steering Committee for Revision of AACR
LC : Library of Congress
LOD : Linked open data
M&S : Modelling and Simulation
MAD : Manual of Archival Description
MARC : Machine Readable Catalog
MARC AMC : Machine Readable Catalog Archival and Manuscripts Control
NISTF : National Information Systems Task Force
OWL : Web Ontology Language
PCDS : Planning Committee on Descriptive Standards
PIAAF : Pilote d'Interopérabilité des Autorités Archivistiques Françaises
POG : Plan d’orientation général
PROV-O : PROV Ontology
PSBP : Section on Professional Standards and Best Practices
RAD : Rules for Archival Description
RDF : Resource Description Framework
RiC : Records in Contexts
RiC-CM : RiC Conceptual Model
RiC-O : RiC Ontology
SAA : Society of American Archivists
SGML : Standard Generalized Markup Language
SIAF : Service interministériel des Archives de France
SKOS : Simple Knowledge Organization System
SNC : Système Nerveux Central
SNP : Système Nerveux Périphérique
SSHRC : Social Sciences and Humanities Research Council
UKMARC : United Kingdom Machine Readable Catalogue

URI : Uniform Resource Identifier

USMARC : United States Machine Readable Catalog

VITAM : Valeurs immatérielles transmises aux archives pour mémoire

W3C : World Wide Web Consortium

WGSAD : Working Group on Standards for Archival Description

XML : Extensible Markup Language

XSLT : eXtensible Stylesheet Language Transformations

Sommaire

TABLE DES SIGLES ET ABREVIATIONS UTILISES 7

INTRODUCTION GÉNÉRALE 1

LA NORMALISATION DE LA DESCRIPTION ARCHIVISTIQUE DE 1961 A 2008 5

1. Des origines bibliothéconomiques 7
2. Le rayonnement du monde anglo-saxon pour la normalisation de la description archivistique 16
3. ICA, chronique de la description archivistique internationale 28

CONCLUSION DE L'ETAT DES CONNAISSANCES 39

BIBLIOGRAPHIE 42

ETAT DES SOURCES 58

RECORDS IN CONTEXTS : CAP SUR LE WEB 3.0 64

1. RiC : origines, conception, implémentation 65
2. De la théorie à la pratique, première expérimentation de RiC à l'échelle nationale 84
3. Représentation sous forme de graphe : un moyen d'optimiser la compréhension contextuelle ? 100

CONCLUSION DE L'ETUDE DE CAS 110

CONCLUSION GÉNÉRALE 113

ANNEXES 116

1. Entretiens 116
2. Questionnaires 201

TABLE DES MATIERES 230

INTRODUCTION GÉNÉRALE

Le chapitre consacré aux instruments de recherche de la quatrième édition de l'*Abrégé d'archivistique* — ouvrage fondamental de tout archiviste en devenir — indique dès ses premières lignes que, « La rédaction des instruments de recherche constitue sans aucun doute l'un des cœurs du métier de l'archiviste, car ces derniers permettent ensuite au public de connaître l'existence des documents et d'y avoir accès »[1]. Déjà en 1998, lors d'un article paru dans *La Gazette des archives*, Vincent Doom soulignait l'enjeu d'un tel exercice, « Pourquoi traiter aujourd'hui de la description archivistique ? Parce qu'elle se situe au cœur de nos préoccupations, parce qu'elle est la source et le fondement des instruments de recherche, eux-mêmes vecteurs de la connaissance de nos ressources et donc de l'ouverture de nos archives à la recherche »[2].

L'arrivée du World Wide Web de Tim Berners-Lee en 1989 marque un tournant pour les sciences de l'information, changement auquel l'archivistique n'échappe pas. En effet, la toile nous fait entrer dans « l'âge des fichiers »[3] ainsi que le présente Jean-Michel Salaün, professeur en sciences de l'information à l'Ecole normale supérieure de Lyon. Plusieurs générations de web se succèdent. Du web 1.0 — ou web traditionnel — qui ne permettait qu'une navigation simple, nous passons dans les années 2000 au web 2.0 — dit aussi web social — qui permet alors le partage de contenus de tous types : textes, images, vidéos, sons… A partir de 2010, le web 3.0 — plus connu sous le nom de web sémantique — marque un nouveau cap : celui de pouvoir lier les connaissances entre elles, on parle alors de *linked data*. Autant d'étapes qui invitent les professionnels de l'information — bibliothèques, musées, services d'archives — à revoir leurs méthodes de travail. De la documentarisation, c'est-à-dire pour l'archiviste le fait de classer, indexer, décrire un document, on passe à la redocumentarisation, soit à la création de métadonnées. Composé du préfixe grec *meta* qui indique l'autoréférence, une métadonnée peut donc se définir comme étant une donnée sur une autre donnée. Utile pour organiser la connaissance, elle apparaît comme un outil d'aide à la recherche pour l'utilisateur final. Il est alors bon de revenir sur les trois dimensions qui constituent tout document, et comment ces dernières se voient être transformées avec l'arrivée du web sémantique.

Un collectif de chercheurs issus de diverses disciplines — Roger T. Pédauque — a défini la tridimensionnalité du document selon les modalités suivantes[4] : les *modalités*

[1] Association des archivistes français, *Abrégé d'archivistique, Principes et pratiques du métier d'archiviste*, Paris, Association des archivistes français, 2020, p. 177.
[2] Vincent Doom, « Description et analyse archivistiques ou la nécessité d'une normalisation », *La Gazette des archives*, n°182-183,1998, p. 232.
[3] Jean-Michel Salaün, « La redocumentarisation, un défi pour les sciences de l'information », *Études de communication*, vol. 30, 2007, p. 2, [en ligne], disponible sur http://journals.openedition.org/edc/428 (consulté le 27 janvier 2021).
[4] *Ibid.* p. 4.

anthropologiques qui font référence à la lisibilité de la forme et du signe, les *modalités intellectuelles*, comprenez l'intelligibilité du texte ou du contenu, et enfin les *modalités sociales*, qui se réfèrent à la médiation du document. Jean-Michel Salaün note alors que, « Cette triade peut être résumée en français de façon mnémotechnique par trois participes passés : vu, lu et su »[5]. Une triade que l'arrivée du web sémantique vient quelque peu bouleverser. Au sujet du « vu » — des *modalités anthropologiques* — il écrit que « La stabilité des supports a été remplacée par celle des structures dans des formats de fichiers qui séparent le contenu de sa mise en forme ». Pour le « lu » — les *modalités intellectuelles* — il mentionne que, « Les thésaurus laissent la place […] aux ontologies ». Enfin, pour le « su » — les *modalités sociales* — il précise que « Le document numérique appartient maintenant à un espace régi par les lois des grands nombres : parce qu'il est mis en relation avec un nombre quasi-infini de ses semblables et parce qu'il est potentiellement visible par un nombre non-fini de lecteurs ». Aussi, parce que le web sémantique repose sur une information structurée et liée, la normalisation apparaît comme un passage obligé pour garantir la qualité des données partagées. Pour la description archivistique, les avantages sont multiples, ainsi que le rappelait Alice Motte en 2015[6] : la « production d'une description selon une structure identique et commune », « la professionnalisation des archivistes » et l' « accès du public à une présentation des données homogènes quel que soit le service où il effectue des recherches ».

La première partie de notre étude portera donc sur les prémisses du cadre normatif international pour la description archivistique. Nous reviendrons donc sur les normes qui ont inspiré les travaux du Conseil international des Archives (ICA), et cela implique d'étudier l'influence exercée par les bibliothèques. Ces normes étant toutes issues du monde anglo-saxon, il peut s'avérer intéressant de questionner les motivations individuelles sous-jacentes dans l'élaboration d'un cadre normatif : quels en sont les acteurs ? quels sont les évènements clés dans l'évolution de la normalisation ? comment ces normes sont-elles reçues ? Aussi, l'objet de notre première partie est de comprendre l'enjeu de la normalisation pour la description archivistique, ainsi que le contexte historique et technologique dans lequel elle s'est inscrite de 1961 à 2008.

Les TIC sont en constante évolution. Et parce qu'elles sont au cœur des systèmes de description archivistique, l'ICA a souvent dû procéder à la révision de ses normes, ainsi que le rappellent certains membres de l'EGAD, « To keep current with the changing understanding and the ongoing transformation of communication technologies that provide the foundation for

---

[5] Jean-Michel Salaün, *Vu, lu, su. Les architectes de l'information face au monopole du web*, 2012, p. 26, [en ligne], disponible sur https://halshs.archives-ouvertes.fr/halshs-01979526/document (consulté le 7 avril 2021).

[6] Alice Motte, « La normalisation de la description archivistique : enjeux et actualités », *La Gazette des archives*, n°238, 2015, p. 122.

CC BY NC ND

archival description systems, ICA has periodically reviewed the standards »[7]. La seconde partie de notre étude porte donc sur *Records in Contexts*, la nouvelle norme de description de l'ICA pour la description archivistique. Cette dernière dote la discipline d'un modèle conceptuel et l'inscrit dans le web sémantique. C'est la volonté de mieux articuler les quatre normes internationales qui, en 2012, motive l'ICA à constituer un Groupe d'experts sur la description archivistique — l'EGAD — dévolu à cette tâche. La méthodologie adoptée pour l'élaboration de ce nouveau cadre normatif suscite l'intérêt : comment parvenir à un niveau d'abstraction suffisamment élevé pour convenir autant aux Archives nationales du Danemark qu'aux archives municipales de Tours ? comment concilier le plus de traditions archivistiques possibles ? comment s'opère le choix des diverses entités ? comment s'organise la construction de l'implémentation technique ?

De la volonté d'homogénéiser la description archivistique pour le professionnel, s'ensuit la réception de la norme par l'utilisateur final. La finalité de *Records in Contexts* sera de présenter les données archivistiques sous la forme de graphe ; comment ce type de visualisation peut-il influer sur la compréhension de la description contextuelle ? quels sont les effets exercés sur nos processus cognitifs ? comment ces derniers traitent-ils l'information perçue sous forme graphique ? En d'autres termes, quel est l'intérêt du graphe pour la compréhension et l'exploitation des ressources archivistiques ?

[7] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *The American Archivist*, vol. 76 (2), 2013, p. 567.

LA NORMALISATION DE LA DESCRIPTION ARCHIVISTIQUE DE 1961 A 2008

« La norme a pour objectif d'harmoniser des pratiques ou d'aligner des techniques. Instrument d'influence porté par des réseaux, elle a envahi à divers degrés l'univers de l'archiviste, son outillage intellectuel et technique »

Bénédicte Grailles[8]

Normaliser n'est pas un acte innocent. Ainsi que l'illustre cette citation de Bénédicte Grailles — maîtresse de conférences en archivistique à l'université d'Angers — la norme revêt un enjeu de pouvoir. En effet, cette dernière permet à tout acteur à la tête d'une initiative de normalisation d'imposer sa vision sur un champ d'action donné et ainsi servir ses intérêts propres ; une idée qui transparaît dans la citation suivante, « [S]tandards are always relative to the infrastructure within/upon/sometimes against which they are implemented »[9]. On comprend alors que tout processus normatif se doit d'être remis en contexte pour en comprendre pleinement la portée.

La description bibliographique fait figure de mère pour la description archivistique. Aussi, pour comprendre le génotype des normes de description archivistique actuelles, il nous paraît nécessaire de revenir tout d'abord sur le cadre normatif de la description bibliographique. Les principes de Paris adoptés en 1961 lors de la conférence sur les principes de catalogage, en tant que fruit d'un premier consensus international, constituent notre point de départ. Quelles en sont les spécificités héritées par la description archivistique ? Comment, à partir d'un contexte bibliothéconomique, ces normes se sont-elles transposées aux problématiques archivistiques ? Comment ont-elles été reçues et mises en pratique par les professionnels ? Quels effets ont-elles eu sur l'accès aux instruments de recherche par l'utilisateur final ?

Après investigation de l'influence de la description bibliographique sur la description archivistique, il conviendra d'analyser celle que revêt les pays de culture anglo-saxonne — le Royaume-Uni, le Canada et plus particulièrement les Etats-Unis — sur le paysage normatif de la discipline à l'échelle internationale. Il s'agira de mieux appréhender les traditions anglo-saxonnes intimement liées aux travaux normatifs de l'ICA, ainsi que l'hégémonie américaine sur les

[8] Bénédicte Grailles, « Normalisations, standardisations, modélisations : acculturations professionnelles ou uniformisations identitaires ? », *La Gazette des archives*, n°259, 2020, p. 51.
[9] Susan Leigh Star, Martha Lampland, « Reckoning with standards », *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Cornell University Press, Ithaca, 2009, p.7.

normes d'implémentation technique. Cette partie sera également l'opportunité de revenir sur les changements induits par ces nouvelles mesures pour la profession et sur l'évolution de l'accès aux ressources archivistiques.

1. Des origines bibliothéconomiques

Retracer l'histoire de la normalisation internationale en matière de description archivistique ne peut se faire sans revenir sur celle de la description bibliographique. L'adoption des principes de Paris en 1961, qui figurent comme « the broad draft of an outline code »[10], sont un passage obligé pour comprendre les nombreuses évolutions qui en ont découlé et dont résulte le cadre normatif actuel, « [The Paris principles] has become the basis for subsequent work on the revision of cataloguing codes in many countries ». La portée de ces principes sera mise en lumière à travers l'étude des diverses normes qui ponctuent la pratique de la description bibliographique.

1.1. Les règles de catalogage anglo-américaines de 1967

Première norme publiée depuis les principes de Paris, les *Anglo-American Cataloging Rules*[11] (AACR) parues en 1967 en empruntent la majeure partie des caractéristiques, « Except for one or two instances the new code follows the Paris Principles closely »[12]. C'est particulièrement vrai en ce qui concerne la première partie de la norme — *Entry and Heading* — relative aux choix et à la forme des vedettes et des notices.

Pour les notices des œuvres écrites par plusieurs auteurs par exemple, AACR suit à la lettre les points 10.1 et 10.2 des principes de Paris, qui indiquent respectivement : « if one author is represented in the book as the principal author […] the main entry for the work should be made under the name of the principal author », « if no author is represented as the principal author, the main entry should be made under the author named first on the title-page, if the number of authors is two or three », « if no author is represented as the principal author, the main entry should be made under […] the title of the work if the number of authors is more than three »[13]. En revanche, AACR s'éloigne du point 10.3 relatif aux œuvres dérivées — des œuvres composées de parties de plusieurs autres œuvres — qui mentionne que la notice principale se doit d'être faite sous le titre de l'œuvre dérivée si elle en possède un, sinon sous le nom de son auteur. L'ordre se trouve alors inversé, « Collections […] are covered by rule 5, which states that they are to be entered under the compiler […] otherwise under the title »[14].

Cependant, bien qu'étant le fruit d'un travail collectif entre les Etats-Unis, le Canada et le Royaume Uni, des écarts de pratiques ont conduit à la production d'une édition propre à

---

[10] Leonard Jolley, « International Conference on Cataloguing Principles II. Thoughts after Paris », *Journal of Documentation*, vol. 19, n°2, 1963, p. 62.
[11] Le nom de l'édition américaine de la norme a été retenu par souci de cohérence ; en effet, la version américaine est parue en janvier 1967, tandis que la version britannique a été publiée dans le courant du mois de novembre de la même année.
[12] James A. Tait, « The Anglo-American Cataloguing Rules, 1967 », *Library Review*, vol. 21, n°2, 1967, p. 69.
[13] International Conference on Cataloguing Principles, *Statement of Principles*, Paris, 1961.
[14] James A. Tait, « The Anglo-American Cataloguing Rules, 1967 », *art. cit.*, p. 70-71.

CC BY NC ND

chaque tradition archivistique, « key differences in practice […] resulted in the production of separate British and North American editions of the document »[15] ; la version américaine sous le nom de Anglo-American *Cataloging* Rules et la version anglaise sous celui de Anglo-American *Cataloguing* Rules. Parmi les désaccords relevés peuvent être citées les vedettes des notices principales de collectivités. En effet, l'AACR souhaitait initialement réaliser ces notices sous le nom des collectivités ; un problème transparaît alors, « The entry of all corporate bodies directly under their names would lead to accumulations of entries under such common words as 'University', 'Institute' and so on »[16]. Pour y remédier, l'édition américaine fait le choix de nouvelles mesures, « […] rules 98 and 99, which prescribe entry under place for certain specified types of institution whose names consist of a common word or phrase followed by the name of a place below national level »[17]. Cependant, la solution ne convient pas aux Britanniques qui préfèrent adopter le nom conventionnel de la collectivité en guise de vedette, « The British committee, to their credit, appear to have refused to accept these exceptions. They have been much more adroit, and have used an extension to rule 63, which favours entry under conventional names »[18]. Ces différences de pratiques sont telles que « Each text contained an appendix listing rules for entry and heading that differed in the other version »[19]. Le jeu des sept différences se poursuit au niveau de la seconde partie de la norme — intitulée *Description* — où, de nouveau, traditions américaines et britanniques ne parviennent pas à tomber d'accord. Cette partie apparaît comme le cœur de leurs divergences, « The largest area of divergence is undoubtedly in Part II, Description, where the rules […] are in many respects different from the American equivalents »[20]. Un exemple parmi d'autres concerne les dates à indiquer où, « the Americans would still prefer the imprint date of the copy being catalogued […] while the British would appear to favour the edition date followed by the imprint date, if different »[21].

La volonté de procéder à la publication de deux exemplaires de la même norme est intimement liée aux acteurs à la tête du processus de normalisation, « the dominant roles appear to be those of the American Library Association and the Library of Congress, with the (British) Library Association toiling along in their wake »[22]. Aussi, il est aisé de comprendre que les choix opérés l'ont été en faveur d'une tradition archivistique dominante ; celle des Etats-Unis, portée par la Bibliothèque des Congrès. L'influence de cette dernière est telle que, « it had been decided that no rule should be included in the North-American text of AACRI without the approval of the

---

[15] Brian Dobreski, « History of the Anglo-American tradition », *Anglo-American library cataloging*, 2020, [en ligne], disponible sur https://www.isko.org/cyclo/cataloging#3.3 (consulté le 15 avril 2021).
[16] James A. Tait, « The Anglo-American Cataloguing Rules, 1967 », *art. cit.*, p. 72.
[17] *Ibid.*
[18] *Ibid.*
[19] Joint Steering Committee for the development of RDA, « A brief history of AACR », [en ligne], disponible sur http://rda-jsc.org/archivedsite/history.html (consulté le 20 avril 2021).
[20] P. R. Lewis, « The Anglo-American Cataloguing Rules (British text) : A note of welcome », *Journal of Documentation*, vol. 24, n°1, 1968, p. 57.
[21] James A. Tait, « The Anglo-American Cataloguing Rules, 1967 », *art. cit.*, p. 73.
[22] P. R. Lewis, « The Anglo-American Cataloguing Rules (British text) : A note of welcome », *art. cit.* p. 55.

LC [Library of Congess] »[23]. Cependant, la réception des deux versions diffère quelque peu, et penche en faveur du texte britannique, « With the increasing interest in the automation of library catalogues, this section of the British text may well prove to have a particular value in its own right »[24], tandis que la version américaine se voit reprocher un manque de considération envers les pratiques émergentes relatives aux nouvelles technologies de l'information, « we have a new code based firmly on the needs and requirements of the card catalogue […] There is little evidence, apart from a mention in the preface, that the compilers of the new code have considered the implications of the computer for the library catalogue »[25].

1.2. L'arrivée de l'automatisation, quels impacts sur la normalisation de la description bibliographique ?

Parallèlement à l'AACR s'est développé MARC — acronyme de *MAchine Readable Catalog* —, un format d'échange de données informatisées relatives à la description bibliographique. L'impact de ce format n'est pas des moindres puisque « MARC has significantly contributed to the growth of library automation and the development of online catalogues »[26]. La Bibliothèque du Congrès fournit de nombreux catalogues "papier" aux bibliothèques du pays, ainsi que l'illustre la citation suivante « cards provided by the LC formed the 'stock in trade of a very big business' »[27]. Aussi, face à l'importante quantité à diffuser, l'institution envisage de se munir de nouveaux outils. L'automatisation paraît alors comme la solution la plus appropriée pour renforcer la rapidité et l'efficacité du processus, « Faced with an enormous amount of catalogue cards to distribute among the American libraries, the Library of Congress started, in 1964, the project MARC (Machine Readable Cataloguing) with the objective of using automation in the bibliographic processing services »[28]. L'idée de réaliser un projet pilote — MARC I — est proposée en 1965 à la suite de deux conférences réunissant divers professionnels du milieu qui ont désigné la Bibliothèque du Congrès « to act as the distributing agency for the new machine-readable cataloguing »[29], un choix qui vient asseoir l'influence de l'institution en la matière, mais légitime, « LC was a logical choice, since it had long influenced traditional catalog standardization through its printed card service and book catalogs »[30].

MARC I se développe de 1966 à 1968. Dès la première année, le projet suscite l'intérêt de la Bibliographie nationale britannique (BNB pour *British National Bibliography*) qui, après

---

[23] Philip Bryant, « Progress in documentation », *Journal of Documentation*, vol. 36, n°2, 1980, p. 144.
[24] P. R. Lewis, « The Anglo-American Cataloguing Rules (British text) : A note of welcome », *art. cit.* p. 57.
[25] James A. Tait, « The Anglo-American Cataloguing Rules, 1967 », art. cit., p. 74.
[26] Rahmatollah Fattahi, « A comparison between the online catalogue and the card catalogue », *Library Review*, vol. 44, n°2, 1995, p. 45.
[27] Philip Bryant, « Progress in documentation », *art. cit.*, p. 144.
[28] Fernanda Campos, Inês Lopes et Rosa Galvão, « MARC formats and their use: an overview », *Program*, vol. 29, n°4, 1995, p. 445.
[29] Anthony Long, « UK MARC and US/MARC: A brief history and comparison », *Journal of Documentation*, vol. 40, n°1, 1984, p. 2.
[30] Mortaza Kokabi, « The internationalization of MARC. Part I: the emergence and divergence of MARC », *OCLC Systems and Services*, vol. 12, n°1, 1996, p. 23.

CC BY NC ND

avoir déterminé si le format pouvait également être adapté à ses ressources, lance le BNB MARC Project en 1967[31]. Une nouvelle fois, deux formats sont publiés ; le format USMARC — aussi identifié comme LCMARC (LC pour *Library of Congress*) — et le format UKMARC — également nommé BNBMARC —. Une question se pose : pourquoi ne pas avoir élaboré qu'un seul format ? Développé en parallèle des règles de catalogage de 1967, MARC en hérite les mêmes problématiques, « Differing texts of Anglo-American Cataloguing Rules (AACR), also led to some divergence between the two formats »[32]. Parmi elles nous retrouvons celle des vedettes des notices principales, « Entries on the B.N.B MARC tape will be catalogued in accordance with the British text of the Anglo-American Cataloguing Rules, 1967 »[33], mais aussi l'ordre dans lequel doivent figurer les éléments de la notice, « The differences between the two AACR texts in rule 23 A, B, C, and especially D for court rules and their impact on the order of the elements in the two MARC formats »[34]. Des différences de terminologie sont également à mentionner, « "leader" in USMARC format versus "record label" in UKMARC format, or "bibliographical level" in US format versus "class of record" in British format »[35].

Au-delà des différences émanant de la substance-même des deux versions des règles de catalogage de 1967, leur degré d'insertion d'un format MARC à un autre diverge également. Côté états-uniens, « The USMARC format was never as deeply rooted in AACR »[36], tandis que son homologue britannique quant à lui, a été pensé à travers la norme de 1967, « Richard Coward, who headed the UK MARC research project, stated 'AACR standards in cataloguing are the cornerstone of the UK MARC record' »[37]. Il peut alors s'avérer intéressant de se questionner sur l'origine de ce lien étroit entre UKMARC et les pratiques bibliothéconomiques de la Bibliothèque Britannique. Serait-ce dû au degré d'implication dans les processus de normalisation ? C'est en tout cas ce que suggère Ross Bourne de la Bibliothèque Britannique, « while our Noth American friends are full participants in the maintenance of MARC, AACR and other standards, […] British libraries have been perhaps too independent and less convinced of the need to co-operate through standardization than have the North Americans »[38]. L'auteur de l'article va plus loin en évoquant la possible perte d'identité et un sentiment de non-légitimité ressenti par le Royaume-Uni, « I believe we suffer in this country (or countries, if you prefer) from no longer being sure about our identity in relation to the rest of the world »[39], un sentiment qui serait dû à la perte de son rayonnement mondial avec le déclin de l'Empire britannique, « the concept of the British Commonwealth was a very real one. Britain was at the heart of a major

---

[31] Mortaza Kokabi, « The internationalization of MARC. Part I: the emergence and divergence of MARC », *art. cit.*, p. 29.
[32] Mortaza Kokabi, « The internationalization of MARC. Part I: the emergence and divergence of MARC », *art. cit.*, p. 30.
[33] *Ibid.*, p. 29.
[34] *Ibid.*, p. 30.
[35] *Ibid.*
[36] *Ibid.*
[37] Philip Bryant, « Progress in documentation », *art. cit.*, p. 139.
[38] Ross Bourne, « Common MARC, or vivent les différences? », *Library Review*, vol. 45 n°2, p. 26.
[39] Ross Bourne, « Common MARC, or vivent les différences? », *art. cit.*, p. 27.

CC BY NC ND

international community »[40]. Un empire dont l'ensemble de la population s'élevait à 412 millions au début du siècle dernier[41] et qui a vu son déclin s'accélérer à la suite de la crise du Canal de Suez en 1956, « the Suez Crisis cruelly demonstrated Britain's fall from great power status »[42]. Aussi, parce que la domination de la Bibliothèque du Congrès en la matière a déjà été démontrée plus haut, l'USMARC s'est imposé comme un modèle de référence sur la scène internationale, « Since the issuance of the MARC II format by the Library of Congress, over twenty countries have developed and implemented MARC formats »[43].

1.3. *International Standard Bibliographic Description* (ISBD), un pas vers la réconciliation des pratiques américaines et britanniques ?

« The successful introduction of MARC, and progress on the development of format recognition programs for creating machine-readable records, were two factors which 'quickened the desire to further the work begun by the Paris Conference' »[44]. Aussi, la Fédération internationale des associations et institutions de bibliothèques (IFLA pour *International Federation of Library Associations*) organise dès 1969 une réunion internationale visant à rassembler les experts en catalogage. Cette réunion, qui a lieu à Copenhague, a pour objectif « de faire le point depuis les travaux de la Conférence internationale sur les principes de catalogage réunie à Paris à l'Unesco en 1961, de passer en revue les développements récents dans le domaine du catalogage, d'étudier les progrès rendus possibles par la normalisation internationale des règles et la mécanisation »[45]. Le fruit de cette réunion sera l'adoption du programme intitulé *International Standard Bibliographic Description* (ISBD) avec trois objectifs principaux, « The objective was to identify components in a bibliographic description, their preferred order, and the necessary punctuation »[46]. L'arrivée de l'automatisation rend possible l'idée d'une normalisation internationale en la matière, « dans la perspective des catalogues automatisés, les bibliothèques nationales les plus importantes envisagent maintenant de clore leurs catalogues traditionnels »[47]. Parmi elles, il y a bien évidemment la Bibliothèque du Congrès des Etats-Unis, qui en 1967, se montrait frileuse sur le sujet : « Because of the size of its catalogs and because changes on its printed cards would affect many others libraries, LC was particularly

---

[40] *Ibid.*
[41] Angus Maddison, *The World Economy: A Millennial Perspective*, Paris, OECD Publishing, 2001, p. 97
[42] Denis Judd, « From the Great Exhibition to the Festival of Britain: the decline and fall of the British empire », *History Extra*, 2021, [en ligne], disponible sur https://www.historyextra.com/period/modern/the-decline-and-fall-of-the-british-empire/ (consulté le 24 avril 2021).
[43] Katharine D. Morton, « The MARC formats: an overview », *The American Archivist*, vol. 49, n°1, 1986, p. 26.
[44] Philip Bryant, « Progress in documentation », *art. cit.*, p. 139.
[45] Roger Pierrot, « Etat présent de la normalisation du catalogage dans la perspective de l'automatisation », *Bulletin des bibliothèques de France*, n°3, 1970, p. 127.
[46] Joint Steering Committee for the development of RDA, « A brief history of AACR », *art. cit.*
[47] Roger Pierrot, « Etat présent de la normalisation du catalogage dans la perspective de l'automatisation », *art. cit.*, p. 129.

sensitive to the economic and practical effects of a new code »[48]. Par ailleurs, l'automatisation rend nécessaire la révision des normes de catalogage puisque « La conférence de 1961 n'avait pu, à sa date, envisager les problèmes posés par l'automatisation et s'était limitée à la vedette, sans considérer l'ensemble de la notice »[49]. Aussi, c'est lors de cette réunion qu'ont vu le jour deux groupes de travail, « l'un pour préparer l'édition commentée des principes de la Conférence de Paris en 1961, l'autre pour la normalisation de la notice bibliographique »[50].

Le second s'est réuni à plusieurs reprises, une première fois à Londres en octobre 1969, et une seconde fois à Paris en mai 1970. Lors de la première réunion, le groupe de travail « a décidé de laisser de côté […] les questions assez délicates concernant le titre et de concentrer son attention sur les autres parties de la notice »[51]. Michael Gorman, bibliothécaire britannique et secrétaire chargé de rédiger les travaux du groupe de travail, soumet le document à commentaires et reçoit « 21 réponses, émanant de 20 pays »[52]. Aussi la seconde réunion, « commença par examiner ces commentaires, ce qui l'amena à préciser les limites de son travail, d'une part, et à expliciter certaines notions, d'autre part »[53]. Michael Gorman rédige alors un nouveau document intitulé *Standard bibliographic description (for single volume and multi-volume monographs)*, une révision « distribuée pour être examinée à la réunion de la Commission de l'unification des règles de catalogage, lors du 36ème Conseil général de la F.I.A.B à Moscou, à la fin d'août 1970 »[54]. Des difficultés à parvenir à un consensus international planent toujours sur certains points. En effet, le groupe de travail doit encore établir une liste d'éléments obligatoires minimums afin d'en connaître le nombre à conserver, « Si, pour leurs besoins propres, des bibliothèques veulent abréger cette notice »[55]. De plus, la question de la ponctuation demeure, « C'est le vœu de tous d'aboutir à un codage qui permette une reconnaissance sans ambiguïté des divers éléments, par les bibliothécaires et par la machine, quelle que soit la langue de la notice. […] Mais déjà les ponctuations varient beaucoup d'une langue à l'autre. […] Tout au moins est-on d'accord sur l'usage des crochets pour indiquer une addition, des… pour indiquer une suppression »[56].

Malgré cela, la norme *International Standard Bibliographic Description* voit le jour en 1971, dix ans après la Conférence de Paris qui en avait esquissé les principes. Cette norme répond à trois exigences. La première concerne le fait que les notices bibliographiques produites dans une

---

[48] Wyllis E. Wright, « The *Anglo-American Cataloguing Rules*: a historical perspective », *Library Resources & Technical Services*, vol. 20, n°1, 1976, p. 44.
[49] Roger Pierrot, « Etat présent de la normalisation du catalogage dans la perspective de l'automatisation », *art. cit.*, p. 128.
[50] *Ibid.*, p. 127.
[51] Suzanne Honoré, « Réunion du groupe de travail sur la normalisation de la notice bibliographique. Londres, 27-27 octobre 1969 », *Bulletin des bibliothèques de France*, n° 3, 1970, p. 134.
[52] Suzanne Honoré, « Deuxième réunion du groupe de travail sur la normalisation de la description bibliographique. Paris 23-24 mai 1970 », *Bulletin des bibliothèques de France*, n° 9-10, 1970, p. 528.
[53] *Ibid.*
[54] *Ibid.*, p. 530.
[55] *Ibid.*
[56] Suzanne Honoré, « Deuxième réunion du groupe de travail sur la normalisation de la description bibliographique. Paris 23-24 mai 1970 », *art. cit.*, p. 530.

langue donnée doivent être aisément comprises par les personnes ne parlant pas la même langue ; la seconde qu'elles puissent être intégrées au sein de fichiers contenant des notices issues d'autres pays ; la troisième que les notices papier puissent être aisément converties en un format informatique[57]. La norme, à ce moment centrée sur les monographies — ISBD(M) —, est par la suite adaptée à plusieurs types de documents. Ainsi en 1977, Michael Gorman déclare alors que, « the ISBD system is widely used throughout the world »[58].

1.4. Les règles de catalogage anglo-américaines de 1978

« It had not been long before criticism had been voiced of the failure of AACRI to deal adequately with description […] The result was a number of amendments; the adoption by the ALA [American Library Association] and the LA [Library Association] of ISBD(M) as a revised Chapter 6 »[59]. Michael Gorman lui-même, « hoped that 'the incorporation of the ISBD into Chapter Six of AACR might lead to an agreed text of that chapter in the British and American version of the rules' »[60]. Aussi, si la normalisation internationale en la matière joue les médiateurs entre les deux versions de l'AACR1, elle n'est pas le seul facteur ayant conduit à la révision du texte en 1978. En effet, il est important de souligner le rôle joué par l'automatisation, et dans ce contexte, par la BNB qui, en sortant le format MARC de son environnement d'origine — celui de la Bibliothèque du Congrès des Etats-Unis —, l'inscrit dans un contexte plus large, « At BNB we prefer to talk about a machine-readable bibliographic record because a bibliographic record is required in many library operations other than cataloguing »[61]. Enfin, la décision de la Bibliothèque du Congrès des Etats-Unis de renoncer à sa pratique de « superposition » au regard des vedettes — c'est-à-dire au fait de conserver les points d'accès déjà en place et de recourir aux méthodes britanniques uniquement si les points d'accès sont inexistants — marque le début du processus de révision des AACR1, « International standardization and LC's decision to abandon its practice of "superimposition" (whereby obsolete forms of name were perpetuated in catalogs) were two of the driving forces behind the organization in 1974 of a tripartite meeting "consisting of one delegate each from the three 'Anglo-American' countries […] to draw up a new memorandum of agreement and to complete the planning of the project for a second edition of AACR" »[62].

De cette rencontre naît le *Joint Steering Committee for Revision of AACR* (JSCAACR), le groupe de travail pour la révision des règles de catalogage anglo-américaines de 1967. Les objectifs à atteindre au cours de ce processus de révision sont inscrits dans le mémorandum de 1966 — Memorandum of Agreement on Catalog Code Revision — étant donné que le groupe de

---

[57] Philip Bryant, « Progress in documentation », *art. cit.*, p. 140-141.
[58] *Ibid.*, p. 141.
[59] *Ibid.*, p. 145.
[60] *Ibid.*
[61] *Ibid.*, p. 138-139.
[62] Ralph W. Manning, « The Anglo-American Cataloguing Rules and their future », *Library Resources & Technical Services*, vol. 44, n°3, p. 130.

travail n'envisage pas une refonte des AACR, mais plutôt une continuité, « This second edition does not supersede their work [the 1967 edition], but continues it : for, in spite of the changes in presentation and content which it introduces, these are still the Anglo-American Cataloguing Rules, having the same principles and underlying objectives as the first edition »[63]. Parmi eux, outre la réconciliation des deux versions des AACR1 par l'intégration au sein du nouveau texte de tous les amendements réalisés auparavant et des nouvelles suggestions d'amendements de la part du JSCAACR, il s'agit également de prendre en considération, les idées de changement émanant des autres pays qui utilisent eux aussi les règles de catalogage anglo-américaines de 1967 ; la finalité étant d'ancrer davantage cette nouvelle édition dans le paysage normatif international[64]. Parmi les premiers travaux du JSCAACR paraissent les lignes directrices du groupe, au nombre de quatre. Ces dernières visent à inscrire la révision des AACR en conformité avec les Principes de Paris de 1961 et l'ISBD(M). Elles prévoient également d'adopter une attention particulière à l'égard des évolutions des processus d'automatisation et de définir le traitement des supports d'information autre que les livres[65].

Pour parvenir à une révision complète et optimale, le JSCAACR traite l'édition de 1967 par catégorie de règles à aborder[66]. Le résultat de cet effort collectif est visible en 1978 avec la parution des AACR2 en un exemplaire commun qui réconcilie les différentes pratiques. Une des premières différences remarquables est le changement de l'organisation interne de la norme, « AACRI dealt first with point of entry to the catalogue, and then went on to give rules for the description of monographs […]. This arrangement has been totally reversed in AACR2 »[67]. Une autre modification transparaît de manière plus subtile, puisque désormais les règles sont toutes édictées à l'impératif positif[68].

Bien que le groupe de travail ait atteint ses objectifs initiaux, quelques critiques vis-à-vis du texte produit ont tout de même été émises. En effet, il lui est tout d'abord reproché de manquer de régularité quant au choix de la terminologie utilisée, « In AACR 2 the terms "items", "works" and "publications" all occur in RI.0 General rules p. 13, but only "item" is defined in the glossary »[69]. L'auteur liste quatre caractéristiques auxquelles une norme de catalogage se doit de répondre : une définition claire de ce qu'est une entité bibliographique et des documents à indexer en format standard, un format détaillé auquel seront indexées les données jugées nécessaires à l'élaboration d'une description formelle de tout type de document ainsi que des

---

[63] Michael Gorman, Paul W. Winkler, *Anglo-American Cataloguing Rules, Second Edition, 1988 Revision*, Chicago, 1988, p. xix.
[64] Carol R. Kelm, « The historical development of the second edition of the Anglo-American Cataloging Rules », *Library Resources & Technical Services*, vol. 22, n°1, 1978, p. 23-24.
[65] Michael Gorman, Paul W. Winkler, *Anglo-American Cataloguing Rules, Second Edition, 1988 Revision*, *op. cit.*, p. xxi.
[66] Carol R. Kelm, « The historical development of the second edition of the Anglo-American Cataloging Rules », *art. cit.*, p. 27.
[67] Philip Bryant, « Progress in documentation », *art. cit.*, p. 145.
[68] Carol R. Kelm, « The historical development of the second edition of the Anglo-American Cataloging Rules », *art. cit.*, p. 27.
[69] J.A. Shinebourne, « A Critique of AACR », *Libri*, vol. 29, n°1, 1979, p. 233.

directives sur la façon dont les données relatives aux caractéristiques doivent être inscrites dans le format[70]. Il déclare ensuite que, « Unfortunately in AACR 2 only (3) and (4) could be said to be adequately accomplished but cannot be followed easily because of the failure of (1) and (2). »[71]. Il ajoute un peu plus loin que non seulement il n'y a pas de définition claire de ce qu'est une entité bibliographique, mais qu'à cela s'ajoute le fait que certains éléments de la description bibliographique sont eux-mêmes mal définis[72]. A titre d'exemple, les règles encadrant la mention d'édition — « edition statement » — et le complément du titre — « other title information » — se confondent, raison pour laquelle l'auteur précise, au sujet des AACR de 1967 que, « "British text" and "North American text" do not distinguish editions of a publication: they distinguish two texts »[73] ; il s'agit bien de deux textes différents, et non d'un texte original et d'une copie. Les conséquences potentielles de cette façon de procéder seraient visibles pour l'utilisateur final lors de ses recherches, « Misleading edition statements, could easily cause fruitless bibliographic searches for editions that do not exist »[74] des désagréments qui auraient pu être évités si dès le départ la norme avait défini les termes à employer pour pouvoir les utiliser à bon escient.[75]

De cette absence de définitions claires résulte une norme au volume conséquent dû aux nombreuses alternatives de catalogage présentées ; on est donc loin des quatre caractéristiques énoncées plus haut auxquelles une norme de catalogage se doit se répondre, ce que l'auteur déplore à maintes reprises, « what is more desirable is a code that articulates clear principles, and provides sound definitions […] That is an ideal, maybe unattainable, to which AACR 2 comes nowhere near »[76]. Une critique qui avait déjà été faite aux AACR de 1967, « It is perhaps disappointing that the new code is so large. […] one might have hoped that the new code would have been less bulky »[77].

Enfin, outre la substance même de la norme, c'est son implémentation qui est également critiquée car les changements qu'elle implique sont jugés couteux et demandent du temps pour être mis en place[78]. Raison pour laquelle l'implémentation de la norme a eu lieu trois ans après sa publication, « Anglo-American Cataloguing Rules, 2nd ed. (AACR2), was published in 1978, but implementation was first postponed to 1980, then to 1981 »[79].

---

[70] *Ibid.*
[71] *Ibid.*
[72] *Ibid.*
[73] *Ibid.*, p. 239-240.
[74] *Ibid.*, p. 241.
[75] *Ibid.*
[76] *Ibid.*, p. 244.
[77] James A. Tait, « The Anglo-American Cataloguing Rules, 1967 », *art. cit.*, p. 74.
[78] Philip Bryant, « Progress in documentation », *art. cit.*, p. 146.
[79] Arlene G. Taylor, « Implementing AACR and AACR2: a personal perspective and lessons learned », *Library Resources & Technical Services*, vol. 56, n°3, p. 124.

2. Le rayonnement du monde anglo-saxon pour la normalisation de la description archivistique

Dans les années 1970, la description archivistique est dépourvue d'un cadre normatif qui lui est propre. La raison ? La profession juge inutile la production de normes en la matière[80], les archivistes de l'époque étant convaincus que la pratique diverge d'une administration à une autre[81]. L'idiosyncrasie — c'est ainsi qu'est désignée l'absence d'homogénéité en termes de pratiques[82] — constitue cependant un écueil pour l'évolution de la description archivistique, « a more immediate and compelling reason for archivists to abandon some of their more individualistic ways is the low tolerance that automated systems have for idiosyncrasy and individualism »[83]. Une idiosyncrasie qui peut être due, par ailleurs, au manque d'une définition claire du rôle et des objectifs de la description archivistique à cette époque[84].

Cependant, ce n'est pas le seul facteur justifiant de l'absence de travaux en la matière. En effet, des difficultés d'ordre économique sont à relever, « the cost of consensus standards development is high for the often resource-poor archival profession », sans compter l'aspect chronophage de l'exercice, « The costs in travel, time, printing, and distribution are not trivial »[85]. Les bibliothèques ont quant à elles débuté ce processus de normalisation bien plus tôt que les archivistes, et en raison des nombreux points communs entre les deux professions, le cadre normatif bibliographique a fourni les bases nécessaires au développement du paysage normatif de la description archivistique[86], plus important encore, l'utilisation des normes de nos voisins, « helped us to see and understand that we are part of a larger information management community »[87]. Aussi, afin que la profession prenne la place qui lui revienne au sein de cette communauté[88], une série de normes propres à la discipline voit le jour.

2.1. Lutte contre l'idiosyncrasie : les efforts pionniers des Etats-Unis

En 1977 est créée la *National Information Systems Task Force* (NISTF), un comité de la Société des archivistes américains (SAA) qui étudie la question d'un système d'information

---

[80] Kathleen D. Roe, « Out of the primordial ooze: the genesis of U.S. archival descriptive standards », *The American Archivist*, vol. 74, 2011, p. 3.
[81] *Ibid.*, p. 4.
[82] Nancy A. Sahli, « Interpretation and application of the AMC format », *The American Archivist*, vol. 49, n°1, 1986, p. 10.
[83] Steven Hensen, « The Use of standards in the application of the AMC format », *The American Archivist*, vol. 49, n°1, 1986, p. 32.
[84] Lisa B. Weber, « Archival description standards: concepts, principles, and methodologies », *The American Archivist*, vol. 52, 1989, p. 511.
[85] *Ibid.*, p. 512.
[86] *Ibid.*, p. 506-507.
[87] Steven Hensen, « The First shall be first: APPM and its impact on American archival description », *Archivaria*, vol. 35, 1993, p. 70.
[88] *Ibid.*

national afin de rendre les archives plus facilement accessibles aux utilisateurs finaux[89]. Aussi parce que la problématique de l'amélioration de l'accès aux archives constitue le centre d'intérêt du comité, ce dernier travaille sur la production d'un cadre normatif pour la description archivistique et sur la simplification de l'adoption des techniques automatisées[90].

Parmi les premiers travaux d'envergure de la NISTF, le rapport d'Elaine D. Engst de 1980 intitulé *Standards elements for the description of archives and manuscripts: a report to the SAA Task Force on National information systems* tient une place de choix. Publié à la suite d'une étude qui visait à déterminer la diversité des pratiques en termes de description archivistique, les conclusions de l'enquête viennent mettre à bas cette idée d'idiosyncrasie puisque seul un petit nombre d'unités d'information est répertorié, preuve que les pratiques se ressemblent plus que les archivistes ne le pensent, « a relatively small number of descriptive data elements accounted for all of the descriptive practice of American archives »[91]. De ce résultat encourageant pour la normalisation des pratiques, le NISTF établit un dictionnaire des éléments de données — le *Date Elements Dictionary* —, « [the dictionary] proved definitively that archivists all indeed collected the same information, although they might call the elements by different names »[92]. Le dictionnaire du NISTF est perçu par ses membres comme étant un cadre normatif complet et permissif, « The standard was comprehensive because it included all types of repositories and records management activities, and it was permissive because it aimed to standardize current practice rather than establish prescriptive norms »[93]. Le dictionnaire est reconnu en 1982 par le Conseil de la SAA comme étant le cadre normatif de référence pour la description archivistique[94].

La prise de conscience qu'une normalisation de la discipline est possible — voire souhaitable — permet le passage d'un nouveau cap en 1983 lorsque la Bibliothèque du Congrès des Etats-Unis dévoile *Archives, Personal Papers, and Manuscripts* (APPM), première norme sur le contenu des données pour la description archivistique, qui puise ses origines dans les AACR2. L'influence est telle qu'elle en vient à être perçue comme étant une version révisée de cette dernière, « variously known as "AACR2½" or AACR2R »[95]. Cependant, APPM est novatrice puisque reposant sur trois principes qui répondent aux spécificités du traitement des archives. Le premier n'est autre que celui de la provenance, « APPM recognizes the primacy of provenance

---

[89] Richard H. Lytle, « An Analysis of the Work of the National Information Systems Task Force », *The American Archivist*, vol. 47, n°4, 1984, p. 360.
[90] Susan E. Davis, « Descriptive standards and the archival profession », *Cataloging & Classification Quaterly*, vol. 35, n°3-4, 2003, p. 298.
[91] *Ibid.*, p. 362.
[92] *Ibid.*
[93] *Ibid.*
[94] Working Group on Standards for Archival Description, « Archival Description Standards: Establishing a Process for Their Development and Implementation: Report of the Working Group on Standards for Archival Description. », *The American Archivist*, vol. 52, n°4, 1989, p. 469.
[95] Steven Hensen, « The First shall be first: APPM and its impact on American archival description », *art. cit.*, p. 65.

in archival description »[96]. Un principe visible à travers le choix de la vedette principale qui se fait sous le nom du producteur d'archives ; un point sur lequel APPM se conforme aux AACR2[97]. Le second principe fait référence à l'unité et reconnaît qu'un document ne fait sens que lorsqu'il est lu à travers le prisme de son contexte de production[98]. Aussi, APPM permet un catalogage respectueux des différents niveaux d'un fonds, et en ce sens, fait office de pionnier en la matière, « Previous attempts to provide cataloging rules for manuscripts and archives invariably made the mistake of assuming that, for cataloging purposes, the analog to the book was the individual manuscript »[99]. Enfin, APPM souligne la différence intrinsèque de la description archivistique en rappelant que, contrairement à la description bibliographique qui repose sur l'identification des informations figurant sur la page de titre[100], la description archivistique quant à elle, consiste en de la déduction d'information[101]. Un fait qui n'a rien de surprenant au vu des documents traités qui n'ont pas été initialement créés pour être publiés[102]. Aussi, bien que toujours liée aux AACR2, APPM constitue un pas en avant pour la profession, « The acceptance by archivists of library-based descriptive standards was a vital step — all the more so, because it helped us to see and understand that we are part of a larger information management community »[103].

La même année, l'aventure de la NISTF s'achève pour laisser la place au *Committee on Archival Information Exchange* (CAIE) dont les objectifs sont de maintenir le dictionnaire des éléments de données, mais également de promouvoir et soutenir des moyens d'échange et de diffusion d'informations relatives aux fonds d'archives. Aussi, en 1984 apparaît le format USMARC AMC, une norme relative à la structure des données directement empruntée aux bibliothèques. En effet, le CAIE a jugé plus économique de réutiliser un format existant en essayant de l'adapter au mieux aux besoins des archivistes, plutôt que d'en construire un de toutes pièces[104]. C'est la raison pour laquelle USMARC AMC emprunte la même ossature que son homologue bibliothéconomique, « the AMC format, like the other USMARC formats, has different parts, each of which contains a particular kind of data. These include the leader, the record directory, control fileds, and variable data fields »[105]. Les deux premières parties du format sont à destination de l'ordinateur et contiennent des informations générales concernant le document ainsi que les paramètres nécessaires à son traitement informatique. La partie intitulée « control fields » fournit les informations utiles à l'utilisateur lorsque ce dernier effectue des recherches. Enfin, « variable data fields » est la partie qui permet de structurer les données grâce aux

---

[96] Steven Hensen, « The First shall be first: APPM and its impact on American archival description », *art. cit.*, p. 67.
[97] *Ibid.*
[98] *Ibid.*
[99] Steven Hensen, « The Use of standards in the application of the AMC format », *art. cit.*, p. 35.
[100] *Ibid.*
[101] Steven Hensen, « The First shall be first: APPM and its impact on American archival description », *art. cit.*, p. 68.
[102] Steven Hensen, « The Use of standards in the application of the AMC format », *art. cit.*, p. 35.
[103] Steven Hensen, « The First shall be first: APPM and its impact on American archival description », *art. cit.*, p. 70.
[104] Nancy A. Sahli, « Interpretation and application of the AMC format », *art. cit.*, p. 12.
[105] *Ibid.*

différentes balises proposées[106]. Cependant, USMARC AMC se distingue de son homologue bibliographique car construit à l'aide du *Date Elements Dictionary* ; il s'adapte par conséquent aux besoins des archives. Aussi, il n'est pas surprenant que le format USMARC AMC fasse figure de fenêtre sur le monde pour les archivistes, « there may be much to be gained in terms of standardizing the way archivists communicate, both with each other and the world at large »[107] et ce en dépit de la maintenance partagée. En effet, il s'agit d'un procédé réalisé à la fois par la Société des archivistes américains, mais également par la Bibliothèque du Congrès, un manque à gagner en termes d'autonomie[108]. Une participation active des archivistes est pourtant essentielle si ces derniers souhaitent que le format demeure un outil adapté aux pratiques de la communauté professionnelle[109].

L'adoption rapide de USMARC AMC, l'absence d'une norme pour implémenter le format et d'une définition claire et partagée de ce qu'est la description archivistique, ainsi que le manque de coordination entre les divers projets de normalisation[110] ont conduit à la création du *Working Group on Standards for Archival Description* (WGSAD) en 1988, un groupe de travail indépendant de toute organisation dont l'objectif premier est de fournir un processus d'examen des normes de description archivistiques requises, « Devising the tools and procedures for evaluating, adopting, and maintaining description standards for the profession, therefore, became the primary mission of the Working Group on Standards for Archival Description »[111]. Au-delà de l'aspect pragmatique de sa mission, le WGSAD constitue une nouvelle étape de l'affirmation des archivistes au sein de la communauté des sciences de l'information ; à défaut de ne pouvoir compter ni sur une association professionnelle, ni sur une institution pour cultiver son paysage normatif, le devenir de la description archivistique repose alors sur l'initiative personnelle[112]. Un fait qui témoigne de la forte volonté des archivistes à améliorer leurs pratiques, « Do all professions rely so heavily on this kind of commitment to develop their professional practice standards? They do not. Really, they do not! »[113]. Le WGSAD est un phare pour la description archivistique.

En parallèle, l'implémentation de USMARC AMC a conduit la norme APPM à une révision, les éléments de cette dernière n'étant pas en accord avec ceux du format, « one of the purposes of this expanded and revised second edition of APPM is to reconcile its descriptive elements with

---

[106] Nancy A. Sahli, « Interpretation and application of the AMC format », *art. cit.*, p. 13.
[107] Steven Hensen, « The Use of standards in the application of the AMC format », *art. cit.*, p. 33.
[108] *Ibid.*
[109] Katharine D. Morton, « The MARC formats: an overview », *art. cit.*, p. 29.
[110] Working Group on Standards for Archival Description, « Archival Description Standards: Establishing a Process for Their Development and Implementation: Report of the Working Group on Standards for Archival Description. », *art. cit.*, p. 432-433.
[111] *Ibid.*
[112] Kathleen D. Roe, « Out of the primordial ooze: the genesis of U.S. archival descriptive standards », *art. cit.*, p. 10.
[113] *Ibid.*

corresponding fields in the USMARC format for Archival and Manuscripts Control »[114]. Aussi, lesdits éléments ont été renommés, tandis que d'autres ont été ajoutés[115]. En effet, certains sujets n'avaient pas été traités six ans auparavant, tels que le choix et la forme des vedettes et des points d'accès[116]. Par ailleurs, outre le fait de souhaiter accorder la norme au format USMARC AMC, il est fort probable que la volonté d'augmenter la contribution des archives au sein des systèmes d'information bibliographiques soit également un élément clé dans le processus de révision de APPM[117]. Le résultat du processus de révision paraît en 1989, et la seconde édition de l'APPM — APPM2 — est alors promue par la SAA ainsi que le suggère la neuvième recommandation du rapport effectué par le WGSAD, « Recommendation 9: The Society of American Archivists should endorse *Archives, Personal Papers, and Manuscripts* (APPM) as a standard for archival descriptive cataloging »[118].

2.2. Rupture avec les origines bibliographiques : la vision britannique

De l'autre côté de l'Atlantique, l'approche du processus de normalisation pour la description archivistique diverge considérablement. Tout commence au début des années 1980 avec la *Society of Archivists* — l'association professionnelle des archivistes — qui étudie la question d'un nouveau cadre normatif pour la discipline ; des réflexions qui prennent la forme d'un groupe de travail — le *Archival Description Project* — en 1984[119], « the Archival Descrption Project rejected the use of the *Anglo-American Cataloguing Rules* »[120]. Et en après deux ans de travail paraît un manuel — *A Manual of Archival Description* — qui, selon Michael Cook, l'un de ses auteurs, a davantage été publié dans l'objectif de faire naître la réflexion sur le sujet, « [MAD1] was published in 1986, largely for the purpose of generating discussion »[121]. Un manuel présenté aux archivistes britanniques à l'occasion de nombreux séminaires[122]. Aussi, les professionnels du Royaume-Uni font le choix de s'émanciper du cadre normatif des bibliothèques, puisque MAD1 « has rejected the bibliographic model »[123].

---

[114] Heather MacNeil, « HENSEN (comp.), Archives, Personal Papers, and Manuscripts: Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries », *Archivaria*, vol. 30, 1990, p. 153.
[115] *Ibid.*
[116] Working Group on Standards for Archival Description, « Archival Description Standards: Establishing a Process for Their Development and Implementation: Report of the Working Group on Standards for Archival Description. », *art. cit.*, p. 470.
[117] Heather MacNeil, « HENSEN (comp.), Archives, Personal Papers, and Manuscripts: Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries », *art. cit.*, p. 153.
[118] Working Group on Standards for Archival Description, « Archival Description Standards: Establishing a Process for Their Development and Implementation: Report of the Working Group on Standards for Archival Description. », *art. cit.*, p. 470.
[119] Michael Cook, « The British move toward standards of archival description: the "MAD" standard », *The American Archivist*, vol. 53, n°1, 1990, p. 132.
[120] Wendy M. Duff, Kent M. Haworth, « The Reclamation of archival description: the Canadian perspective », *Archivaria*, vol. 31, 1990, p. 27.
[121] Michael Cook, « The British move toward standards of archival description: the "MAD" standard », *art. cit.*, p. 132.
[122] *Ibid.*
[123] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *Archivaria*, vol. 35, 1993, p. 58.

Pour autant, il n'est pas question pour les Britanniques de prendre totalement leurs distances avec les pratiques états-uniennes. En effet, comme par le passé avec la normalisation de la description bibliographique, le Royaume-Uni s'appuie à nouveau sur les travaux réalisés par les Etats-Unis, « In terms of standards development, it could be argued that we rather shamelessly stole those of the others! »[124]. L'histoire se répète donc avec le format UKMARC que les Britanniques adaptent aux spécificités des archives, « the Society of Archivists […] commissioned a report on archival description in 1988, carried out by Michael Cook, which called for an archival version of UKMARC: UKMARC AMC »[125]. Aussi, le format UKMARC AMC voit le jour en 1989, et accompagne la publication de la seconde version du *Manual of Archival Description* (MAD2). Malheureusement, le format ne connut pas le succès de son homologue américain. Un échec dû au refus de la bibliothèque nationale du Royaume-Uni d'ajouter UKMARC AMC au format bibliographique UKMARC[126], et qui a influencé la position de des archives nationales britanniques, « Subsequently, and in part because of this opposition, the National Council on Archives rejected UK MARC as a possible format for the exchange of archival data in its March 1990 report on IT development »[127].

La norme de description MAD2 en revanche connaît plus de succès ; succès qui repose sur plusieurs facteurs. Tout d'abord, les règles qu'elle édictent sont plus précises et plus nombreuses[128] que celles proposées par MAD1. De plus, MAD2 — tout en continuant d'inscrire la description archivistique hors du cadre des bibliothèques — permet toutefois aux archivistes de comprendre la terminologie utilisée par les sciences de l'information, indispensable à cette époque alors que ces dernières reconnaissent des domaines communs avec les archives, « now that the information sciences are recognizing areas of interdependence it must be advantageous for archivists to understand a language which has hitherto seemed to us arcane »[129]. Aussi, MAD2 se présente comme une norme permettant d'affirmer des principes archivistiques au travers d'un langage et d'une méthodologie de travail propre aux sciences de l'information[130].

MAD2 apporte une approche nouvelle pour les archivistes de cette époque en attirant leur attention sur quatre points avant tout travail de description archivistique : tout d'abord la disposition des archives et l'utilisation d'un système de classification, puis les niveaux de disposition et de description, le degré de détail et de précision de la description, ainsi que la possibilité de modifier une description[131]. Le second fait référence à la hiérarchisation des

---

[124] William Stockting, « Commentary : a U.K. perspective », *The American Archivist*, vol. 74, 2011, p. 30.
[125] *Ibid.*, p. 28.
[126] Alan Hopkinson, *UKMARC AMC: Unpublished Draft Rev 4.0: UK MARC Format for Archives and Manuscripts Control (UK MARC AMC)*, Middlesex University Research Repository, 1995, p. 1, [en ligne], disponible sur https://eprints.mdx.ac.uk/5196/2/ukamc1.pdf (consulté le 5 mai 2021).
[127] *Ibid.*
[128] Michael Cook, « The British move toward standards of archival description: the "MAD" standard », *art. cit.*, p. 132.
[129] David Robinson, « Theory, pratice and a Manual of Archival Description », *Journal of the Society of Archivists*, vol. 12, n°1, 1991, p. 1.
[130] *Ibid.*, p. 2.
[131] Michael Cook, « The British move toward standards of archival description: the "MAD" standard », *art. cit.*, p. 133-135.

niveaux évoquées plus haut, qui se traduit au sein de MAD2 par l'attribution d'un numéro en fonction du niveau de classement traité. Le numéro 1 correspond au niveau de « management », il permet de désigner plusieurs fonds rassemblés en raison d'un trait commun tel que le caractère public ou privé des archives qu'ils contiennent. Le numéro 2 fait référence au fonds, le 3 à la série, le 4 au dossier et le 5 à la pièce[132]. La première ligne de chaque niveau de classement se compose de trois éléments : une cote, un titre et des dates extrêmes qui constituent ensemble un premier point d'accès pour aider l'utilisateur final à retrouver les documents.

Le troisième point relatif au degré de détail et de précision de la description mérite également d'être développé. En effet, il repose sur trois règles : la règle de la représentation, la règle de la recherche d'information et la règle contre les biais. La première indique que tout instrument de recherche étant réalisé pour répondre à un besoin spécifique, la forme de ce dernier se doit d'être choisie de manière à répondre au mieux au besoin exprimé. La seconde règle précise que tous les instruments de recherche se doivent de contenir les mots clés nécessaires à l'utilisateur final pour qu'il accède aux ressources archivistiques souhaitées. Enfin la troisième et dernière règle garantit l'authenticité de l'instrument de recherche qui se doit de refléter fidèlement le contenu et le contexte de production des documents d'archives[133]. Une fois ces trois points à l'esprit, le travail de description archivistique peut commencer, et l'archiviste peut alors s'appuyer sur un tableau contenant des éléments de données pour mener à bien la réalisation de son instrument de recherche. Cependant, son chemin vers la réalisation d'un instrument de recherche ne sera pas sans difficulté.

En effet, MAD2 s'est vue reprocher son manque de précision sur des sujets pourtant essentiels, « MAD seems to have two modes of expression, the general rule and the vague generalisation […] There is no balancing of pros and cons where decisions are not clearcut »[134]. L'auteur illustre ses propos en signalant que le chapitre intitulé *The function of a finding aid system* ne fait pas état de bon nombre des fonctions réalisées par un instrument de recherche, « [the chapter] does not describe several of the specific functions a finding-aid system performs in acting as a representation of the archive, such as owner's receipt, public relations tool or inventory against loss, misplacement or theft »[135]. MAD2 n'apporte pas d'élément pour déterminer la typologie d'instrument de recherche à adopter — synthétique ou analytique — puisqu'elle ne présente aucun des facteurs à prendre en compte pour effectuer ce choix, « The aim of security from theft or protection from unnecessary handling […] might be cited as justifications for detailed item-by-item description. Constraints of time, similarity of content […] might be used as arguments for less detailed description. […] These, together with resources of staff and time, govern content of the finding-aid and the depth of description which is

---

[132] Michael Cook, « The British move toward standards of archival description: the "MAD" standard », *art. cit.*, p. 133-134.
[133] *Ibid.*, p. 134-135.
[134] David Robinson, « Theory, pratice and a Manual of Archival Description », *art. cit.*, p. 2.
[135] *Ibid.*

appropriate »[136]. Aussi, parce que MAD2 semble ne pas parvenir à apporter d'exemples concrets pour l'application des règles édictées, son intitulé se voit être remis en cause, « the title 'The Theory of Archival Description' would be better »[137]. Théorique, c'est encore ce qui lui est reproché face à la règle qui énonce d'attribuer un numéro particulier en fonction du niveau de classement à traiter. Bien que faisant sens sur le papier, la pratique fait émerger des incohérences d'ordre varié : amalgames possibles entre les niveaux 2 et 3, difficulté à déterminer l'organe de tutelle du service d'archives pour les niveaux 1 et 2, terminologie confuse entre les niveaux 4 et 5[138].

Les membres de l'*Archival Description Project* sont cependant convaincus que MAD2 pourrait beaucoup apporter aux archivistes du monde entier, et suggèrent que leurs travaux devraient bénéficier d'une visibilité plus grande, « The MAD project team concludes that there is a great potential in the descriptive standards they are putting forward, and that they should be made known internationally »[139]. Pourtant, les principes sur lesquels reposent la norme sont déjà largement adoptés en Europe en 1991, et ce peut-être en raison de la faible présence des pratiques issues des bibliothèques au sein de la communauté archivistique de ce côté-ci de l'Atlantique[140]. Mais les membres de l'*Archival Description Project* sont persuadés qu'un terrain d'entente avec les pratiques états-uniennes en termes de description archivistique est possible, « The apparent conflict between the MAD approach, which structures finding aid systems for use in repositories, and the United-States standards, which structure bibliographic descriptions of archives, can be resolved »[141].

2.3. L'initiative canadienne, un compromis entre l'approche américaine et britannique ?

La réflexion autour de la normalisation de la description archivistique canadienne débute un peu plus tôt que son voisin ou le Royaume-Uni, puisque déjà en 1975, « se met progressivement en place une structure nationale de coopération vers la normalisation de la description archivistique »[142]. Il faut cependant attendre le début des années 1980 pour voir le processus prendre de la vitesse. En effet, c'est à cette époque que le *Bureau of Canadian Archivists* (BCA) reçoit de la part du *Social Sciences and Humanities Research Council* (SSHRC) une importante subvention de 97 250$ devant servir à financer le *Canadian Working Group on*

---

[136] David Robinson, « Theory, pratice and a Manual of Archival Description », *art. cit.*, p. 2.
[137] *Ibid.*, p. 3.
[138] *Ibid.*, p. 3-4.
[139] Michael Cook, « The British move toward standards of archival description: the "MAD" standard », *art. cit.*, p. 138.
[140] Michael Cook, « MAD2: Reassessing the Experience », *Archivaria*, vol. 35, 1993, p. 16.
[141] Michael Cook, « The British move toward standards of archival description: the "MAD" standard », *art. cit.*, p. 138.
[142] Claire Sibille de Grimoüard, « Normes internationales et influence française », *La Gazette des archives*, n°256, 2019, p. 19-20.

Archival Descriptive Standards[143]. Bien qu'un parallèle entre les objectifs canadiens et états-uniens pourrait être établi, le contexte de travail et les résultats obtenus divergent[144]. Au sujet du contexte de travail, les archivistes canadiens ne disposent que d'une année pour proposer plusieurs projets de description archivistique à la communauté professionnelle[145].

Cependant, l'ampleur de la mission est reconnue trop importante, « it would be impossible, not to mention foolhardy, to attempt to develop standards in a year »[146] et le groupe de travail a finalement opté pour la préparation d'un rapport — intitulé *Toward Descriptive Standards* — qui voit le jour en 1985. Ce rapport, riche de trente-cinq recommandations organisées en huit catégories, constitue le fondement de tous futurs travaux normatifs pour la description archivistique au Canada. La première catégorie énonce que le processus normatif de la description archivistique doit se réaliser selon le principe du respect des fonds. En effet, la recommandation n°1 appelle au respect de la provenance en énonçant que tout archiviste canadien se doit de réaliser la description au niveau du fonds, « The concept of fonds is inextricably related to provenance and the principle of provenance »[147]. Une recommandation qui reflète l'approche nommée *Total Archives* appliquée au Canada, c'est-à-dire le fait de collecter à la fois les archives publiques et privées et sous toutes leurs formes, qu'ils s'agissent de dessins architecturaux, de microfilms, de cartes, etc. ; raison pour laquelle RAD peut être, « characterized as "Total Description" »[148]. Les recommandations n°3, n°4 et n°6 suivent la notion de respect de l'ordre originel d'un fonds, « levels of arrangement will determine the level of description, and accurate descriptions should represent the relative levels of arrangement »[149]. Enfin, la recommandation n°5 suggère que la description se réalise du général au particulier pour faciliter la compréhension contextuelle, « Users must know the context in which the records they are consulting have been created »[150].

Lors de sa dernière réunion, le *Canadian Working Group on Archival Descriptive Standards* demande au BCA d'établir un comité pour la normalisation de la description archivistique afin de s'assurer que le processus normatif prenne appui sur ces recommandations[151]. La recommandation 31 du rapport de 1985 encadre la procédure, « a) committee members should be known to have expertise in and practical working knowledge of the issue concerned; b) each

---

[143] Wendy M. Duff, Kent M. Haworth, « The Reclamation of archival description: the Canadian perspective », *art. cit.*, p. 29.
[144] *Ibid.*
[145] *Ibid.*
[146] Canadian Working Group on Archival Description, *Toward Descriptive Standards*, Ottawa, 1985, p. 5.
[147] Hugo Stibbe, « Implementing the Concept of Fonds: Primary Access Point, Multilevel Description and Authority Control », *Archivaria*, vol. 34, 1992, p. 109.
[148] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *art. cit.*, p. 56.
[149] Wendy M. Duff, Kent M. Haworth, « The Reclamation of archival description: the Canadian perspective », *art. cit.*, p. 32.
[150] Planning Committee on Descriptive Standards, « Preface to the 1990 edition », *Rules for Archival Description*, Ottawa, 2008, p. xviii.
[151] Kent M. Haworth, « The Development of descriptive standards in Canada: a progress report », *Archivaria*, vol. 34, 1992, p. 75.

committee should have at least one representative from the Public Archives of Canada »[152]. Le comité — le *Planning Committee on Descriptive Standards* (PCDS) — est ainsi créé en 1986 et est composé de deux représentants de l'Association des archivistes du Québec (AAQ) ainsi que de deux représentants de l'Association of Canadian Archivists (ACA)[153]. Lors de sa première réunion en janvier 1987, le PCDS s'agrandit et accueille le secrétaire général du BCA et un représentant des archives nationales canadiennes[154]. Puis, plusieurs groupes de travail sont créés par le comité, chacun d'entre eux étant assigné à la description d'un support d'archivage à tous les niveaux de la description : fonds, série, dossier et pièce[155]. La même année entre en vigueur le *National Archives of Canada Act*, un texte qui permet alors aux Archives nationales canadiennes d'accorder des fonds pour soutenir la profession dans la normalisation de la description archivistique[156] ; se conjugue à cela le travail déjà réalisé dans le rapport de 1985 et toutes les circonstances sont propices pour mener à bien la normalisation des pratiques en la matière[157]. Ainsi, le PCDS et ses groupes de travail récoltent le fruit de leur labeur avec la parution de la norme intitulée *Rules for Archival Description* (RAD) — en 1990. Une norme qui semble parvenir à un terrain d'entente entre les pratiques états-uniennes et britanniques[158].

En effet, RAD semble dialoguer avec les deux traditions archivistiques, « RAD has formulated its rules based on the archival principle of *respect des fonds* and the bibliographic structure of the ISBD(G) »[159]. La norme ISBD ayant elle-même inspirée la révision des AACR, il n'est pas surprenant de constater que RAD en emprunte l'ossature, « The structure of this work [RAD] follows very carefully that of AACR2 with Part I devoted to rules for Description and Part II to Headings, Uniform Titles, and References »[160]. A l'image de APPM2 donc, RAD se compose d'une partie dédiée à la description du fonds et d'une seconde partie consacrée aux points d'accès[161]. La norme canadienne va plus loin encore que APPM2, en respectant fidèlement la structure de ISBD(G), « APPM (Rule 0.10): "… a collection's inclusive dates are considered an integral part of its title." RAD retains date information in the ISBD(G) area — Area 4: Publication, Distribution, etc., Area prescribed for date information »[162]. En revanche, RAD fait écho à MAD au sujet de la description à plusieurs niveaux, « RAD, like MAD, places special emphasis on multilevel description »[163]. RAD se distingue cependant de la vision états-unienne et de

---

[152] Kent M. Haworth, « The Development of descriptive standards in Canada: a progress report », *art. cit.*, p. 75.
[153] Wendy M. Duff, Kent M. Haworth, « The Reclamation of archival description: the Canadian perspective », *art. cit.*, p. 30.
[154] *Ibid.*, p. 30-31.
[155] *Ibid.*, p. 31.
[156] *Ibid.*, p. 30.
[157] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *art. cit.*, p. 55.
[158] Keith Stotyn, « Planning Committee on Descriptive Standards, Rules for Archival Description », *Archivaria*, vol. 32, 1991, p. 163.
[159] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *art. cit.*, p. 56.
[160] Steven L. Hensen, « RAD, MAD, and APPM: the search for Anglo-American standards for archival description », *Archives and Museum Informatics*, vol. 5, n°2, 1991, p. 3.
[161] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *art. cit.*, p. 58.
[162] *Ibid.*, p. 62.
[163] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *art. cit.*, p. 59.

l'approche britannique en n'ayant pas de format associé, « RAD does not prescribe products […] it does not provide guidance on the types of finding aids archives should develop »[164]. Selon le PCDS, cela est du ressort des institutions[165].

Reflet des pratiques archivistiques canadiennes en matière de description archivistique, le développement de RAD se différencie des procédés normatifs états-unien et britannique pour avoir été pensé et conçu selon un processus consultatif, « Rules for Archival Description is a microcosm of the nation which has produced it »[166]. Cette spécificité est inscrite à la recommandation n°32 du rapport *Toward Descriptive Standards*, l'objectif étant d'engager le plus d'archivistes possible et que la normalisation ne soit pas que l'affaire d'un petit groupe, « Unlike Great Britain or the United States, where descriptive standards development has been largely the initiative and responsibility of one or two individuals, in Canada working groups are responsible for the presentation of a report to the PCDS »[167]. Cette même recommandation permet non seulement de rendre compte des besoins et spécificités des pratiques canadiennes mais aussi d'engager doucement la description archivistique canadienne sur la voie de la normalisation internationale, en suggérant que ces propositions de normes soient approuvées et adoptées par les organismes internationaux concernés[168]. Une suggestion qui fait écho à la recommandation n°2, inscrite dans la catégorie des principes sur lesquels doit reposer tout processus de normalisation, qui suggère que le BCA demande à l'ICA d'élaborer et de solliciter l'approbation d'adopter un cadre normatif relatif à la description de fonds d'archives[169].

Ce terme, « fonds », est directement emprunté au français afin d'éviter des problèmes de confusions terminologiques liés aux expressions « record group », « manuscript group » ou encore « collection »[170]. RAD étant le reflet des pratiques archivistiques canadiennes, il n'est pas surprenant d'y déceler les couleurs de l'Hexagone, « [RAD] is a blend of French-Canadian and English-Canadian archival thought »[171]. En ce sens, le Canada fait figure de passerelle entre les diverses pratiques en matière de description archivistique, « Au fil des années, le Canada revendique une place d'acteur de premier plan sur la scène internationale. Il […] entretient des relations panaméricaines et dans le cadre du Commonwealth, tout en assumant des relations "privilégiées" avec la France et le Royaume-Uni »[172] ; raison pour laquelle le Canada s'intéresse de près au développement de la normalisation de la description archivistique sur le plan international[173]. Il en est le précurseur, « The proposal to develop international standards of

---

[164] Planning Committee on Descriptive Standards, « Preface to the 1990 edition », *op. cit.*, p. xix.
[165] *Ibid.*, p. xx.
[166] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *art. cit.*, p. 56.
[167] *Ibid.*
[168] Kent M. Haworth, « The Development of descriptive standards in Canada: a progress report », *art. cit.*, p. 75-76.
[169] *Ibid.*, p. 76-77.
[170] Canadian Working Group on Archival Description, *Toward Descriptive Standards*, Ottawa, 1985, p. 7.
[171] Kent M. Haworth, « The voyage of RAD: from the Old World to the New », *art. cit.*, p. 56
[172] Bénédicte Grailles, « Archivistiques canadienne et française : échanges, influences ou transferts ? », *Le théâtre du patrimoine (France-Canada)*, Paris, CTHS, 2008, p. 55-56.
[173] Kent M. Haworth, « The Development of descriptive standards in Canada: a progress report », *art. cit.*, p. 77.

description is credited to Canadian archivists »[174], et ce pour plusieurs raisons. Tout d'abord de par les progrès en matière de description archivistique visibles à travers RAD, mais également en tant qu'hôte du 12ème Congrès du Conseil international des Archives en 1992[175] ; aussi, il convoque — conjointement avec l'ICA — les experts en description archivistique des quatre coins du globe à Ottawa en 1988. Une réunion qui rassemble alors vingt-neuf experts issus d'une quinzaine de pays et de trois organisations non gouvernementales : l'ICA, l'IFLA et l'ISO[176], et qui signe le début d'une normalisation à l'échelle internationale.

[174] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *Comma*, vol. 2011, n°2, 2011, p. 51.
[175] *Ibid.*
[176] Christine Nougaret, « Vers une normalisation internationale de la description des archives. La norme ISAD(G) du Conseil international des archives », *La Gazette des archives*, n°169, 1995, p. 275.

3. ICA, chronique de la description archivistique internationale

En reconnaissant la nécessité d'établir un cadre normatif international en termes de description archivistique, le groupe d'experts de 1988 demande à l'ICA la mise en place d'un groupe de travail pour mener à bien ce projet. Une requête entendue par l'organisation, « En décembre 1989, le Conseil international des archives accédait à cette requête en organisant à Paris, avec l'Unesco, une réunion consultative restreinte afin d'établir un plan d'action international à long terme pour l'élaboration de normes de description »[177] ; réunion dont les membres ne sont autres qu'un « noyau d'experts de la réunion d'Ottawa »[178]. Les experts en arrivent tous à la même conclusion : avant d'entrer dans le vif du sujet, la discipline doit se doter de principes sur lesquels s'appuyer, « This points to the need for an international congress of archivists to produce a Statement of Principles such as those endorsed by librarians in Paris in 1961 »[179]. L'année 1990 marque un tournant, non seulement un protocole d'accord fait des Archives nationales canadiennes le chef d'orchestre de ce pojet de grande envergure, « The National Archives of Canada's Office of Archival Descriptive Standards has agreed to coordinate the international effort by serving as the ICA Secretariat for international descriptive standards initiatives »[180], mais surtout, c'est au mois de septembre de cette même année, à Wroclaw, que voit le jour la commission *ad hoc* sur les normes de description (DDS)[181]. Une commission — principalement composée des membres de la réunion de Paris en 1989 — où huit pays sont représentés parmi lesquels figurent l'Espagne, les États-Unis, la France, la Malaisie, le Portugal, le Royaume-Uni, la Suède et bien entendu le Canada[182]. La promesse d'un cadre normatif international en la matière n'est alors plus très loin.

3.1. De la matière première au produit fini : normalisation de la description archivistique

1.1.1. De l'importance des principes

La première séance plénière de la DDS se tient l'année suivante, en 1990 à Höhr-Grenzhausen, avec pour objectif de définir un Énoncé des principes ; principes qui, à l'instar d'une boussole, se doivent d'amener la description archivistique à bon port, « a Statement of Principles respecting archival description was drafted first, to provide a foundation on which to

---

[177] Christine Nougaret, « Vers une normalisation internationale de la description des archives. La norme ISAD(G) du Conseil international des archives », *art. cit.*, p. 275.
[178] *Ibid.*
[179] Wendy M. Duff, Kent M. Haworth, « The Reclamation of archival description: the Canadian perspective », *art. cit.*, p. 32.
[180] *Ibid.*, p. 35.
[181] Christine Nougaret, « Vers une normalisation internationale de la description des archives. La norme ISAD(G) du Conseil international des archives », *art. cit.*, p. 276.
[182] *Ibid.*

proceed with the drafting of rules »[183]. La première version est soumise à commentaires au sein de l'ICA[184], et l'année suivante, « un sous-comité de 4 membres (Canada, États-Unis, Royaume-Uni) élabore à Liverpool une ébauche des règles générales »[185]. Les deux textes — l'Énoncé des Principes et celui comportant les règles de description archivistique — sont présentés lors de la seconde session plénière de la DDS à Madrid en 1992 où ils sont adoptés, puis diffusés à la communauté professionnelle internationale pour réflexion avant discussion au cours du 12ème Congrès de l'ICA[186].

En septembre 1992 à Montréal, les congressistes se voient l'opportunité d'échanger directement avec les membres de la DDS grâce à la mise en place d'une séance publique[187]. Les discussions sont animées et les premiers efforts normatifs controversés, « There was opposition to some of the principles, from the United States and the UK but mainly from Australia »[188]. L'objet du désaccord se trouve être le concept de fonds et le fait que la description doive avoir lieu après que le classement du fonds ait été réalisé[189] ; une recommandation qui découle directement des principes de RAD, « A second axiom follows from the observance of *respect des fonds*: the way archives depends on their arrangement. […] A fonds cannot be described until it has been arranged »[190]. A l'issue du congrès, la DDS compte alors un nouveau membre, Chris Hurley, représentant de l'Australie ; une décision à la fois théorique — Chris Hurley apparaît comme le porte-parole des points de vue divergents[191] — et politique — l'Australie est membre du Commonwealth, et la signature de la Déclaration de Harare en 1991 rappelle l'importance du lien qui en unit les membres, « there are shared principles for peace, equality, prosperity and democracy »[192] —. Pour autant, le texte n'est pas révisé au cours de la troisième séance plénière de la DDS de 1993 à Stockholm, mais n'est suivi d'aucune publication[193]. Une décision perçue par Chris Hurley comme un moyen de couper court aux débats, « a manœuvre to cut off further discussion »[194]. Une décision non sans conséquence puisque bon nombre d'archivistes ont, suite à cela, tourné le dos à *International Standard Archival Description – General* (ISAD/G)[195] ; norme

---

[183] Wendy M. Duff, Kent M. Haworth, « The Reclamation of archival description: the Canadian perspective », *art. cit.*, p. 33.
[184] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 52.
[185] Christine Nougaret, « Vers une normalisation internationale de la description des archives. La norme ISAD(G) du Conseil international des archives », *art. cit.*, p. 276.
[186] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 52.
[187] *Ibid.*
[188] *Ibid.*
[189] *Ibid.*
[190] Planning Committee on Descriptive Standards, « Preface to the 1990 edition », *op. cit.*, p. xviii.
[191] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 52.
[192] Françoise Grellet, *A Cultural guide - Précis culturel des pays du monde anglophone*, Paris, Nathan, 2015, p. 101.
[193] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 52.
[194] Jennifer Bunn, « Developing descriptive standards: a renewed call to action », *Archives & Records*, vol. 34, n°2, p. 238.
[195] Jennifer Bunn, « Developing descriptive standards: a renewed call to action », *art. cit.*, p. 238.

qui quant à elle, fut pourtant révisée à la lumière des commentaires reçus pendant le congrès au cours de la troisième séance plénière du groupe de travail[196]. La même année, la norme est approuvée par la 29e Conférence internationale de la Table ronde des archives (CITRA), puis par l'ICA en 1994. Le texte est alors publié au sein de *Janus* — la revue scientifique de l'ICA — en anglais et en français[197] ; une publication qui signe l'entrée officielle d'ISAD/G au sein de la communauté professionnelle.

1.1.2. ISAD/G, une norme couteau suisse

ISAD/G s'inscrit dans une longue lignée de normes issues du monde anglo-saxon, « For the development of ISAD(G), the Commission drew heavily on the previous experiences of standards development already undertaken — in APPM, MAD and RAD — »[198] ; la volonté de prendre en compte le plus de traditions archivistiques possibles anime cependant le groupe de travail[199]. Un héritage qui se traduit dans l'Énoncé des principes à travers trois préceptes : l'affirmation du respect des fonds « especially in view of the increasing complexity of administrative structures, an explanation of the context in which the material was created is an important aspect of archival description »[200]. Un principe d'autant plus essentiel dans un contexte où les structures administratives se complexifient, « also in view of this complexity, it may sometimes be necessary for the archivist to regard the fonds as an intellectual rather than a physical concept »[201]. Un second précepte rappelle que les niveaux de classement et les niveaux de description sont étroitement liés, « the Commission has set out a hierarchical model of the levels of arrangements for the fonds and its constituents parts. These are levels of description, with differing degrees of detail, appropriate to each level of arrangement »[202] ; un précepte qui implique que la description se réalise du général au particulier, « the principles that archival description proceeds from the general to the specific is the practical consequence of the principles of respect des fonds »[203] ; enfin le dernier principe concerne la description à plusieurs niveaux, « A hierarchical format is convenient for showing the different levels of description »[204]. Le résultat final est une norme de 26 éléments descriptifs répartis en 6 zones, un outil pour les archivistes à qui, « elle offre une structure cohérente, concise et complète »[205] mais également

---

[196] Christine Nougaret, « Vers une normalisation internationale de la description des archives. La norme ISAD(G) du Conseil international des archives », *art. cit.*, p. 277.
[197] *Ibid.*
[198] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 52.
[199] *Ibid.*, p. 52-53.
[200] International Council on Archives, « Statement of Principles regarding archival description », *Archivaria*, vol. 34, 1992, p. 10.
[201] *Ibid.*
[202] *Ibid.*, p. 11.
[203] *Ibid.*, p. 13.
[204] *Ibid.*, p. 14.
[205] Christine Nougaret, « D'ISAD/G à ISAAR : un nouvel outil du Conseil International des Archives pour la description archivistique », *La Gazette des archives*, n°176, 1997, p. 102.

pour l'utilisateur final pour qui il s'agit d'un « outil clair et homogène d'appréhension des sources archivistiques complexes et diversifiées »[206].

1.1.3. Un format de description archivistique fait *par* des archivistes *pour* des archivistes[207]

Cependant, ISAD/G peine à s'implanter aux Etats-Unis car au même moment se développe le *Berkeley Finding Aid Project* (BFAP)[208], un projet ayant pour objectif de doter la discipline d'un format d'échange de données, « the excitement and debate it generated in U.S. archival circles served to divert attention and effort from archival data content standards to data structures »[209]. Le résultat ne se fait pas attendre puisque déjà en 1993 nait une ébauche de définition de type de document, la *FindAid DTD*[210]. Deux ans plus tard, la *Commission on Preservation and Access* (CPA) et la UC Berkeley library organisent une conférence afin d'évaluer ce premier résultat et de préparer son devenir[211]. C'est un succès, et la conférence se clôture sur la conviction d'impliquer davantage d'experts en description archivistique au sein du projet, un élément essentiel pour Daniel Pitti qui est alors à la tête de l'équipe, « Standards are the products of communities, not of individuals working in splendid isolation, and the development process is as much a political exercise as it is an intellectual and technical undertaking »[212].

L'équipe expérimente la *FindAid DTD* au sein du Bentley Library Research Fellowship Program de l'université du Michigan, « Daniel Pitti bravely assembled a group of seven archivists and one expert in Standard Generalized Markup Language (SGML) in July 1995 to critique and revise his data content model »[213]. Une expérience qui se réalise donc main dans la main avec des archivistes experts en description archivistique — parmi lesquels figurent des membres du NISTF et du WGSAD[214] —, et l'informaticien Steven J. DeRose[215], reconnu pour avoir contribué au développement de la norme ISO 8879 de 1986. De cette réunion naît un accord relatif aux principes de conception nommé *Ann Arbor Accords* — du nom de la ville où se situe l'université du Michigan —, et la *FindAid DTD* est renommée *Encoded Archival Description* (EAD)[216]. A la suite de cette réunion paraît la version 0.1 de l'EAD ; une version et un accord qui sont en septembre 1995 présentés à la CAIE lors de la réunion annuelle de la SAA, « CAIE was invited

---

[206] *Ibid.*
[207] Kris Kiesling, « EAD as an archival descriptive standard », *The American Archivist*, vol. 60, 1997, p. 347.
[208] William E. Landis, « Overcoming the bibliographic conundrum in archival description », *The American Archivist*, vol. 74, 2011, p. 14.
[209] *Ibid.*, p. 15.
[210] Daniel Pitti, « Encoded Archival Description : The development of an encoding standard for archival finding aids », *The American Archivist*, vol. 60, 1997, p. 279.
[211] *Ibid.*
[212] *Ibid.*
[213] Janice E. Ruth, « Encoded Archival Description: a structural overview », *The American Archivist*, vol. 60, 1997, p. 311.
[214] William E. Landis, « Overcoming the bibliographic conundrum in archival description », *art. cit.*, p. 15.
[215] Daniel Pitti, « Encoded Archival Description : The development of an encoding standard for archival finding aids », *art. cit.*, p. 280.
[216] Daniel Pitti, « Encoded Archival Description : The development of an encoding standard for archival finding aids », *art. cit.*, p. 280.

to become formally involved in the development of EAD. CAIE agreed and created the EAD Working Group (EADWG) »[217]. Le rôle de l'EADWG est de contribuer à l'élaboration d'un format pour les instruments de recherche, de procéder à l'examen de la DTD EAD présentée et des directives d'application relatives au format[218] ; « Before EAD can be endorsed as a formal SAA standard, it must pass through a fairly rigorous review process »[219], l'objectif final étant de faire approuver la DTD EAD d'abord par le comité des normes de la SAA, puis par le Conseil de l'association professionnelle. Pour ce faire, l'EAD doit s'appuyer sur une documentation officielle, « The EAD documentation will comprise four technical documents: the DTD, the tag library, application guidelines, and possibly a compendium of practice »[220].

Le premier document technique est ainsi la DTD elle-même. Mais en quoi consiste-t-elle exactement ? Comment cela fonctionne-t-il ? Pourquoi avoir choisi de créer un tel modèle de données alors que le pays dispose déjà du format MARC AMC développé en 1984 par la CAIE ? SGML, le langage qui sous-tend la DTD, est lui aussi une norme au même titre que le format MARC, « It comprises a formal set of conventions in the public domain, and thus is not owned by and thereby dependent on any hardware or software producer »[221]. Le groupe de travail est conscient de l'importance de ce critère s'ils veulent que la DTD qu'ils élaborent perdure ; le domaine public fait ainsi office d'assurance vie pour le futur format. Cependant, SGML présente de nombreux avantages que MARC AMC ne peut pas offrir. En effet, le format MARC présente une structure plate peu adaptée à la représentation d'une information structurée hiérarchiquement, « MARC accomodates hierarchically structured information very poorly »[222], ce que à quoi SGML répond parfaitement, « unlike MARC, SGML accomodates hierarchically interrelated information at as many levels as needed »[223]. Enfin, SGML s'avère être également plus approprié en raison de l'absence d'une limitation du nombre de caractères utilisables[224] contrairement au format MARC, « MARC records are limited to a maximum lenght of 100,000 characters »[225] ; un problème dès lors que bon nombre d'instruments de recherche sont plus conséquents que cela. Le dernier argument du SGML réside dans le marché beaucoup plus important qu'il représente[226]. Ainsi, le langage de balisage descriptif sur lequel repose SGML — un langage qui permet une description des données — semble parfaitement répondre à la volonté du groupe de travail de pouvoir représenter la complexité propre aux instruments de recherche tout en permettant une manipulation pointue desdits instruments[227].

---

[217] *Ibid.*
[218] *Ibid.*
[219] Kris Kiesling, « EAD as an archival descriptive standard », *art.cit.,* p. 348.
[220] *Ibid.*, p. 349.
[221] Daniel Pitti, « Encoded Archival Description : The development of an encoding standard for archival finding aids », *art. cit.*, p. 276.
[222] *Ibid.*, p. 275.
[223] *Ibid.*, p. 276.
[224] *Ibid.*
[225] *Ibid.*, p. 275.
[226] Daniel Pitti, « Encoded Archival Description : The development of an encoding standard for archival finding aids », *art. cit.*, p. 276.
[227] *Ibid.*, p. 275.

Le groupe de travail doit alors s'approprier le fonctionnement d'une DTD : comment cela fonctionne-t-il ? Comment l'adapter aux besoins de la communauté professionnelle en matière de description archivistique ? Pour le comprendre, il est intéressant de souligner le parallèle entre DTD et grammaire. En effet, à l'image d'un ensemble de règles qui déterminent l'emploi correct d'une langue, la DTD définit la structure d'un document : des champs de données nommés « éléments », leurs attributs, ainsi que leur agencement. Pour ce faire, le groupe de travail s'appuie sur les *Ann Arbor Accords* de 1995, « these principles or accords stated the group's philosophy and outlined the DTD's intent, parameters, and structural features »[228]. Le résultat obtenu prend en compte deux types d'information : des informations relatives à l'instrument de recherche lui-même traduites à travers les éléments <eadheader> et <frontmatter>, et des informations relatives au fonds d'archives transcrites au sein de l'élément <archdesc>[229].

Enfin, la DTD a été pensée et conçue comme un format susceptible de répondre aux besoins de la communauté professionnelle internationale et en ce sens, ne pouvait ignorer le développement de la normalisation internationale en la matière, « Although EAD began its development in the United States based on archival descriptive practices used in this country, its progress has been closely followed by members of the international archival committee »[230] ; un intérêt visible à travers la présence de Sharon Gibbs Thibodeau, membre de la DDS de 1990 à 1997, au sein de l'équipe de Daniel Pitti[231]. Afin de faciliter l'adoption du format à l'international, le groupe de travail a opté pour un langage neutre au sein de la DTD, « Terms such as collection, personal papers, archives, series, fonds, and record group were avoided in favor of "generic terms like unit and component that are not specific to any individual setting or institution." »[232]. Une comparaison entre la DTD et la norme de l'ICA témoigne de la possibilité pour l'EAD d'encoder un instrument de recherche réalisé avec ISAD/G[233], « This compatibility between EAD and the ISAD(G) affirms the validity of the EAD structure and suggests a role for EAD in fostering the international exchange of archival descriptions »[234]. La version 1.0 de la DTD EAD est publiée en août 1998.

Cependant, la DTD à elle-seule ne peut suffire pour prétendre au statut de cadre normatif, et est accompagnée d'un dictionnaire des balises — *EAD Tag Library* — afin de déterminer les éléments que peut contenir la DTD ainsi que de directives d'application — *EAD Application Guidelines* — afin de s'assurer de la bonne mise en œuvre du format. L'importance de ces documents est d'autant plus grande au vu de l'idiosyncrasie qui régnait il n'y a pas si longtemps sur la profession, « For many decades, the archival way of thinking was characterized by the

---

[228] Janice E. Ruth, « Encoded Archival Description: a structural overview », *art. cit.,* p. 313.
[229] *Ibid.*, p. 318.
[230] *Ibid.*, p. 316.
[231] *Ibid.*
[232] *Ibid.*, p. 314.
[233] *Ibid.*, p. 317.
[234] *Ibid.*

above-mentioned motto: "my archives are unique" — unifying these unique "universes" seems like very real challenges and job »[235]. Un défi humain, « do we have staff able to work on it ? »[236] possible à relever à en croire le projet britannique *Access to Archives* (A2A), « [A2A] provides a retroconversion service and access to descriptive metadata about archives held by local archives, museums and libraries throughout England »[237].

En 2002, le contexte du web 2.0 et l'apparition de l'*Extensible Markup Language* (XML) en 1999 — un langage jugé plus facile à utiliser sur le web[238] — conduit le groupe de travail à imaginer et créer la version 2.0 de la DTD EAD qui paraît alors en 2002. Le langage XML qui sous-tend la DTD de 2002 est en effet plus adapté au contexte archivistique que le *HyperText Markup Language* (HTML) paru en 1993. Ce dernier relève du langage de balisage procédural, c'est-à-dire qu'il décrit la mise en forme d'un document ; un langage peu adapté aux archives, « HTML peine à gérer les documents de gros volume ou fortement structurés, où la recherche plein texte est insuffisante. Il ne permet pas de vérifier le contenu des données, et ne donne donc pas de possibilité d'aider l'internaute dans ses requêtes »[239]. De son côté, le XML constitue un langage de balisage descriptif qui est à même de refléter la description à plusieurs niveaux, « Tout document XML est un arbre d'éléments, permettant de décomposer le texte en unités d'information »[240]. Autre avantage de l'XML : il est plus simple d'utilisation que le SGML, « SGML is complicated for both application developers and the users of the applications »[241].

3.2. La description contextuelle sur la voie de l'émancipation : vers une logique de réseau de relations

3.2.1. Vers l'indépendance de la description du producteur

Peu de temps après a parution de la première édition d'ISAD/G, de nouveaux besoins émergent en matière de description archivistique. En effet, bien qu'étant reconnu comme un élément central de la description au sein des principes de 1992, la notion de provenance qui gravite autour du principe de respect des fonds est peu mise en avant au sein de la norme de 1994, « Le constat des limites d'ISAD(G) pour produire une description fine et structurée des producteurs — seuls deux éléments s'y consacrent — conduit au développement d'une nouvelle norme »[242]. Cette nouvelle norme, *International Standard Archival Authority Record for*

---

[235] Hubert Wajs, « The Polish road to EAD », *Comma*, vol. 2005, n°2, 2005, p. 4.
[236] *Ibid.*
[237] William Stocking, « 'A Flexible friend': the role of Encoded Archival Description (EAD) in the promotion of access to archives in the United Kingdom », *Comma*, vol. 2005, n°2, 2005, p. 1.
[238] Daniel Pitti, « Encoded Archival Description : The development of an encoding standard for archival finding aids », *art. cit.*, p. 278.
[239] Claire Sibille de Grimoüard, « Les instruments de recherche électroniques », *Comma*, vol. 2006, n°1-2, 2006, p. 3.
[240] *Ibid.*, p. 4.
[241] Daniel Pitti, « Encoded Archival Description : The development of an encoding standard for archival finding aids », *art. cit.*, p. 278.
[242] Alice Motte, « La normalisation de la description archivistique : enjeux et actualités », *art. cit.*, p. 123.

Corporate Bodies, Persons and Families (ISAAR/CPF), est pensée dès 1993 où un sous-groupe de la DDS se penche sur l'élaboration d'une première ébauche inspirée des notices bibliographiques[243]. Le sous-groupe de travaille se réunit alors à Liverpool en 1993 puis à La Haye en 1994 où la première version est soumise à commentaires ; des commentaires sur lesquels les membres se sont appuyés pour réviser une dernière fois la norme en 1995 lors d'une réunion à Paris avant de pouvoir la publier en 1996[244].

Une question demeure : pourquoi avoir fait le choix de la notice d'autorité ? Deux raisons à cela : tout d'abord, le constat que le nom de la provenance, quel qu'il soit, puisse évoluer à travers le temps et nécessite donc d'être normalisé « pour en garantir l'identification et permettre ainsi des recherches pertinentes »[245]. Deuxièmement, l'indépendance de la description du producteur « assure une plus grande souplesse dans l'élaboration, la mise à jour et l'utilisation de ces différentes descriptions », il est alors possible, par exemple, d'apporter des modifications sur un producteur sans avoir à effectuer lesdites modifications pour toutes les descriptions d'archives le concernant. Une notice d'autorité contient alors deux éléments essentiels, « le contrôle d'autorité du nom utilisé »[246] et « un cadre de description structurée présentant le producteur de façon exhaustive »[247]. La nécessité d'un contrôle d'autorité du nom est d'autant plus nécessaire dans un contexte où les stuctures administratives se complexifient comme vu précédemment, « Providing authority control is not unlike the process of creating order and meaning out of a bowl of alphabet soup. Authority control provides the means to create intellectual order out of the chaos of documents received in archival repository »[248]. Un contrôle d'autorité dont la forme n'est pas dictée par la norme de description des producteurs ; les membres de la DDS ayant préféré laisser les conventions nationales en vigueur cadrer les pratiques en la matière[249].

3.2.2. Le 13ème Congrès international des archives et la création du Comité sur les normes de description : quels apports pour la discipline ?

En 1996, l'organisation de la normalisation de la description archivistique évolue. En effet, l'Assemblée générale du 13ème Congrès international des archives, qui se tient alors à Pékin, fait le choix de dissoudre la DDS pour le remplacer par un Comité permanent sur les normes de

---

[243] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 53.
[244] *Ibid.*
[245] Christine Nougaret, « D'ISAD/G à ISAAR : un nouvel outil du Conseil International des Archives pour la description archivistique », *art.cit.*, p. 102.
[246] *Ibid.*, p. 104.
[247] Alice Motte, « La normalisation de la description archivistique : enjeux et actualités », *art. cit.*, p. 123.
[248] Cynthia J. Durance, « Authority Control: Beyond a bowl of alphabet soup », *Archivaria*, vol. 35, 1993, p. 38.
[249] Christine Nougaret, « D'ISAD/G à ISAAR : un nouvel outil du Conseil International des Archives pour la description archivistique », *art.cit.*, p. 104.

description — *Committee on Descriptive Standards* (CDS) —[250]. Un comité qui compte dorénavant un représentant d'Amérique latine et qui a quelque peu revu son fonctionnement puisque désormais, « the initial distinction between Commission members, experts or advisory members was diluted and eventually disappeared. In addition, group decisions were always taken by consensus and not by vote which […] ensured that the positions taken were accepted by all »[251] ; un pas de plus vers une normalisation plus représentatrice des traditions archivistiques existantes.

La nature permanente — et par conséquent plus stable — du groupe de travail permet la création de nouvelles normes et la révision des premières[252]. Aussi le CDS s'attelle à la révision d'ISAD/G, avec l'aide de 33 commentaires reçus de 25 pays différents[253]. Parmi les objectifs de cette révision, il y a la nécessité de fournir plus d'exemples concrets pour illustrer l'application de la norme, d'apporter une plus grande clarté au niveau de l'introduction, de faire mention du concept des dates d'accumulation, de favoriser une meilleure articulation entre la norme et ISAAR/CPF mais également de mettre à disposition des directives d'application pour la description des documents électroniques[254]. Bien que la liste des éléments à revoir soit longue, le groupe entend éviter tout changement radical afin de ne pas nuire à l'adoption de la norme sur la scène internationale, « Recognising that ISAD/G still needed to be accepted and applied throughout the world, CDS chose to avoid radical changes »[255]. La seconde édition d'ISAD/G paraît finalement en 1999, et est rendue publique lors du 14ème Congrès de l'ICA à Séville en 2000[256]. Une nouvelle édition suivie de près par la révision de la norme de description des producteurs.

En effet, la révision d'ISAAR/CPF débute l'année suivante, en 2001, à Bruxelles. Cinq ans après sa première parution, le bilan est peu concluant, « Unlike ISAD, the ISAAR standard was little known and even less used — archivists not only found it complex but also had difficulty understanding what good would come of it and why it was important »[257]. L'un des changements majeurs effectués par le CDS est l'ajout d'une zone de relations ; un élément alors révolutionnaire qui permet de relier les notices d'autorité entre elles, mais aussi de les lier aux instruments de recherche, « Un réseau de liens entre notices d'autorités et ressources associées se crée qui, exploité par informatique, permet de nouveaux modes de recherche, décloisonnés et plus lisibles »[258].

---

[250] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 53.
[251] *Ibid.*
[252] *Ibid.*
[253] *Ibid.*, p. 54.
[254] *Ibid.*
[255] *Ibid.*
[256] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 54.
[257] *Ibid.*
[258] Alice Motte, « La normalisation de la description archivistique : enjeux et actualités », *art. cit.*, p. 124.

3.2.3. Un format pour la description du producteur

Une exploitation facilitée par le développement d'un format d'échange de données pour la description contextuelle, « steps must also be taken to ensure that the standard [ISAAR/CPF] could be understood, accepted and implemented. It was therefore seen as essential to complement EAD with a format that could represent the new version of ISAAR(CPF) »[259]. Les principes du format — une DTD XML à l'image de celle de l'EAD 2002 — sont alors élaborés par l'université de Toronto et sont nommés *Principles and criteria for a model for archival context description* ou Principes de Toronto[260]. Le format structure toute notice d'autorité selon deux parties : <eacheader> et <condesc>. A l'image de l'élément <eadheader>, <eacheader> comporte toutes les informations relatives à la notice d'autorité elle-même, notamment son <eacid>, soit son identifiant unique, ou encore son <mainhis>, un élément qui décrit l'historique de maintenance de la notice d'autorité ; deux sous-éléments obligatoires[261]. L'élément <condesc> quant à lui ne comporte qu'un seul sous-élément obligatoire, <identity>, un élément qui permet non seulement d'associer un ou plusieurs noms utilisés par le producteur, mais également d'associer les formes parallèles du nom dudit producteur[262].

A l'instar de son homologue pour la description archivistique, l'EAC est également maintenu par la SAA, « We currently have an odd arrangement wherein the international archives organization, ICA, develops and maintains international descriptive standards, while a national archives organization, SAA, develops and maintains international encoding standards »[263]. La raison de ce monopole provient du CAIE établi par la SAA pour assurer la maintenance du *Data Elements Dictionary* et de MARC AMC. Aussi, lorsque l'EAD a émergé, il est paru naturel qu'il soit à son tour placé sous la responsabilité de l'association professionnelle des archivistes états-uniens[264]. Un schéma répété avec l'EAC.

3.2.4. Influence du 15ème Congrès de l'ICA sur la normalisation de la description archivistique

Le 15ème Congrès de l'ICA qui se tient à Vienne en 2004 signe une nouvelle étape pour l'ICA. En effet, une restructuration complète de l'organisation a lieu, avec notamment la fusion du CDS et du Comité d'évaluation des normes ; le résultat de cette union est la Section sur les

---

[259] Per-Gunnar Ottosson, « Implementing ISAAR(CPF): Encoded Archival Context (EAC) », *Comma*, vol. 2005, n°2, 2005, p. 1.
[260] *Ibid.*
[261] Claire Sibille de Grimoüard, « L'EAD et l'EAC, deux DTD pour la description archivistique », *Archives de France*, 2004, p. 29, [en ligne], disponible sur http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2004/pin20040504ead_eac.pdf (consulté le 22 février 2021).
[262] *Ibid.*, p. 31.
[263] Michael Rush, « The archival network: you don't get to describe records without making a few standards », *The American Archivist*, vol. 74, 2011, p. 26.
[264] Kathleen D. Roe, « Out of the primordial ooze: the genesis of U.S. archival descriptive standards », *art. cit.*, p. 6.

normes et les bonnes pratiques — *Section on Professional Standards and Best Practices* (PSBP) — ; une section dont l'objectif est de rassembler encore davantage de professionnels et d'étendre son champ d'action au-delà des normes de description et de leur évaluation[265]. Parmi les aspirations du PSBP figurent : optimiser la coordination entre les normes développées et les textes relatifs aux bonnes pratiques, gérer le processus par lequel les normes et les bonnes pratiques sont adoptées par l'ICA, encourager les projets ayant pour objectif de développer de nouvelles normes et/ou les normes existantes sont revues et mises à jour au besoin, faciliter l'échange d'expériences professionnelles entre différentes traditions archivistiques, promouvoir les normes développées et les bonnes pratiques qui les accompagnent, ainsi qu'assurer la liaison avec d'autres associations professionnelles de l'information et d'autres organismes internationaux de normalisation[266]. Les trois premiers objectifs sont retranscrits au sein du *Process for Developing and Endorsing a Standard*, un texte ayant pour objectif de guider le groupe de travail dans l'élaboration de nouvelles normes pour la description archivistique[267].

Le groupe de travail constate alors de nouveaux besoins en la matière et se penche sur l'élaboration d'une nouvelle norme. En effet, « l'élément "fonctions" (5.2.5) d'ISAAR(CPF) se révèle insuffisant pour décrire en détail et de façon structurée des variations de compétences fréquentes chez les productions d'archives »[268] ; un premier constat qui, conjugué au « besoin de constitution et de gestion autonome d'un référentiel de description des fonctions »[269], donnent naissance à la *International Standard for Describing Functions* (ISDF).

Partant de ce même constat s'est développée la *International Standard for Describing Institutions with Archival Holdings* (ISDIAH), une norme pour faciliter les recherches de l'utilisateur final, « avant de se déplacer, le chercheur désire savoir où sont conservées les archives, qui contacter, quels sont les horaires d'ouverture et de consultation, quels sont les autres services offerts comme les facilités de reproduction »[270]. Une norme aux couleurs de l'Espagne puisque « précédée par un format d'échange développé par le ministère de la Culture espagnol, l'EAG (Guide archivistique encodé) dans le cadre d'un projet de portail, le *Censo Guia* »[271]. Les deux nouvelles normes — ISDF et ISDIAH — sont communiquées lors du 16ème Congrès de l'ICA à Kuala Lumpur en 2008 ; conjuguées à ISAAR/CPF, elles permettent « une plus grande plasticité pour restituer au public des informations complexes et contextualisées les unes par rapport aux autres »[272].

---

[265] Vitor Manoel Marques da Fonseca, « The ICA description standards: the history of their creation and efforts to disseminate them », *art. cit.*, p. 55.
[266] *Ibid.*, p. 55-56.
[267] *Ibid.*, p. 56.
[268] Alice Motte, « La normalisation de la description archivistique : enjeux et actualités », *art. cit.*, p. 124.
[269] *Ibid.*
[270] Claire Sibille de Grimoüard, « Les normes internationales de description archivistique : origines, développements, perspectives », *La Gazette des archives*, n°228, 2012, p. 84.
[271] *Ibid.*, p. 85.
[272] Alice Motte, « La normalisation de la description archivistique : enjeux et actualités », *art. cit.*, p. 125.

CC BY NC ND

Conclusion de l'état des connaissances

La nature récente de l'histoire de la normalisation de la description archivistique est due à l'idiosyncrasie qui a longtemps dominée la profession. Si l'on ajoute à cela l'aspect coûteux du processus — qu'il s'agisse de moyens ou de temps —, on obtient des pratiques qui diffèrent d'un service d'archives à un autre. La situation évolue cependant à la fin des années 1970 où le paysage normatif de la discipline se dessine peu à peu. Le souhait de faciliter l'accès aux archives, la prise de conscience de faire partie d'une grande communauté — celle des sciences de l'information — et la volonté d'en devenir un acteur à part entière constituent les multiples facteurs à l'origine de ce changement. Riches de leur expérience en la matière et en raison des similitudes entre les deux professions, les bibliothèques nous ont montré la voie à suivre. Aussi, il n'est pas étonnant de déceler en nos normes un air de déjà-vu : de AACR2 à ISAD/G, il n'y a qu'un pas.

Si le but de cet état des connaissances est de comprendre l'évolution de notre cadre normatif, il nous permet également de mieux saisir les enjeux de pouvoir qui sous-tendent le processus de normalisation. Qu'il s'agisse de description bibliographique ou archivistique, une tension est palpable — parfois presque électrique — entre les pays impliqués. Les deux visions, nord-américaine et britannique, qui se font face tout au long du processus normatif de la description bibliographique se retrouvent au cœur de la normalisation de la description archivistique. Un tel impérialisme pourrait amener certaines traditions archivistiques moins — voire non — représentées à entrer dans une catégorie de pratiques "en voie d'extinction" en raison de choix normatifs qui leur seraient défavorables : « cultural nuances can easily be lost or subsumed by a dominant perspective »[273]. Une idée de rapport de force manifeste à de nombreuses reprises au cours de l'histoire de la description archivistique ; Michael Cook, en déclarant l'absence de rival sérieux qui puisse concurrencer la diffusion de MAD2[274], en est un exemple parmi tant d'autres.

La menace de disparition de certaines pratiques est d'autant plus vraie que la composition des groupes de travail est souvent la même en raison de la crédibilité apportée par l'expertise des personnes impliquées[275]. Aussi, la normalisation peut vite apparaître comme un cercle restreint de spécialistes, « Standardization is closely linked to expertise and is usually motivated by the view that there are some persons who know best. […] Reference to expert knowledge is often used to give standardization legitimacy »[276]. Un avantage, si l'on considère l'atmosphère cordiale — si ce n'est amicale — dans laquelle le processus normatif peut par ailleurs se dérouler

---

[273] Alan Bell, « Standards and standards culture: understanding the nature and criticisms of standardisation », *Comma*, vol. 2011, n°2, 2011, p. 33.
[274] Michael Cook, « MAD2: Reassessing the Experience », *art. cit.*, p. 16.
[275] Steven Hensen, « Concluding Thoughts from the Chair », *The American Archivist*, vol. 74, 2011, p. 32.
[276] Bengt Jacobsson, « Standardization and Expert Knowledge », *A World of Standards*, Oxford, Oxford University Press, 2002, p. 47.

et dont le résultat peut bénéficier[277] : processus plus rapide, écoute active, recherche de compromis, norme plus facilement adoptée.

Pour autant, il est également important de rappeler que la normalisation peut aussi être source de différends au sein même des groupes de travail ainsi que l'illustre Steven Hensen, « It was quite a process, not nearly so friendly as subsequent efforts at standards »[278] voire d'un processus presque violent : « despite the blood, the tears, the pushing and shoving that took place sometimes »[279]. De même qu'il est nécessaire de souligner que tous les postes ne se valent pas au sein desdits groupes de travail. En effet, certaines places sont plus stratégiques que d'autres, comme la présidence ou le secrétariat, « car ils permettent de déterminer à la fois l'agenda et le rythme de travail, deux ressources déterminantes pour influer sur le résultat final »[280].

Ainsi le rayonnement des pratiques issues du monde-anglosaxon au sein de l'ICA n'est plus à démontrer, « la question du standard, […] se pose d'abord sous l'influence et le modèle de la sphère anglo-saxonne et nord-américaine ; elle mûrit dans les années 1990 pour exploser réellement dans les années 2000 »[281]. Les années 2000 voient en effet apparaître un nouvel acteur de taille : l'Australie, qui domine alors le paysage normatif du *records management* à travers l'*International Organization for Standardization* (ISO)[282] et remet en question les normes élaborées par l'ICA : « In my opinion then, there is still no standard for archival description. […] Why else would there be in existence today two seemingly opposed sets of standards; those also developed under the aegis of the ICA and those developed under that of the ISO? »[283]. La nécessité de regarder au-delà des deux types de normes qui encadrent la discipline — les normes relatives au contenu des données, c'est-à-dire les normes de description archivistique et les normes relatives à la structure des données, appelées aussi formats d'échange de données — paraît nécessaire pour surmonter l'opposition qui semble s'être dressée entre les deux familles de normes. Pour y parvenir, la perspective d'un modèle conceptuel pour réconcilier les cadres normatifs apparaît idéale : « I think it would be worth exploring the emerging concept of interoperability as a possible way of both transcending the distinction between information exchange and archival description and furthering common ground between the archival world and that of technology specialists. […] I am intrigued by the idea of 'conceptual interoperability', which comes from a model developed within the discipline of Modelling and Simulation

---

[277] Steven Hensen, « Concluding Thoughts from the Chair », art. cit., p. 32.
[278] Steven Hensen, « Audience Comments and Questions », *The American Archivist*, vol. 74, 2011, p. 34.
[279] Jean Dryden, « Audience Comments and Questions », *The American Archivist*, vol. 74, 2011, p. 34.
[280] Olivier Borraz, « Les normes, instruments dépolitisés de l'action publique », *Gouverner par les normes*, Paris, Presses de Sciences Po, 2004, p. 128.
[281] Bénédicte Grailles, « Normalisations, standardisations, modélisations : acculturations professionnelles ou uniformisations identitaires ? », *art. cit.*, p. 55.
[282] Jennifer Bunn, « Developing descriptive standards: a renewed call to action », *Archives & Records*, vol. 34, n°2, p. 238.
[283] *Ibid.*, p. 241.

(M&S) »[284]. Une perspective qui soulève de nombreuses questions : quelles traditions archivistiques seront représentées ? Quels seront les étapes du processus normatif et les acteurs impliqués ? Est-ce que ce modèle conceptuel nous permettra de dialoguer mieux encore avec les communautés professionnelles apparentées ? Comment vont se traduire les métadonnées archivistiques ? Enfin et surtout, quels services pourra-t-il rendre à l'utilisateur final ?

[284] *Ibid.*, p. 244.

BIBLIOGRAPHIE

1. LA NORMALISATION

BELL (Alan), « Standards and standards culture: understanding the nature and criticisms of standardisation », *Comma*, vol. 2011, n°2, 2011, p. 25-38.

BEYEA (Marion), « Introduction : ICA et normes », *Comma*, vol. 2011, n°2, 2011, p. 3-5.

BORRAZ (Olivier), « Les normes, instruments dépolitisés de l'action publique », *Gouverner par les normes*, Paris, Presses de Sciences Po, 2004, p. 123-161.

COTTIN (Michel), « La famille des normes ISO sur le records management », *La Gazette des archives*, n°228, 2012, p. 119-133.

HOLMBLAD (Jacob), « Les normes font tourner le monde », *ISO Focus+*, vol. 1, n°2, 2010, 1 p.

JACOBSSON (Bengt), « Standardization and Expert Knowledge », *A World of Standards*, Oxford, Oxford University Press, 2002, p. 40-49.

LEIGH STAR (Susan), LAMPLAND (Martha), *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Cornell University Press, Ithaca, 2009, 244 p.

MADAY (Charlotte), TAILLEFER (Marion), « Les métadonnées du records management du point de vue des normes ISO », *La Gazette des archives*, n°228, 2012, p. 191-197.

MEISSONNIER (Antoine), ROQUES (Rémy), « L'archiviste, les normes et le droit », *La Gazette des archives*, n°240, 2015, p. 135-151.

PETERSON (Trudy), « Access principles: developing a statement of best practice », *Comma*, vol. 2011, n°2, 2011, p. 93-100.

REED (Barbara), « The ISO standards process and implications of ISO for ICA standards », *Comma*, vol. 2011, n°2, 2011, p. 123-130.

ROUSSEL (Stéphanie), « Le champs normatif de l'archivage électronique », *La Gazette des archives*, n°228, 2012, p. 59-76.

SIBILLE DE GRIMOÜARD (Claire), « Normes internationales et influence française », *La Gazette des archives*, n°256, 2019, p. 63-70.

STAR (Susan Leigh), LAMPLAND (Martha), « Reckoning with standards », *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Cornell University Press, Ithaca, 2009, p. 3-24.

2. LA DESCRIPTION BIBLIOGRAPHIQUE

BOURNE (Ross), « Common MARC, or vivent les différences? », *Library Review*, vol. 45, n°2, p. 26.

BRYANT (Philip), « Progress in documentation », *Journal of Documentation*, vol. 36, n°2, 1980, p. 133-163.

CAMPOS (Fernanda), LOPES (Inês) et GALVÃO (Rosa), « MARC formats and their use: an overview », *Program*, vol. 29, n°4, 1995, p. 445-459.

DOBRESKI (Brian), « History of the Anglo-American tradition », *Anglo-American library cataloging*, 2020, [en ligne], disponible sur https://www.isko.org/cyclo/cataloging#3.3 (consulté le 15 avril 2021).

FATTAHI (Rahmatollah), « A comparison between the online catalogue and the card catalogue », *Library Review*, vol. 44, n°2, 1995, p. 44-58.

GORMAN (Michael), « The origins and making of the ISBD: a personal history, 1966-1978 », *Cataloging & Classification Quarterly*, vol. 52, 2014, p. 821-834.

HONORÉ (Suzanne), « Réunion du groupe de travail sur la normalisation de la notice bibliographique. Londres, 27-27 octobre 1969 », *Bulletin des bibliothèques de France*, n°3, 1970, p. 134-135.

HONORÉ (Suzanne), « Deuxième réunion du groupe de travail sur la normalisation de la description bibliographique. Paris 23-24 mai 1970 », *Bulletin des bibliothèques de France*, n° 9-10, 1970, p. 528-530.

HOPKINSON (Alan), *UKMARC AMC: Unpublished Draft Rev 4.0: UK MARC Format for Archives and Manuscripts Control (UK MARC AMC)*, Middlesex University Research Repository, 1995, 61 p., [en ligne], disponible sur https://eprints.mdx.ac.uk/5196/2/ukamc1.pdf (consulté le 5 mai 2021).

JOINT STEERING COMMITTEE FOR THE DEVELOPMENT OF RDA, « A brief history of AACR », [en ligne], disponible sur http://rda-jsc.org/archivedsite/history.html (consulté le 20 avril 2021).

JOLLEY (Leonard), « International Conference on Cataloguing Principles II. Thoughts after Paris », *Journal of Documentation*, vol. 19, n°2, 1963, p. 47-62.

KELM (Carol R.), « The historical development of the second edition of the Anglo-American Cataloging Rules », *Library Resources & Technical Services*, vol. 22, n°1, 1978, p. 23-24.

KOKABI (Mortaza), « The internationalization of MARC. Part I: the emergence and divergence of MARC », *OCLC Systems and Services*, vol. 12, n°1, 1996, p. 21-31.

LEWIS (P. R.), « The Anglo-American Cataloguing Rules (British text) : A note of welcome », *Journal of Documentation*, vol. 24, n°1, 1968, p. 55-58.

LONG (Anthony), « UK MARC and US/MARC: A brief history and comparison », *Journal of Documentation*, vol. 40, n°1, 1984, p. 1-12.

MANNING (Ralph W.), « The Anglo-American Cataloguing Rules and their future », *Library Resources & Technical Services*, vol. 44, n°3, p. 129-134.

MORTON (Katharine D.), « The MARC formats: an overview », *The American Archivist*, vol. 49, n°1, 1986, p. 26.

NATIONAL ARCHIVES OF AUSTRALIA, *Australian Recordkeeping Metadata Standard (AGRkMS)*, version 2.2, 2015, 126 p., [en ligne], disponible sur https://www.naa.gov.au/sites/default/files/2019-09/AGRkMS-Version-2.2-June-2015_tcm16-93990_1.pdf (consulté le 15 mai 2021).

PIERROT (Roger), « Etat présent de la normalisation du catalogage dans la perspective de l'automatisation », *Bulletin des bibliothèques de France*, n°3, 1970, p. 127-129.

SHINEBOURNE (J.A.), « A Critique of AACR », *Libri*, vol. 29, n°1, 1979, p. 231-259.

STOCKTING (William), « Commentary : a U.K. perspective », *The American Archivist*, vol. 74, 2011, p. 27-31.

TAIT (James A.), « The Anglo-American Cataloguing Rules, 1967 », *Library Review*, vol. 21, n°2, 1967, p. 69-74.

TAYLOR (Arlene G.), « Implementing AACR and AACR2: a personal perspective and lessons learned », *Library Resources & Technical Services*, vol. 56, n°3, p. 122-126.

WRIGHT (Wyllis E.), « The Anglo-American Cataloguing Rules: a historical perspective », *Library Resources & Technical Services*, vol. 20, n°1, 1976, p. 44.

3. L'EMPIRE BRITANNIQUE

GRELLET (Françoise), *A Cultural guide - Précis culturel des pays du monde anglophone*, Paris, Nathan, 2015, 192 p.

JUDD (Denis), « From the Great Exhibition to the Festival of Britain: the decline and fall of the British empire », *History Extra*, 2021, [en ligne], disponible sur https://www.historyextra.com/period/modern/the-decline-and-fall-of-the-british-empire/ (consulté le 24 avril 2021).

MADDISON (Angus), *The World Economy: A Millennial Perspective*, Paris, OECD Publishing, 2001, 384 p.

4. LA DESCRIPTION ARCHIVISTIQUE

ASSOCIATION DES ARCHIVISTES FRANÇAIS, *Abrégé d'archivistique, Principes et pratiques du métier d'archiviste*, Paris, Association des archivistes français, 2020, 348 p.

BEYA (Marion), « Introduction, ICA et normes », *Comma*, vol. 2011, n°2, 2011, p. 3-5.

BISSON (Marie), GOLOUBKOFF (Anne), « Les notices d'autorités en XML-TEI : un outil pour l'accroissement collaboratif de connaissances et l'indexation d'éditions de sources », *Tabularia*, 2020, [en ligne], disponible sur http://journal.openedition.org/tabularia/4176 (consulté le 1 novembre 2020).

BUNDSGAARD (Inge), SODE-PEDERSEN (Anders), « Danish strategies for new methods of digital archival description and access », *Comma*, vol. 2004, n°1, 2004, p. 67-76.

BUNN (Jennifer), « Developing descriptive standards: a renewed call to action », *Archives & Records*, vol. 34, n°2, p. 235-247.

CANADIAN WORKING GROUP ON ARCHIVAL DESCRIPTION, *Toward Descriptive Standards*, Ottawa, 1985, 192 p.

COOK (Michael), « The British move toward standards of archival description: The "MAD" standard », *The American Archivist*, vol. 53, 1990, p. 130-138.

COOK (Michael), « MAD2: Reassessing the Experience », *Archivaria*, vol. 35, 1993, p. 15-23.

DA FONSECA (Vitor), « The ICA description standards: the history of their creation and efforts to disseminate them », *Comma*, vol. 2011, n°2, 2011, p. 49-58.

DAVIS (Susan E.), « Descriptive standards and the archival profession », *Cataloging & Classification Quaterly*, vol. 35, n°3-4, 2003, p. 298.

DOOM (Vincent), « Description et analyse archivistiques ou la nécessité d'une normalisation », *La Gazette des archives*, n°182-183, 1998, p. 232-246.

DRYDEN (Jean), « Developing international standards: Lessons from the CUSTARD project », *Comma*, vol. 2005, n°2, 2005, p. 1-6.

DRYDEN (Jean), « Audience Comments and Questions », *The American Archivist*, vol. 74, 2011, p. 34.

DUFF (Wendy), HAWORTH (Kent), « The Reclamation of archival description: the Canadian perspective », *Archivaria*, vol. 31, 1991, p. 26-35.

DURANCE (Cynthia J.), « Authority Control: Beyond a bowl of alphabet soup », *Archivaria*, vol. 35, 1993, p. 38-46.

GRAILLES (Bénédicte), « Archivistiques canadienne et française : échanges, influences ou transferts ? », *Le théâtre du patrimoine (France-Canada)*, Paris, CTHS, 2008, p. 47-66.

GRAILLES (Bénédicte), « Normalisations, standardisations, modélisations : acculturations professionnelles ou uniformisations identitaires ? », *La Gazette des archives*, n° 259, 2020, p. 51-67.

HAIDER (Salman), « Anglo-American Cataloguing Rules », *Encyclopedia Britannica*, 2017, [en ligne], disponible sur https://www.britannica.com/topic/library/Catalog-standardization#ref336275 (consulté le 3 mars 2021).

HAWORTH (Kent), « The Development of descriptive standards in Canada: a progress report », *Archivaria,* vol. 34, 1992, p. 75-90.

HAWORTH (Kent), « The Voyage of RAD: From the Old World to the New », *Archivaria,* vol. 35, 1993, p. 55-63.

HENSEN (Steven), « The Use of standards in the application of the AMC format », *The American Archivist*, vol. 49, n°1, 1986, p. 31-40.

HENSEN (Steven), « RAD, MAD, and APPM: The Search for Anglo-American Standards for Archival Description », *Archives and Museum Informatics,* vol. 5, n°2, 1991, p. 2-5.

HENSEN (Steven), « The First shall be first: APPM and its impact on American archival description », *Archivaria,* vol. 35, 1993, p. 64-70.

HENSEN (Steven), « Concluding Thoughts from the Chair », *The American Archivist*, vol. 74, 2011, p. 32.

HENSEN (Steven), « Audience Comments and Questions », The American Archivist, vol. 74, 2011, p. 34-35.

INTERNATIONAL COUNCIL ON ARCHIVES, « Statement of Principles regarding archival description », *Archivaria*, vol. 34, 1992, p. 8-15.

KIESLING (Kris), « EAD as an archival descriptive standard », *The American Archivist*, vol. 60, 1997, p. 344-354.

LYTLE (Richard H.), « An Analysis of the Work of the National Information Systems Task Force », *The American Archivist*, vol. 47, n°4, 1984, p. 360.

MACNEIL (Heather), « HENSEN (comp.), Archives, Personal Papers, and Manuscripts: Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries », *Archivaria*, vol. 30, 1990, p. 153.

MOTTE (Alice), « La normalisation de la description archivistique : enjeux et actualités », *La Gazette des archives*, n°238, 2015, p. 121-128.

NOUGARET (Christine), « Vers une normalisation internationale de la description des archives. La norme ISAD (G) du Conseil international des archives », *La Gazette des archives*, n°169, 1995, p. 274-292.

NOUGARET (Christine), « D'ISAD/G à ISAAR : un nouvel outil du Conseil international des archives pour la description archivistique », *La Gazette des archives,* n°176, 1997, p. 102-122.

NOUGARET (Christine), GALLAND (Bruno), *Les instruments de recherche dans les archives*, Paris, La Documentation française : Direction des archives de France, 1999, 259 p.

OTTOSSON (Per-Gunnar), « Implementing ISAAR (CPF): encoded archival description (EAC) », *Comma*, vol. 2005, n°2, 2005, p. 1-4.

PLANNING COMMITTEE ON DESCRIPTIVE STANDARDS, « Preface to the 1990 edition », *Rules for Archival Description*, Ottawa, 2008, p. xv-xxi.

ROBINSON (David), « Theory, pratice and a Manual of Archival Description », *Journal of the Society of Archivists*, vol. 12, n°1, 1991, p. 1-6.

ROE (Kathleen D.), « Out of the primordial ooze: the genesis of U.S. archival descriptive standards », *The American Archivist*, vol. 74, 2011, p. 3-11.

RUSH (Michael), « The archival network: you don't get to describe records without making a few standards », *The American Archivist*, vol. 74, 2011, p. 20-27.

SAHLI (Nancy A.), « Interpretation and application of the AMC format », *The American Archivist*, vol. 49, n°1, 1986, p. 9-20.

SALAÜN (Jean-Michel) *Vu, lu, su. Les architectes de l'information face à l'oligopole du web*, Paris, La Découverte, 2012, 73 p., [en ligne], disponible sur https://halshs.archives-ouvertes.fr/halshs-01979526/document (consulté le 7 avril 2021).

SHEPHERD (Elizabeth), WEST (Victoria), « Are ISO 15489-1:2001 and ISAD(G) compatible? », *Records Management Journal*, vol. 13 (1), 2003, p. 9-23, vol.13 (2), 2003, p. 62-69.

SIBILLE DE GRIMOÜARD (Claire), « L'EAD et l'EAC, deux DTD pour la description archivistique », *Archives de France*, 2004, 33 p., [en ligne], disponible sur http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2004/pin20040504ead_eac.pdf (consulté le 22 février 2021).

SIBILLE DE GRIMOÜARD (Claire), « Les instruments de recherche électroniques », *Comma*, vol. 2006, n°1-2, 2006, p. 1-11.

SIBILLE DE GRIMOÜARD (Claire), GNESSOUGOU BAROAN-DIOUMENCY (Padré Lydie), « Une nouvelle norme internationale pour décrire les fonctions des producteurs d'archives : ICA-ISDF », *Comma*, vol. 2008, n°2, 2008, p. 113-123.

SIBILLE DE GRIMOÜARD (Claire), « ICA-Atom, un outil pour appliquer les normes de description », *Comma*, vol. 2011, n°2, 2011, p. 59-72.

SIBILLE DE GRIMOÜARD (Claire), « Les normes internationales de description archivistique : origines, développements, perspectives », *La Gazette des archives*, n°228, 2012, p. 77-90.

SIBILLE DE GRIMOÜARD (Claire), « Implementation of EAC-CPF (Encoded Archival Context-Corporate bodies, Persons, Families) in France: towards the development of national authority files », *Comma*, vol. 2012, n°2, 2012, p. 105-114.

SIBILLE DE GRIMOÜARD (Claire), « D'hier à aujourd'hui : les évolutions de la description archivistique », *La Gazette des archives*, n°247, 2017, p. 117-123.

STIBBE (Hugo), « Implementing the Concept of Fonds: Primary Access Point, Multilevel Description and Authority Control », *Archivaria*, vol. 34, 1992, p. 109-137.

STOCKTING (Bill), « 'A flexible friend': the role of encoded archival description (EAD) in the promotion of access to archives in the United Kingdom », *Comma*, vol. 2005, n°2, 2005, p. 1-11.

WAJS (Hubert), « The polish road to EAD », *Comma*, vol. 2005, n°2, 2005, p. 1-4.

WALSH (G. Mark), « The disconnect between archival descriptive technique and records management taxonomies », *Comma*, vol. 2017, n°2, 2019, p. 35-42.

WEBER (Lisa B.), « Archival description standards: concepts, principles, and methodologies », *The American Archivist*, vol. 52, 1989, p. 511.

YEO (Geoffrey), « Archival description in the era of digital abundance », *Comma*, vol. 2013, n°2, 2013, p. 15-26.

5. L'ENVIRONNEMENT NUMERIQUE DES ARCHIVES

BACHIMONT (Bruno), « L'archive et la massification des données : une nouvelle raison numérique », *La Gazette des archives*, n°245, 2017, p. 27-43.

BÉNEL (Aurélien), « Archives numériques et construction du sens ou "Comment échapper au Web sémantique ?" », *La Gazette des archives*, n°245, 2017, p. 173-187.

CARAVACA (María), VIOLA (Anne-Marie), « The costs of a free, open-source software implementation: building ICCROM, a web-based archival catalogue using ICA-AtoM », *Comma*, vol. 2011, n°2, 2011, p. 73-86.

DELTOUR (Jean-Pierre), « ArchivisTIC, Métadonnées / Mets ta donnée ! / Mais… t'as donné (fin) », *Archivistes !*, n°115, 2015, p. 10-11.

GUYON (Céline), LONGIN (Cyril), ZELLER (Jean-Daniel), « Open data et archivage électronique, quelles convergences ? », *La Gazette des archives*, n°240, 2015, p. 385-396.

LE DEUFF (Olivier), *La documentation dans le numérique*, Villeurbanne, Presses de l'Enssib, 2014, 223 p.

LEE (Christopher), « Archival application of digital forensics methods for authenticity, description and access provision », *Comma*, vol. 2012, n°2, 2012, p. 133-140.

LEMIEUX (Victoria), « Using information visualization and visual analytics to achieve a more sustainable future for archives. A survey and critical analysis of some developments », *Comma*, vol. 2012, n°2, 2012, p. 55-70.

LEVY (Violette), coord. « Les Archives et le web sémantique », [dossier thématique] *Archivistes !*, n°136, 2021, p. 19-33.

LUQUET (Jean), « Diffuser et préserver : pourquoi le numérique révolutionne les archives », *La Gazette des archives*, n°207, 2007, p. 199-207.

MAUREL (Dominique), CHEBBI (Aïda), « Towards negotiated governance of digital records: individual and collective information practices in organizations », *Comma*, vol. 2013, n°1, 2013, p. 15-28.

MEDDEB (Sami), « L'archivage à l'ère du numérique : défis et solutions », *Comma*, vol. 2016, n°1-2, 2016, p. 205-208.

MUÑOZ (Francisco Javier Crespo), « Las nuevas realidades de la reprografía en archivos : de los metadatos a la gestión de procesos », *Comma*, vol. 2017, n°2, 2019, p. 53-62.

POMART (Julien), « La visualisation des données archivistiques : les premiers pas, et ensuite ? », *La Gazette des archives*, n°248, 2017, p. 97-112.

RENÉ-BAZIN (Paule), « Vers une informatique archivistique », *La Gazette des archives*, n°129, 1985, p. 114-124.

SAJUS (Bertrand), LEROI (Marie-Véronique), « Le développement du web des données culturelles : les enjeux pour le ministère de la Culture et de la Communication », *I2D - Information, données & documents*, vol. 53, 2016, p. 46-47.

SALAÜN (Jean-Michel), « La redocumentarisation, un défi pour les sciences de l'information », *Études de communication*, vol. 30, 2007, p. 13-23.

SALAÜN (Jean-Michel), *Vu, lu, su. Les architectes de l'information face au monopole du web*, 2012, [en ligne], disponible sur https://halshs.archives-ouvertes.fr/halshs-01979526/document (consulté le 7 avril 2021).

SIBILLE DE GRIMOÜARD (Claire), NICHELE (Baptiste), « Le Standard d'échange de données pour l'archivage (SEDA), un outil structurant pour l'archivage », *La Gazette des archives*, n°240, 2015, p. 153-164.

WALKER (Helen), « Guidance for implementing globally harmonised software specifications for records », *Comma*, vol. 2011, n°2, 2011, p. 87-91.

WESTEEL (Isabelle), « Indexer, structurer, échanger : métadonnées et interopérabilité », *Numériser et mettre en ligne*, Paris, Presses de l'enssib, 2010, p. 100-123.

WETERINGS (Jorien), « Changing theory into practice: playing the Metadata Game », *Comma*, vol. 2012, n°2, 2012, p. 85-96.

WHITELAW (Mitchell), « Towards generous interfaces for archival collections », *Comma*, vol. 2012, n°2, 2012, p. 123-132.

6. L'UTILISATEUR FINAL DES ARCHIVES

BROCHIER (André), « Aider le chercheur lors de la mise en ligne d'instruments de recherche sur Internet : l'expérience des Archives nationales d'outre-mer », *Comma*, vol. 2008, n°2, 2008, p. 31-38.

CLAVERT (Frédéric), DANIEL (Johanna), FLECKINGER (Hélène), GRANDJEAN (Martin), IDMHAND (Fatiha), « Histoire et humanités numériques : nouveaux terrains de dialogue entre les archives et la recherche », *La Gazette des archives*, n°245, 2017, p. 121-134.

DUFF (Wendy), « Beyond name collecting: PhD Students' information seeking behaviour in physical and virtual archives », *Comma*, vol. 2005, n°3, 2005, p. 1-12.

POUBLANC (Sébastien), « Les jeunes historiens rêvent-ils d'archives numériques ? », *La Gazette des archives*, n°253, 2019, p. 37-51.

WILLIAMS (Caroline), « Allies or aliens? What is the relationship between academic and practice-based research? », *Comma*, vol. 2008, n°2, 2008, p. 79-88.

ZOLOTAREVSKY (Masha), « The information retrieval needs of archival users: a case of study from the Jabotinsky Institute, Israel », *Comma*, vol. 2008, n°2, 2008, p. 47-53.

7. LA TERMINOLOGIE

AMANY (Mohamed), « International trends in standardizing archival terminology: the multilingual archival terminology (MAT) as a model », *Comma*, vol. 2017, n°2, 2019, p. 135-161.

CARRERA (Carmen), MUNDET (José), CORRAL (Isabel), MEDINA (Santiago), « El macrodiccionario de terminología : una herramienta para la normalización en el ámbito de la lengua española », *Comma*, vol. 2008, n°2, 2008, p. 101-112.

LEROI (Marie-Véronique), COUDYZER (Eva), *Terminologies and semantic enrichment*, AthenaPlus, actes de la conférence TEI organisée à Rome en 2013, p. 4, [en ligne], disponible sur https://athenaplus.eu/getFile.php?id=260 (consulté le 18 octobre 2020).

MCCULLOCH (Emma), « Multiple terminologies: an obstacle to information retrieval », *Library Review*, vol. 53, 2004, p. 297-300.

8. LE NUMERIQUE

FRANCART (Thomas), « SHACL (RDF Shapes Constraint Language) : enfin la possibilité de valider des données en RDF », *Sparna Blog*, 2017, [en ligne], disponible sur http://blog.sparna.fr/2017/01/02/shacl-rdf-shapes-constraint-language-enfin-la-possibilite-de-valider-des-donnees-rdf/ (consulté le 28 février 2021).

GANDON (Fabien), *Données liées et Web sémantique : quand le lien fait sens*, actes de la présentation pour les rencontres Inria Industrie « Technologies du Web », 2014, 61 p., [en ligne], disponible sur https://fr.slideshare.net/fabien_gandon/web-semlinkeddata-fabiengandon (consulté le 14 mai 2021).

WORLD WIDE WEB CONSORTIUM, *W3C Opens Data on the Web with SPARQL, Powerful Technology for Querying Distributed and Diverse Data*, W3C, 2008, [en ligne], disponible sur https://www.w3.org/2007/12/sparql-pressrelease.html.en (consulté le 20 mai 2021).

9. LA LINGUISTIQUE

DELBECQUE (Nicole), *Linguistique cognitive, comprendre comment fonctionne le langage*, Bruxelles, De Boeck, 2006, 404 p.

DE SAUSSURE (Ferdinand), *Cours de linguistique générale*, Paris, Payot & Rivages, 2016, 409 p.

LOCKE (John), *An Essay Concerning Human Understanding*, London, Penguin Classics, 1998, 980 p.

10. L'APPROCHE ENCYCLOPEDIQUE

ALEMBERT (Jean Le Rond D'), MALHERBE (Michel), *Discours préliminaire de l'Encyclopédie*, Paris, J. Vrin, 2000, 209 p.

GROULT (Martine), *D'Alembert et la mécanique de la vérité dans l'Encyclopédie*, Paris, Champion, 1999, 504 p.

GROULT (Martine), *L'Encyclopédie ou la création des disciplines*, Paris, CNRS, 2003, 344 p.

REY (Alain), *Miroirs du monde*, Paris, Fayard, 2007, 262 p.

11. LES NEUROSCIENCES ET LES SCIENCES COGNITIVES

BOLOIX (Emmanuelle), « Attention et perception de scènes visuelles », *L'année psychologique*, vol. 107, n°1, 2007, p. 113-151.

BUSER (Pierre), LESTIENNE (Rémy), *Cerveau, information, connaissance*, Paris, CNRS, 2001, 227 p.

DAVID (Alfred), *La vie secrète du cerveau : découvrez ses étonnants pouvoirs !*, Malakoff, Dunod, 2019, 192 p.

DENIS (Michel), *Image et cognition*, Paris, Presses universitaires de France, 1989, 284 p.

DENIS (Michel) (dir.), *La psychologie cognitive*, Paris, Maison des Sciences de l'Homme, 2012, 320 p.

DORTIER (Jean-François) (dir.), *Le cerveau et la pensée, le nouvel âge des sciences cognitives*, Auxerre, Sciences humaines, 2013, 479 p.

DUBUC (Bruno), *Les différents cortex visuels*, Le cerveau à tous les niveaux !, [en ligne], disponible sur https://lecerveau.mcgill.ca/flash/a/a_02/a_02_cr/a_02_cr_vis/a_02_cr_vis.html (consulté le 23 mai 2021).

DUMAS (Thibaud), *Le cerveau*, Paris, Mango, 2018, 128 p.

EHRLICH (Marie-France) (dir.), *Les modèles mentaux, approche cognitive des représentations*, Paris, Sciences cognitives, 1988, 183 p.

FORTIN (Claudette), ROUSSEAU (Robert), *Psychologie cognitive : Une approche de traitement de l'information*, Québec, Presses de l'Université du Québec, 2012, 388 p.

FRITH (Chris), *Comment le cerveau crée notre univers mental*, Paris, O. Jacob, 2010, 306 p.

HOUDE (Olivier), *Le raisonnement*, Paris, Presses universitaires de France, 2014, 127 p.

HUBEL (David H.), *L'oeil, le cerveau et la vision, les étapes cérébrales du traitement visuel*, Paris, Pour la science : diffusion Belin, 1994, 240 p.

INSTITUT DU CERVEAU, *Comprendre le cerveau, son développement, son fonctionnement, sa plasticité et ses pathologies*, 2020, [en ligne], disponible sur https://institutducerveau-icm.org/fr/actualite/comprendre-le-cerveau-et-son-fonctionnement/ (consulté le 23 mai 2021).

LEMAIRE (Patrick), *Abrégé de psychologie cognitive*, Bruxelles, De Boeck, 2006, 137 p.

LEMAIRE (Patrick), DIDIERJEAN (André), *Introduction à la psychologie cognitive*, Bruxelles, De Boeck, 2018, 416 p.

MATLIN (Margaret W.), *La cognition, une introduction à la psychologie cognitive*, Bruxelles, De Boeck Supérieur, 2001, 786 p.

MEUNIER (Jean-Marc), *Mémoires, représentations et traitements*, Malakoff, Dunod, 2019, 213 p.

NACCACHE (Lionel), NACCACHE (Karine), *Parlez-vous cerveau ?*, Paris, Odile Jacob, 2018, 214 p.

ORBAN (Guy), *La vision, mission du cerveau : les trois révolutions des neurosciences*, Paris, Fayard, 2007, 88 p.

PIAGET (Jean), *La psychologie de l'intelligence*, Paris, NOM, 2007, 238 p.

PUDELKO (Béatrice), BASQUE (Josianne), « Logiciels de construction de cartes de connaissances : des outils pour apprendre », 2005, [en ligne], disponible sur https://www.vteducation.org/sites/default/files/images/Yves%20Munn/logiciels-de-construction-de-cartes-de-connaissances.pdf (consulté le 7 avril 2021).

REY (Arnaud), *Psychologie cognitive expérimentale*, Paris, Presses universitaires de France, 2012, 247 p.

RICHARD (Daniel) (dir.), *Mémo visuel de neurosciences : du neurone aux sciences cognitives*, Malakoff, Dunod, 2021, 172 p.

RICHARD (Jean-François), *Les activités mentales, de l'interprétation de l'information à l'action*, Paris, A. Colin, 2004, 428 p.

ROSSI (Sandrine), VAN DER HENST (Jean-Baptiste), *Psychologies du raisonnement*, Bruxelles, De Boeck, 2007, 328 p.

12. LA CARTE HEURISTIQUE

BUZAN (Tony), *Développez votre intelligence avec le mind mapping*, Paris, Alisio, 2018, 223 p.

FISHER (Kathleen M.), « Semantic networking: the new kid on the block », *Journal of Research in Science Teaching*, vol. 27, n°10, 1990, p. 1001-1018.

MCALEESE (Ray), « The Knowledge Arena as an Extension to the Concept Map: Reflection in Action », *Interactive Learning Environments*, vol. 6, n°3, 1998, p. 251-272

MONGIN (Pierre), DE BROECK (Fabienne), *Enseigner autrement avec le mind mapping, cartes mentales et conceptuelles*, Malakoff, Dunod, 2016, 191 p.

13. LA SEMIOLOGIE GRAPHIQUE

BERTIN (Jacques), *Sémiologie graphique : les diagrammes, les réseaux, les cartes*, Paris, École des Hautes Études en sciences sociales, 2005, 452 p.

PALSKY (Gilles), ROBIC (Marie-Claire), « Aux sources de la sémiologie graphique », *Cybergeo : European Journal of Geography*, 1997, [en ligne], disponible sur https://journals.openedition.org/cybergeo/554 (consulté le 7 avril 2021).

PALSKY (Gilles), « Eléments pour une histoire de la sémiologie graphique avant Jacques Bertin », *Bulletin de l'Association des géographes français*, vol. 80 (2), 2003, p. 183-194.

PALSKY (Gilles), « La Sémiologie graphique de Jacques Bertin a cinquante ans ! » *Visionscarto*, 2017, [en ligne], disponible sur https://visionscarto.net/la-semiologie-graphique-a-50-ans (consulté le 7 avril 2021).

PLANTIN (Jean-Christophe), « L'avènement de la carte comme médiation », *Questions de co mmunication*, 2014, [en ligne], disponible sur https://journals.openedition.org/questionsdeco mmunication/9050 (consulté le 7 avril 2021).

14. LA *DATA VISUALISATION*

BEGUIN-VERBRUGGE (Annette), *Images en texte, images du texte : dispositifs graphiques et communication écrite*, Villeneuve d'Ascq, Presses universitaires du Septentrion, 2006, 313 p.

FEKETE (Jean-Daniel), BOY (Jeremy), « Recherche en visualisation d'information ou dataviz : pourquoi et comment ? », *I2D - Information, données & documents*, vol. 52, n°2, 2015, p. 32-33.

FEW (Stephen), « Data visualization for human perception », *The Encyclopedia of Human-Computer interaction*, Interaction Design Foundation, 2014, [en ligne], disponible sur https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception (consulté le 23 mai 2021).

VINOT (Jean-Luc), CONVERSY (Stéphane), *La continuité, une idée-force de design pour l'architecture graphique des systèmes interactifs*, actes de la 27ème conférence francophone sur l'Interaction Homme-Machine organisée à Toulouse en 2015, 10 p.

ETAT DES SOURCES

1. Sources écrites

1.1. Échelle internationales

CONSEIL INTERNATIONAL DES ARCHIVES, *International Standard Archival Description – General*, 1ère édition, 1994.

CONSEIL INTERNATIONAL DES ARCHIVES, *International Standard Archival Description – General*, 2ème édition, 2000.

CONSEIL INTERNATIONAL DES ARCHIVES, *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, 1ère édition, 1995.

CONSEIL INTERNATIONAL DES ARCHIVES, *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, 1ème édition, 2004.

CONSEIL INTERNATIONAL DES ARCHIVES, *International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, 1ère édition, 1995.

CONSEIL INTERNATIONAL DES ARCHIVES, *International Standard for Describing Functions*, 2008.

CONSEIL INTERNATIONAL DES ARCHIVES, *International Standard for Describing Institutions with Archival Holdings*, 2008.

CONSEIL INTERNATIONAL DES ARCHIVES, *Records in Contexts (RiC), An archival description draft standard*, actes du 16ème Congrès du Conseil international des Archives à Séoul en 2016, [en ligne], disponible sur https://www.ica.org/sites/default/files/session-7.8-ica-egad-ric-congress2016.pdf (consulté le 18 janvier 2021).

CONSEIL INTERNATIONAL DES ARCHIVES, *Records in Contexts, A Conceptual Model for Archival Description, Consultation Draft v0.2* (preview), 2019, [en ligne], disponible sur https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf (consulté le 18 janvier 2021).

CONSEIL INTERNATIONAL DES ARCHIVES, *International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2*, 2021, [en ligne], disponible sur https://www.ica.org/standards/RiC/RiC-O_v0-2.html (consulté le 3 mars 2021).

GUEGUEN (Gretchen), MARQUES DA FONSECA (Vitor Manoel), PITTI (Daniel), SIBILLE DE GRIMOÜARD (Claire), « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *The American Archivist*, vol. 76, n°2, 2013, p. 566-583.

INTERNATIONAL CONFERENCE ON CATALOGUING PRINCIPLES, *Statement of Principles*, Paris, 1961.

INTERPARES TRUST, « Comments on "Records in Context" », *InterPARES Trust blog*, 2016, [en ligne], disponible sur https://interparestrustblog.wordpress.com/2016/12/11/interpares-trust-responds-to-egad-ric/ (consulté le 5 mars 2021).

PITTI (Daniel), STOCKTING (Bill), CLAVAUD (Florence), « An introduction to "Records in Contexts": an archival description draft standard », *Comma*, vol. 2016, n°1-2, 2018, p. 173-188.

1.2. Australie

NATIONAL ARCHIVES OF AUSTRALIA, *Australian Recordkeeping Metadata Standard (AGRkMS)*, version 2.2, 2015, [en ligne], disponible sur https://www.naa.gov.au/sites/default/files/2019-09/AGRkMS-Version-2.2-June-2015_tcm16-93990_1.pdf (consulté le 15 mai 2021).

1.3. Canada

COMITE DE PLANIFICATION DES NORMES DE DESCRIPTION, *Les règles pour la description des documents d'archives*, Ottawa, 1990.

1.4. États-Unis

BIBLIOTHEQUE DES CONGRES, *Archives, Personal Papers and Manuscripts,* 1ère édition, 1983.

GORMAN (Michael), WINKLER (Paul W.), *Anglo-American Cataloguing Rules*, 2nde édition, 1978.

HENSEN (Steven L.), *Archives, Personal Papers and Manuscripts*, 2nde édition, 1989.

SOCIETY OF AMERICAN ARCHIVISTS, *Describing Archives: A Content Standard*, 1ère édition, 2004.

SOCIETY OF AMERICAN ARCHIVISTS, *Describing Archives: A Content Standard*, 4ème édition, 2019.

SOCIETY OF AMERICAN ARCHIVISTS, *Encoded Archival Description*, 1ère édition, 1998.

SOCIETY OF AMERICAN ARCHIVISTS, *Encoded Archival Description*, 2nde édition, 2002.

SOCIETY OF AMERICAN ARCHIVISTS, *Encoded Archival Description*, 3ème édition, 2010.

SOCIETY OF AMERICAN ARCHIVISTS, *Encoded Archival Context*, 1ère édition, 2004.

SOCIETY OF AMERICAN ARCHIVISTS, *Encoded Archival Context*, 2ème édition, 2010.

SUMNER SPALDING (Charles), *Anglo-American Cataloguing Rules*, 1ère édition, 1967.

WORKING GROUP ON STANDARDS FOR ARCHIVAL DESCRIPTION, « Archival Description Standards: Establishing a Process for Their Development and Implementation: Report of the Working Group on Standards for Archival Description. », *The American Archivist*, vol. 52, n°4, 1989, p. 469.

1.5. France

ANGJELI (Anila), CLAVAUD (Florence), ROUSSEL (Stéphanie), « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *La Gazette des archives*, n°245, 2017, p. 157-171.

CLAVAUD (Florence), CHARBONNIER (Pauline), *Records in Contexts aux Archives nationales : enjeux et premières réalisations*, actes de la journée d'étude organisée le 28 janvier 2020 par les Archives nationales à Paris, [en ligne], disponible sur https://f-origin.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_3_RiCauxAN_EnjeuxPremieresRealisations.pdf (consulté le 18 janvier 2021).

ARCHIVES NATIONALES, *Le portail FranceArchives : de l'agrégation d'instruments de recherche à la constitution d'un réservoir de données*, actes de la journée d'étude organisée le 28 janvier 2020 par les Archives nationales à Paris, [en ligne], disponible sur https://f-origin.hypotheses.org/wpcontent/blogs.dir/2167/files/2020/02/20200128_6_PortailFranceArchives_ReservoirDonnees.pdf (consulté le 18 janvier 2021).

CLAVAUD (Florence), « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *Journal of Archival Organization*, vol. 12, n°1-2, 2015, p. 118-142.

CLAVAUD (Florence), *Les référentiels aux Archives nationales*, actes de la rencontre annuelle des correspondants archives des opérateurs et grands corps de l'Etat du 12 octobre 2016, [en ligne], disponible sur https://www.francearchives.fr/file/20b2e2aa9a1f62b2b843239beff36a29468db1f3/static_9370.pdf (consulté le 18 janvier 2021).

CLAVAUD (Florence), « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *La Gazette des archives*, n°254, 2019, p. 59-88.

CLAVAUD (Florence), *Le nouveau standard international de description des archives, Records in Contexts*, actes de la journée d'étude organisée le 28 janvier 2020 par les Archives nationales à Paris, [en ligne], disponible sur https://f.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_2_RecordsInContexts_englishVersionAdded1003.pdf (consulté le 18 janvier 2021).

FRANCART (Thomas), CHARBONNIER (Pauline), *RIC-O converter, un logiciel libre de conversion de métadonnées archivistiques (en EAD et EAC-CPF) en jeux de données conformes à RiC-O*, actes de la journée d'étude organisée le 28 janvier 2020 par les Archives nationales à Paris, [en ligne], disponible sur https://f.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_4_RiCOConverter.pdf (consulté le 18 janvier 2021).

MINISTERE DE LA CULTURE, *Feuille de route stratégique, Métadonnées culturelles et transition Web 3.0*, 2014, 38 p., [en ligne], disponible sur https://www.culture.gouv.fr/var/culture/storage/pub/feuille_de_route__metadonnees_culturelles_et_transition_web_3_0_janvier_2014/files/docs/all.pdf (consulté le 18 mai 2021).

1.6. Roumanie

POPOVICI (Bogdan Florin), « Records in Contexts : vers un nouveau niveau dans la description archivistique ? », *Archives*, vol. 48, 2019, p. 7-39.

POPOVICI (Bogdan Florin), « A broader perspective on records as seen by Records-in-Contexts », *Comma*, vol. 2016, n°1-2, 2018, p. 189-198.

1.7. Royaume-Uni

COOK (Michael), *Manual of Archival Description Archives*, 1ère édition, 1984.

COOK (Michael), PROCTER (Margaret), *Manual of Archival Description Archives*, 2nde édition, 1989.

COOK (Michael), PROCTER (Margaret), *Manual of Archival Description Archives*, 3ème édition, 2000.

GORMAN (Michael), WINKLER (Paul W.), *Anglo-American Cataloguing Rules, Second Edition, 1988 Revision*, Chicago, 1988, 677 p.

2. Sources par voie d'entretiens

La transcription de l'ensemble de ces entretiens est située en annexe 1.

L'entretien avec Madame Claire Sibille de Grimoüard a été réalisé le 5 mars 2021. Il dure 56 minutes et 20 secondes. L'entretien s'est déroulé en appel vidéo. Madame Sibille de Grimoüard est experte en normalisation de la description archivistique et a participé au début des travaux portant sur *Records in Contexts*.

L'entretien avec Monsieur Thomas Francart a été réalisé le 18 mars 2021. Il dure 1 heure, 9 minutes et 43 secondes. L'entretien s'est déroulé au sein de la société Sparna, à Tours. Monsieur Francart est expert en modélisation des connaissances et en gestion des ontologies. Il a développé le logiciel libre RiC-O Converter avec les Archives nationales.

L'entretien avec Monsieur Jean-Séverin Lair a été réalisé le 25 mars 2021. Il dure 56 minutes et 21 secondes. L'entretien s'est déroulé en appel vidéo. Monsieur Lair exprime un point de vue défavorable sur RiC-O. Un point de vue qui s'analyse au prisme de son expertise en informatique et de son expérience de directeur du programme Vitam, premier programme à reposer sur le travail collaboratif d'archivistes et d'informaticiens.

L'entretien avec Madame Florence Clavaud a été réalisé le 26 mars 2021. Il dure 1 heure, 8 minutes et 41 secondes. L'entretien s'est déroulé en appel vidéo. Madame Clavaud est membre exécutif de l'EGAD, responsable de l'équipe de développement de RiC-O et responsable de la mission référentiels et de projets numériques aux Archives nationales.

L'entretien avec Monsieur Edouard Vasseur a été réalisé le 26 mars 2021. Il dure 54 minutes et 59 secondes. L'entretien s'est déroulé en appel vidéo. Monsieur Vasseur manifeste un avis dubitatif au sujet de RiC-O. Un avis qui s'explore au prisme de son expertise en archivistique et de son expérience de responsable fonctionnel du programme Vitam, programme durant lequel il a pu travailler main dans la main avec Frédéric Brégier, le directeur technique du programme.

L'entretien avec Monsieur David Chavalarias a été réalisé le 29 avril 2021. Il dure 48 minutes et 55 secondes. L'entretien s'est déroulé par téléphone. Les recherches de Monsieur Chavalarias se situent à l'intersection des graphes de connaissance et des sciences cognitives.

3. Sources par voie de questionnaires

Les réponses à ces questionnaires sont situées en annexe 2.

Un questionnaire a été soumis à Monsieur Daniel Pitti le 22 mars 2021. Monsieur Pitti est président de l'EGAD.

Un questionnaire a été soumis à Madame Florence Clavaud le 29 mars 2021.

RECORDS IN CONTEXTS : CAP SUR LE WEB 3.0

« Dans cette vaste entreprise de description de toute information existante sur fond d'automatisation, les normes de description et de communication des données sont absolument centrales pour uniformiser les métadonnées et s'assurer que tout le monde se comprenne : (as-tu nor)mé ta donnée ? »

Jean-Pierre Deltour[285]

Ainsi que nous l'avons vu, il existe un lien étroit entre les technologies de l'information et l'évolution du cadre normatif de la description archivisique ; les premières semblent influer sur le dernier. Aussi, l'arrivée du web sémantique ou « web de (méta)données »[286] dans les années 2010 rebat les cartes du jeu. En effet, « Polymorphiques, complexes et omniprésentes, [les métadonnées] sont un enjeu d'avenir considérable tant pour le citoyen que pour les services d'archives »[287] ; raison pour laquelle leur normalisation est cruciale pour en garantir la qualité.

En parallèle, l'ICA s'est engagé dans la révision de ses normes de description archivistique ; un processus débuté en 2008 après le Congrès de Kuala Lumpur avec le remaniement de la PSBP en *Committee on Best Practices and Standards* (CBPS). Le travail de ce comité est conséquent, puisqu'il aboutit en 2012 à deux documents visant à repenser les quatre normes qui encadrent la pratique de la description archivistique : un document — *Relationship in Archival Descriptive Systems* — qui expose des cas de chevauchement ainsi que des incohérences au sein du cadre normatif, suivi d'un rapport — *Progress Report for Revising and Harmonizing ICA Descriptive Standards* — qui détaille les solutions envisageables[288] ; ce à quoi s'ensuit la naissance de l'EGAD — successeur partiel du CBPS[289] — qui enclenche la procédure de révision du paysage normatif de la description archivistique par le développement de *Records in Contexts* (RiC) ; un modèle conceptuel encore en gestation à ce jour.

L'étude de cette norme présente de nombreux intérêts. D'une part, elle permet de mieux appréhender le développement d'un processus normatif à l'échelle internationale ; le premier à envisager une norme pour la description archivistique et son implémentation technique au sein d'un même groupe de travail. D'autre part, étudier le RiC peut permettre de saisir pleinement

---

[285] Jean-Pierre Deltour, « ArchivisTIC, Métadonnées / Mets ta donnée ! / Mais… t'as donné (fin) », *Archivistes !*, n°115, 2015, p. 11.
[286] *Ibid.*
[287] *Ibid.*
[288] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 571.
[289] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°8.

l'enjeu d'inscrire la pratique de la description au sein du web 3.0 ; l'objectif étant de comprendre comment un modèle conceptuel en la matière peut davantage contribuer à positionner les archivistes comme des membres actifs de la communauté des sciences de l'information.

Ainsi, nous nous efforcerons tout d'abord d'exposer l'organisation et la méthodologie de l'EGAD et toutes les problématiques qu'un processus normatif à l'échelle internationale peut soulever : selon quels critères choisir les membres du groupe de travail ? comment inclure le plus de pays possibles au sein de ce processus ? comment parvenir à un consensus par-delà les traditions archivistiques représentées et la barrière de la langue ? Puis nous nous intéresserons à RiC elle-même et à la manière dont elle a été pensée et construite : comment s'opère la refonte des quatre normes existantes ? qu'est-ce que RiC apporte de plus au paysage normatif de la description archivistique ? la norme envisage-t-elle une plus grande entente avec le cadre normatif de l'ISO sur les métadonnées archivistiques ?

L'exemple des Archives nationales (AN) nous permettra d'appréhender au mieux les changements engendrés par la mise en œuvre de cette nouvelle norme au sein d'un service d'archives. Il s'agira d'analyser l'application de RiC à travers le *Pilote d'Interopérabilité des Autorités Archivistiques Françaises* (PIAAF), de procéder à l'examen du logiciel libre RiC-O Converter — un logiciel pour convertir les métadonnées archivistiques EAD et EAC en jeux de données conformes à *Records in Contexts Ontology* (RiC-O) —, ainsi que de se pencher sur les référentiels des AN, et sur les changements effectués et à effectuer pour les enrichir, et ainsi faciliter l'accès aux métadonnées et favoriser l'interopérabilité. Le choix de ce service d'archives n'est pas le fruit du hasard puisqu'il comprend l'un des membres exécutifs de l'EGAD, Florence Clavaud, qui est à la tête de l'équipe de développement de RiC-O et également responsable de la Mission Référentiels des AN ; un poste clé qui va nous permettre une plus grande compréhension des étapes d'élaboration de RiC ainsi que de la manière de se l'approprier.

Enfin, RiC ayant pour vocation de présenter les données archivistiques sous forme de graphe, il s'avère alors intéressant de questionner les effets de ce type de visualisation sur l'appréhension des données archivistiques : la compréhension du contexte de production d'un fonds d'archives est-elle plus efficace sous la forme d'un graphe ? C'est ce que nous essaierons de comprendre par l'étude de la perception visuelle, des systèmes mnésiques ou encore des processus attentionnels. C'est donc une partie qui s'intéresse à la fois aux forces du graphe mais aussi aux limites de nos procédés cognitifs.

1. RiC : origines, conception, implémentation

La volonté de doter la description archivistique d'un modèle conceptuel est en partie due au projet *Functional Requirements for Bibliographic Records object oriented* (FRBRoo) publié en 2009. Ce dernier repose sur le modèle conceptuel des bibliothèques — *Functional Requirements for Bibliographic Records (FRBR)* — paru en 1998, et celui des musées — *Conceptual Reference Model* (CRM ou CIDOC-CRM du nom de *International Committee for Documentation* qui en est

l'un des auteurs) — publié pour le première fois en 1999. La volonté d'inclure les archives au cœur de ce projet suscite l'intérêt de la profession pour la constitution de son propre modèle conceptuel[290]. Un modèle nécessaire pour que — une fois de plus — la description archivistique soit en mesure de dialoguer avec la description bibliographique et muséologique : « it will […] enable us to collaborate better with fellow information professionals […] and position us to ensure that the archives in our care take their proper place in the developing digital world »[291] ; l'objectif à terme étant de permettre un accès intégré à l'expression humaine sous toutes ses formes[292].

1.1. L'EGAD, une mosaïque de traditions archivistiques ?

1.1.1. Constitution et méthodes de travail de l'EGAD

DDS, CDS, PSBP, CBPS, EGAD… les groupes de travail se succèdent au sein de l'ICA, et bien qu'ayant des particularités propres, tous ambitionnent d'atteindre une large représentativité des pratiques archivistiques existantes. Cette représentativité se doit d'être un élément essentiel de tout processus normatif engagé à cette échelle, « Pour mettre en pratique des normes à un niveau international, les principales difficultés résident de toute évidence dans la prise en compte des différentes traditions archivistiques et des différents contextes — physique, culturel et politique, [et] dans les questions de langues et de terminologie »[293]. Cependant, nous avons constaté que l'archivistique dominante est d'origine anglo-saxonne, ce que Claire Sibille de Grimoüard, experte en normalisation de la description archivistique, confirme[294]. Le colonialisme, la guerre froide et la domination culturelle et économique des États-Unis sont autant de facteurs ayant conduits à une domination du monde anglo-saxon dans les pratiques archivistiques ainsi que le rappelle Daniel Pitti, président de l'EGAD[295]. Une domination toujours perceptible ainsi que le souligne le groupe du projet InterPARES Trust[296] et que corrobore Claire Sibille de Grimoüard, « c'est vrai que si on regarde la composition des groupes de normalisation, il y a certains pays qui sont plus actifs que d'autres »[297].

Cependant, « la composition du groupe de RiC est plus diverse par rapport aux groupes précédents »[298] avec la représentation de 14 pays, dont le Brésil, la région administrative

---

[290] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 575.
[291] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *Comma*, vol. 2016, n°1-2, 2016, p. 177.
[292] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 575.
[293] Marion Beya, « Introduction, ICA et normes », *Comma*, vol. 2011, n°2, 2011, p. 3-4.
[294] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.19.16s.
[295] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°18.
[296] InterPARES Trust, « Comments on "Records in Context" », *InterPARES Trust blog*, 2016, p. 3, [en ligne], disponible sur https://interparestrustblog.wordpress.com/2016/12/11/interpares-trust-responds-to-egad-ric/ (consulté le 5 mars 2021).
[297] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.05.41s.
[298] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.19.16s.

spéciale de Hong Kong et des pays d'Europe de l'Est comme l'Autriche ou la Roumanie ; un membre de Côte d'Ivoire était également présent mais a démissionné dans le courant de l'année 2020[299]. L'EGAD s'inscrivant à la suite du CBPS, il hérite de certains de ses membres ; un « héritage » essentiel selon Daniel Pitti « to have continuity between the old and the new »[300]. Les autres membres qui composent le groupe de travail ont été recrutés par Daniel Pitti, selon leur expertise et dans la volonté de représenter au mieux une certaine diversité au sein de l'EGAD[301].

Aussi, puisqu'il est impossible d'envisager un processus de normalisation avec un grand nombre de personnes : « There are 195 countries in the world. I do not think having one representative from each would be realistic or manageable »[302], l'EGAD — comme ses prédécesseurs avant lui — remédie à cette situation en soumettant l'avancée de ses travaux à des appels à commentaires[303].

Des interrogations demeurent pourtant, notamment en ce qui concerne la représentation d'un même pays par plusieurs personnes et l'influence de cette représentation plus importante lors des prises de décisions[304]. Une crainte qui aurait été légitime si le mode de décision du groupe de travail reposait sur le vote, or depuis 1996 et la création de la CDS, les décisions se font par consensus, ainsi que nous l'avons vu précédemment ; une caractéristique dont l'EGAD a hérité[305].

A ces problématiques de représentation s'ajoute le choix de l'anglais comme langue de travail, un choix qui requiert « un effort conceptuel, un effort intellectuel, doublé d'un effort linguistique pour beaucoup d'entre nous »[306] ainsi que le souligne Florence Clavaud. Daniel Pitti en arrive à la même conclusion et précise que la patience se doit d'être de mise, « The overall strategy is to remind native speakers of English to speak slowly, clearly. And to be patient in giving non-native English speakers an opportunity to find the words to express their ideas »[307]. Un choix qui n'est pas neutre selon Claire Sibille de Grimoüard[308]. En effet, bien que l'ICA s'inscrive dans une tradition multiculturelle[309], le choix de l'anglais comme langue de travail soulève des difficultés terminologiques ; des difficultés surmontées par de longues réflexions, « We spent a lot of time carefully analyzing the etymology of terms, and we gravitated towards adopting Latin-based words, as, given the far reach and influences of the Roman Empire, the Latin-based terms tended to have acceptable equivalence terms in other languages »[310].

---

[299] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°9.
[300] *Ibid.*, question n°8.
[301] *Ibid.*
[302] *Ibid.*, question n°10.
[303] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.08.11s.
[304] InterPARES Trust, « Comments on "Records in Context" », *art. cit.*, p. 3.
[305] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°11.
[306] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.16.14s.
[307] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°17.
[308] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.08.11s.
[309] *Ibid.*, 00.11.21s.
[310] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°17.

Par ailleurs, les conflits d'intérêts autrefois discernés lors des processus de normalisation semblent grandement édulcorés au sein du groupe de travail[311] puisque « The overall attitude cultivated in the group is that none of knows everything, but each of us has knowledge to contribute to the deliberations. […] There are a lot of "know a lot" members of EGAD, but no "know it all." The latter is anathema to the success of a group »[312].

Aussi, RiC semble plus sensible à la conservation et à la prise en compte d'un plus grand nombre de traditions archivistiques. La nouvelle norme de l'EGAD ambitionne même d'aller plus loin encore : « RiC is aspirational, that is, it does not canonize existing practice, but lays the foundation for next generation archival description »[313]. Raison pour laquelle après avoir pensé la constitution du groupe de travail, il paraît nécessaire d'évaluer le cadre normatif actuel afin d'en comprendre les forces et les faiblesses. Une mesure nécessaire pour comprendre la raison de cette adoption inégale entre les quatre normes de l'ICA[314] et parvenir à déceler et comprendre ce qui nuit au développement d'une homogénéisation des pratiques. Une étape importante étant donné que l' « on ne fait pas une norme pour le plaisir de faire une norme, il faut que ce soit appliqué. Donc si des pays n'arrivent pas à suivre, c'est un problème »[315]. Un problème non des moindres puisque l'essence même de RiC — norme 4-en-1 qui vise à unifier les pratiques — serait menacée sans la prise en compte de ces pays ; dans cette perspective, RiC ne serait alors qu'une norme de plus, ainsi que l'illustre la bande suivante :

[311] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.18.22s.
[312] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°12.
[313] *Ibid.*, question n°6.
[314] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.15.55s.
[315] *Ibid.*

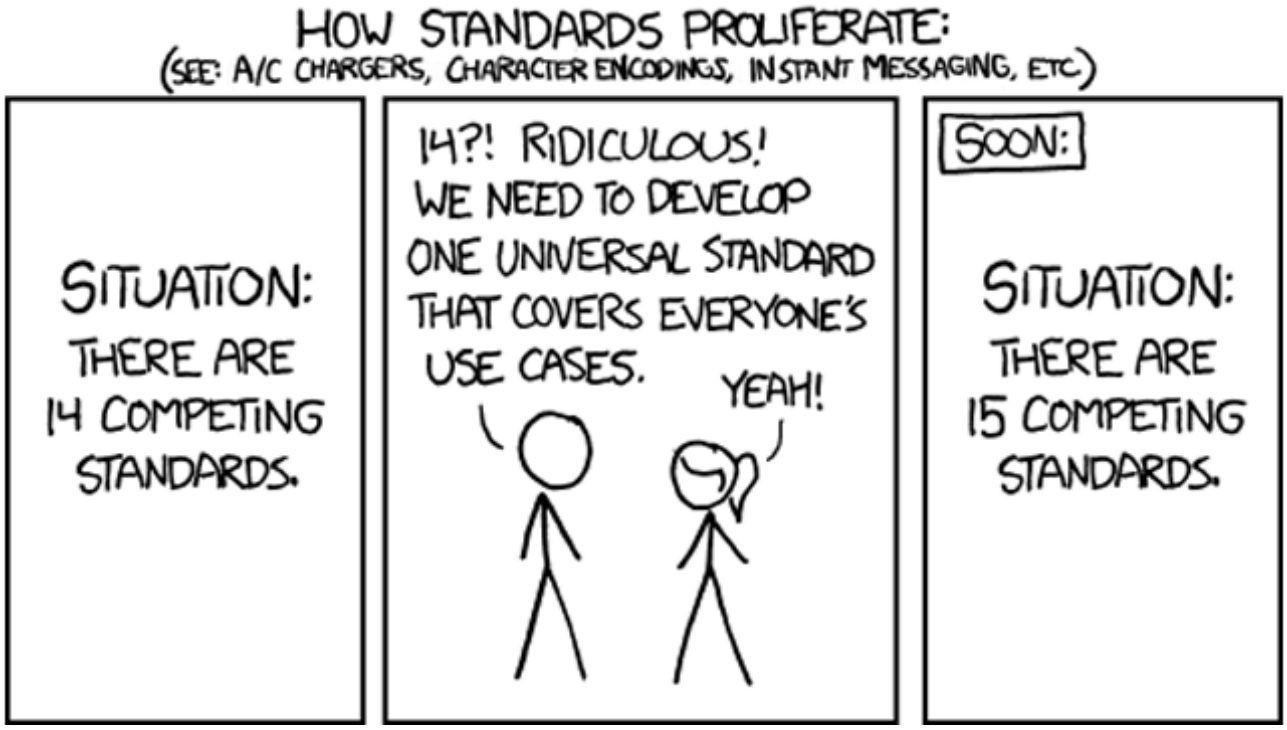

Illustration 1 — Brian Barrett, « Why We Can't All Just Settle on One Stupid Standard », *Gizmodo*, 2011, [en ligne], disponible sur https://gizmodo.com/why-we-cant-all-just-settle-on-one-stupid-standard-5822999 (consulté le 15 mai 2021).

Aussi, reconnaître qu'il y a actuellement une normalisation à deux vitesses est par ailleurs nécessaire pour mener à bien l'adoption de la norme une fois achevée : « il y a des pays qui en sont encore à se poser des questions sur la manipulation d'ISAD(G). Ces pays sont très éloignés d'une réflexion sur la modélisation de la description archivistique et sur l'apport du web de données pour valoriser les inventaires d'archives »[316]. Daniel Pitti est lucide sur le sujet, son expérience avec l'EAD lui a enseigné que l'adoption d'une nouvelle norme se faisait de manière graduelle[317]. Il ajoute que cela relève des devoirs de l'ICA que d'aider les pays qui en ont besoin à s'approprier les normes développées, « This, I think, is a long-term ethical and political challenge for ICA and the international archival community. […] Our collective memory needs to be inclusive of all human history. It the social justice dimension of archival professional principles and practice »[318]. Aussi, « Il ne s'agit […] en aucun cas d'introduire une rupture ; pour permettre la transition, RiC doit être compatible avec ces standards [le cadre normatif actuel de l'ICA] »[319], et ce, afin qu'aucun pays ne se sente dépassé par l'évolution de la normalisation de la description archivistique.

[316] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.20.43s.
[317] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°6.
[318] *Ibid.*
[319] Florence Clavaud, *Le nouveau standard international de description des archives, Records in Contexts*, actes de la journée d'étude organisée le 28 janvier 2020 par les Archives nationales à Paris, p. 6, [en ligne], disponible sur https://f.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_2_RecordsInContexts_englishVersionAdded1003.pdf (consulté le 18 janvier 2021).

1.1.2. Elaborer un modèle conceptuel : par où commencer ?

Les membres de l'EGAD se rencontrent lors de la première réunion du groupe à Bruxelles en novembre 2013 afin de « commencer par trouver des points de consensus »[320] pour définir ses objectifs. Très vite, la nécessité d'élaborer une transposition technique en parallèle du modèle conceptuel se fait ressentir, au risque que le modèle conceptuel ne soit jamais adopté, ainsi que le souligne Florence Clavaud[321]. Claire Sibille de Grimoüard précise que le groupe a donc commencé par réaliser un état des lieux de la réflexion. Les communautés professionnelles apparentées ont bien entendu été une source d'inspiration importante puiqu'en partie à l'origine du processus de développement de RiC comme vu précédemment. A cela s'ajoute des projets nationaux en matière de modélisation de la description archivistique, des travaux sur lesquels l'EGAD s'est appuyé pour élaborer RiC[322]. Parmi eux figurent une initiative australo-néo-zélandaise en 2008 et un projet espagnol en 2012 suivi de près par une approche finlandaise en 2013. Le Président de l'EGAD nous apporte plus de précisions sur le sujet : « Broadly, Australian theory had considerable influence on Activity (or function) and also Mandate. The Spanish model was influential with respect to how it was organized and presented, and in some areas also on specific definitions, for example, Rule. The Finnish model was influential in much the same way as the Spanish model »[323].

L'initiative binationale repose sur la seconde version de la norme australienne intitulée *Australian Government Recordkeeping Metadata Standard* (AGRkMS) et la première version de la norme néo-zélandaise *Technical Specifications for the Electronic Recordkeeping Metadata Standard* ; deux normes qui puisent les origines au sein de la norme *Recordkeeping Metadata Standard for Commonwealth Agencies* publiée en 1999[324]. AGRkMS s'appuie elle-même sur deux normes ISO : la *Australian Standard on Records Management* (AS ISO 15489) et la *Metadata for Records* (AS ISO 23081)[325], deux normes propres au records management dont l'EGAD a tenu compte pour l'élaboration de RiC, un premier pas vers une plus grande harmonie des normes ISO/ICA, et par conséquent, entre la description des archives courantes et intermédiaires et celle des archives définitives. Par ailleurs, AGRkMS s'appuie également sur le projet SPIRT mené par l'université de Monash en Australie et dont le premier résultat est visible dès 2000 : « The SPIRT Recordkeeping Metadata Project, led by Monash University, developed the first multiple-entity approach to recordkeeping metadata »[326] ; un modèle qui identifie cinq

---

[320] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.13.00s.
[321] *Ibid.*
[322] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.05.41s.
[323] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°20.
[324] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 575.
[325] *Ibid.*, p. 576.
[326] National Archives of Australia, *Australian Recordkeeping Metadata Standard (AGRkMS),* version 2.2, 2015, p. 9, [en ligne], disponible sur https://www.naa.gov.au/sites/default/files/2019-09/AGRkMS-Version-2.2-June-2015_tcm16-93990_1.pdf (consulté le 15 mai 2021).

entités : Records, Agent, Business, Mandate et Relationship[327]. Aussi, au sein de l'AGRkMS, « The Relationship entity plays the role of interrelating the other four entities to form a complete description. Relationship is event based, which is to say that relations document specific human activities or events »[328]. Une particulatité sur laquelle se construit RiC.

L'initiative espagnole — *Comisión de Normas Españolas de Descripción Archivística* (CNEDA) — marche dans les pas des AGRkMS en ayant recourt à la modélisation de données selon le modèle Entité-Relation. Elle prend en considération les quatre normes actuelles de l'ICA ainsi que le modèle conceptuel du progiciel open source *Access to Memory* (AtoM)[329]. Comme les AGRkMS, la CNEDA identifie les entités suivantes : Records, Agent, Business et Mandate, auxquelles s'ajoutent l'entité Concept, Object, or Event ainsi que l'entité Place[330].

Enfin, le *Finnish Conceptual Model for Archival Description* se rapproche davantage du modèle qui sous-tend le FRBRoo, tout en tenant compte des quatres normes de l'ICA. Les archivistes finlandais distinguent également les entités Business, Agent, Mandate et Place, auxquelles viennent se greffer les entités Information Resource, LifeCycle Event, Temporal Event, et Subject[331].

Mais ces trois modèles ne sont pas les seules sources d'inspiration de l'EGAD. Au Royaume-Uni par exemple, le projet LOCAH visant à exposer les données — encodées en EAD — du portail britannique *Archives Hub* en *linked open data* (LOD) est également un modèle sur lequel l'EGAD s'est appuyé pour le développement de RiC[332]. A cela se conjuguent le projet InterPARES Trust — qui porte sur la fiabilité des documents et des données numériques —, le dictionnaire de données PREMIS — qui est un modèle pour les métadonnées de préservation — ainsi que le cadre normatif du *World Wide Web Consortium* (W3C) avec l'ontologie PROV (PROV-O) et le langage de représentation des connaissances *Web Ontology Language* (OWL)[333] ; autant de cadres normatifs sur lesquels l'EGAD s'est appuyé pour développer RiC. Aussi, il paraît naturel de s'interroger sur la manière dont le groupe de travail s'inspire et s'approprie ses différents projets issus des quatre coins du globe : de quelle manière notre nouvelle norme va-t-elle se développer ? Quelles vont être ses particularités ? Comment va-t-elle refléter toutes les spécificités et les richesses de la description archivistique ? C'est ce que nous nous efforcerons de comprendre à travers l'étude du modèle conceptuel et de l'ontologie.

---

[327] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 576.
[328] *Ibid.*
[329] *Ibid.*
[330] *Ibid.*
[331] *Ibid.*, p. 577.
[332] *Ibid.*
[333] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°20.

1.2. Redéfinir les fondamentaux de la description archivistique

Avant de présenter le modèle conceptuel — *RiC Conceptual Model* (RiC-CM) — élaboré par l'EGAD, il convient tout d'abord de préciser ce qu'est un modèle conceptuel. Le groupe de travail en apporte une définition claire et concise au sein de son premier rapport — qui fait suite à la réunion du groupe tenue à Bruxelles en 2013 — *Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description*. Ainsi, ce rapport explique qu'un modèle conceptuel représente les principaux concepts, leurs attributs et les relations qui lient ces concepts entre eux, pour un domaine de connaissance donné[334]. Lors de la journée d'étude du 28 janvier 2020 aux Archives nationales, Florence Clavaud précise que « RiC se concentre sur les entités qui constituent le monde des archives, non sur les règles de description de ces entités »[335]. Des règles — ISAD/G, ISAAR/CPF, ISDF et ISDIAH — « élaborées à des époques différentes donc […] mal articulées les unes par rapport aux autres, fatalement »[336].

Au sein de ce rapport, l'EGAD énonce que sa première mission — peut-être la plus difficile — consiste à identifier les entités de haut niveau[337]. Pour ce faire, Daniel Pitti nous explique que « The identification of the core entities was primarily based on an analysis of existing ICA standards augmented by reviewing a variety of professional literature »[338]. Au final, l'EGAD définit 22 entités de haut niveau, ainsi que l'illustre le tableau suivant issu de la version 0.2 de RiC-CM parue en décembre 2019 — la version 0.1 est parue en septembre 2016 — :

[334] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 574.
[335] Florence Clavaud, *Le nouveau standard international de description des archives, Records in Contexts*, *art. cit.*, p. 8.
[336] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.15.55s.
[337] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 579.
[338] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°23.

RiC Entities Hierarchy			
First Level	**Second Level**	**Third Level**	**Fourth Level**
RiC-E01 Thing	**RiC-E02 Record Resource**	RiC-E03 Record Set	
		RiC-E04 Record	
		RiC-E05 Record Part	
	RiC-E06 Instantiation		
	RiC-E07 Agent	RiC-E08 Person	
		RiC-E09 Group	RiC-E10 Family
			RiC-E11 Corporate Body
		RiC-E12 Position	
		RiC-E13 Mechanism	
	RiC-E14 Event	**RiC-E15 Activity**	
	RiC-E16 Rule	RiC-E17 Mandate	
	RiC-E18 Date	RiC-E19 Single Date	
		RiC-E20 Date Range	
		RiC-E21 Date Set	
	RiC-E22 Place		

Tableau 1 — Conseil international des Archives, *Records in Contexts, A Conceptual Model for Archival Description, Consultation Draft v0.2* (preview), 2019, p. 6, [en ligne], disponible sur https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf (consulté le 18 janvier 2021).

Des entités dont les noms ont été souvent fondés sur la terminologie actuelle précise Daniel Pitti[339]. Cependant, lorsque le groupe de travail ne parvenait pas à se mettre d'accord — en raison du défi omniprésent de la traduction — ou lorsque le nom original était imprécis ou ambigu, un autre était choisi ; pour ce faire, le groupe s'est appuyé sur des noms d'origine latine pour surmonter les difficultés induites par la barrière de la langue[340], « We spent a lot of time carefully analyzing the etymology of terms, and we gravitated towards adopting Latin-based words, as, given the far reach and influences of the Roman Empire, the Latin-based terms tended to have acceptable equivalence terms in other languages »[341]. Des difficultés linguistiques qui ne sont pas sans rappeler l'idée émise par John Locke selon laquelle langue et culture sont intimement liées, car la plupart des mots expriment les us et coutumes d'un peuple[342]. Une idée que la vision humboldtienne nous permet de saisir pleinement : « [Humboldt] voit les langues comme des prismes […] chaque langue reflète sa propre vision du monde »[343]. Cela nous permet de comprendre dès lors pourquoi cette étape a demandé un certain temps de concertation aux membres du groupe dont l'objectif est de permettre « une description indépendante de l'histoire de chaque entité archivistique »[344].

---

[339] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°23.
[340] *Ibid.*
[341] *Ibid.*, question n°17.
[342] John Locke, « Chapter XXII, Of Mixed Modes », *An Essay Concerning Human Understanding*, London, Penguin Classics, 1998, p. 370.
[343] Nicole Delbecque, *Linguistique cognitive, comprendre comment fonctionne le langage*, Bruxelles, De Boeck, 2006, p. 164.
[344] Bogdan Florin Popovici, « *Records in Contexts* : vers un nouveau niveau dans la description archivistique ? », *Archives*, vol. 48, n°2, 2019, p. 30.

Parmi ces 22 entités, il y a tout d'abord les entités relatives aux archives elles-mêmes — toutes appartenant à l'entité Record Resource (RiC-E02) —, première "famille" d'entités élaborée par l'EGAD : « We begin then with the "what", that is the records themselves »[345]. Elles sont au nombre de trois : Record Set (RiC-E03), Record (RiC-E04) et Record Part (RiC-E05). L'élément central — Record (RiC-E04) — fait écho à l'unité de description « pièce » d'ISAD/G[346] qui précise qu'il s'agit de la plus petite unité de description envisageable. RiC affine encore cela en proposant l'entité Record Part (RiC-E05), qui offre la possibilité de décrire au niveau d'une partie de la pièce ; réalisable dès lors que la partie en question contient des informations lui permettant d'atteindre une certaine unité physique et intellectuelle[347]. Enfin, l'entité Record Set (RiC-E03) concerne les archives réunies ensemble parce que résultant d'une même activité, d'une même fonction, d'un même sujet ou revêtant une même forme[348]. En ce sens, cette entité témoigne à son tour d'un lien de parenté avec ISAD/G, « The set then may be a "series" of records, as traditionally understood »[349]. Etroitement liée à Record Resource (RiC-E02), l'entité Instantiation (RiC-E06) — qui représente la manifestation physique d'un document d'archives — a exigé un certain temps de réflexion, ainsi que l'illustre Daniel Pitti, « We also spent a lot of time differentiating a Record from the physical Instantiation of the record based on the practical observation that in many cases, for preservation and other reasons, an original instance of a Record may be "re-instantiated" and the new instantiation was considered functionally equivalent, or not »[350]. A cette considération s'ajoute la volonté de proposer des entités de haut niveau qui puissent être étendue par le dictionnaire de données PREMIS pour les instantiations numériques ; une nécessité étant donné que les données relatives à la gestion technique ne sont pas propres aux archives, mais sont aussi un objet d'intérêt pour les communautés professionnelles apparentées[351].

Le second groupe d'entités sur lequel l'EGAD s'est attelé concerne celui des personnes : « We then move to the "who", those who create, manage and use records »[352]. Elles sont au nombre de quatre, toutes agrégées sous l'entité de second niveau Agent (RiC-E07). Les entités Person (RiC-E08) et Group (RiC-E09) — contenant elle-même les entités Family (RiC-E10) et Corporate Body (RiC-E11) — reflètent directement l'élément Type of Entity de la zone d'identification d'ISAAR/CPF. L'entité Position (RiC-E12) partage quant à elle de fortes ressemblances avec l'élément Functions, occupations and activities de la norme de description

---

[345] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 178.
[346] *Ibid.*
[347] Conseil international des Archives, *Records in Contexts, A Conceptual Model for Archival Description, Consultation Draft v0.2* (preview), *op. cit.*, p. 9.
[348] Association des archivistes français, *Abrégé d'archivistique, Principes et pratiques du métier d'archiviste*, *op. cit.*, p. 140.
[349] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 178.
[350] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°23.
[351] *Ibid.*
[352] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 179.

des producteurs et *a fortiori*, avec ISDF elle-même ; elle correspond en effet à une description des fonctions mais diffère de ses aînées pour ne pas y inclure la notion d'activité qui elle est décrite à part au sein de RiC-CM, nous y reviendrons. Enfin, le modèle conceptuel reconnaît qu'un processus ou un système — qui peut tout autant présenter des composants mécaniques et logiciels — agit comme un agent puisque pouvant générer ou modifier des archives[353]. Cette entité se présente sous le nom de Mechanism (RiC-E13).

Ainsi que nous l'avons vu, la notion d'activité dispose d'une entité propre nommée Activity (RiC-E15). Cela permet au modèle conceptuel de proposer une description toujours plus fine. En effet, à l'image d'ISDF en 2008 qui souhaitait proposer une description des fonctions et des activités indépendantes de la norme ISAAR pour faire face aux diverses évolutions administratives et ainsi faciliter les recherches de l'utilisateur final, RiC-CM tamise ISDF et sépare les fonctions des activités pour les mêmes raisons. Conjuguée à l'entité Position (RiC-E12), elle détermine les circonstances dans lesquelles les documents d'archives sont créés : « We […] have the entities that models the "how" and "why" agents act and create records: the business entities »[354].

Une entité — Mandate (RiC-E17) — fait office de pont entre ces « business entities » et les entités Person (RiC-E08), Group (RiC-E09) ainsi que Mechanism (RiC-E13) — car nécessairement créé par l'une des deux première entités[355] — : « Tying the business entities together with agents is Mandate »[356]. La raison à cela est que cette entité fait référence aux règles qui déterminent les fonctions et les activités exercées par une personne[357].

Toutes ces entités — Record Resource, Instantiation, Agent, Activity et Mandate — sont propres à l'univers des archives. Elles sont par ailleurs complétées par des entités partagées avec les communautés professionnelles apparentées telles que les entités Date (RiC-E18) et Place (RiC-E22)[358].

Une fois les entités définies, l'EGAD se livre à l'identification de leurs propriétés[359]. Une étape importante et qui a encore exigé des membres bon nombre de réflexions : « With respect to attributes, we again paid attention to current practices, but also our experience with a variety of projects, and attempted to identify new attributes when existing attributes did not address something we considered important, or we may have differentiated an existing attribute into

---

[353] Conseil international des Archives, *Records in Contexts, A Conceptual Model for Archival Description, Consultation Draft v0.2 (preview)*, op. cit., p. 14.
[354] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 179.
[355] Conseil international des Archives, *Records in Contexts, A Conceptual Model for Archival Description, Consultation Draft v0.2 (preview)*, op. cit., p. 14.
[356] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 179.
[357] *Ibid.*
[358] *Ibid.*, p. 179-180.
[359] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 579.

more than one attribute when the existing was imprecise or ambiguous »[360]. Elles sont au nombre de 41[361]. Parmi elles, il y a tout d'abord celles communes à toutes les entités — Descriptive note, Identifier et Name —, auxquelles s'ajoutent des propriétés qui n'appartiennent qu'à une entité en particulier[362]. L'entité Record Resource (RiC-E02) par exemple comprend treize propriétés qui peuvent être réparties en quatre zones[363] : une zone relative au contenu, une autre pour la représentation, une autre dédiée au support matériel, et une dernière qui rassemble des propriétés qui ont trait à la gestion et à l'utilisation des archives. Les propriétés qu'elles contiennent ne sont pas sans rappeler divers éléments propres à ISAD/G dont voici quelques exemples issus de ces quatre zones : RiC-A38 Scope and content qui emprunte le même nom que son homologue dans ISAD/G, RiC-A25 Language dont l'équivalent est Language/scripts of material, RiC-A35 Record Resource Extent pour l'élément Extent and medium of the unit of description (quantity, bulk, or size), ou encore RiC-A21 History pour l'élément Archival history.

Le même phénomène se produit avec l'entité Agent (RiC-E07) où — pour ne citer que lui —, History est l'élément d'ISAAR/CPF éponyme de RiC-A21 History. Il est également possible d'apercevoir au sein de cette entité des caractéristiques propres à ISDIAH ; en effet, les propriétés relevant de l'entité Corporate body (RiC-E11) permettent « the capture of the information in relation to institutions holding archives […] : services to the public, contact information, operating hours and facilities »[364].

Une fois les entités définies et leurs propriétés identifiées, il convient de dresser la liste des relations possibles entre ces entités[365]. RiC en présente 78 — sans les relations inverses — [366], et Daniel Pitti reconnaît que cette étape a été des plus délicates : « This was perhaps this single most difficult challenge »[367]. La difficulté réside dans le caractère sans fin des relations potentiellement identifiables ; ainsi Daniel Pitti et Florence Clavaud se sont astreints à définir la quantité minimale nécessaire bien que cela représente déjà un nombre conséquent de relations. Pour ce faire, ils ont fait le choix de ne pas décrire les relations sortant du cadre professionnel : « One thing we clearly avoided was identifying relations between persons that were qualitative and transitory: no "best friend forever," and such. Some relations are transitory, but none should be both transitory and qualitative »[368]. Il peut être intéressant de mentionner que la précision

---

[360] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°23.
[361] Florence Clavaud, *Le nouveau standard international de description des archives, Records in Contexts*, *art. cit.*, p. 24.
[362] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 180.
[363] *Ibid.*
[364] *Ibid.*, p. 181.
[365] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 579.
[366] Florence Clavaud, *Le nouveau standard international de description des archives, Records in Contexts*, *art. cit.*, p. 28.
[367] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°26.
[368] *Ibid.*

de RiC est visible jusque dans le temps du verbe de la relation, puisque si cette dernière est considérée comme achevée, alors la relation s'exprime au passé[369].

Aussi, « RiC permet enfin une plus grande précision. Avec 22 entités, 41 attributs et 78 relations (plus les relations inverses) on dispose d'un arsenal permettant d'atteindre une granularité bien plus importante qu'avant »[370] ainsi que le souligne Florence Clavaud. Il est également intéressant de relever que la sortie de la bêta du modèle conceptuel s'accompagne de la version alpha de l'ontologie. Une fois de plus, des questions émergent : RiC-CM a-t-il influé sur le développement de RiC-O — et inversement — ? Comment s'organise la construction d'une telle implémentation technique ? Sur quels principes se fonde-t-elle ? A l'instar de l'EAD pour ISAD/G ou de l'EAC pour ISAAR/CPF, RiC-O est-elle plus précise encore que le modèle conceptuel dont elle découle ? Autant d'interrogations auxquelles nous nous devons d'apporter des réponses.

1.3. « Qui contrôle les métadonnées, contrôle le Web »[371]

Cet adage de Fabien Gandon — représentant de l'Institut national de recherche en informatique et en automatique (INRIA) au W3C — illustre l'enjeu de taille auquel doivent faire face les archivistes, et à plus grande échelle, tous les professionnels en sciences de l'information. Avec le web sémantique, l'utilisation de métadonnées de qualité est indispensable pour garantir l'accès à l'information. Ainsi que nous l'avons vu, la tridimensionnalité du document évolue ; le choix de doter la description archivistique d'une ontologie de domaine n'est pas sans rappeler l'entrée de la dimension intellectuelle du document dans le web sémantique. Un domaine complexe puisque se trouvant à la croisée de plusieurs disciplines : « les chercheurs en informatique et ingénierie des connaissances croisent des préoccupations anciennes, très anciennes même des chercheurs en Sciences de l'homme depuis la philosophie jusqu'aux Sciences de l'information, en passant par les différents courants des Sciences du langage »[372].

1.3.1. Du modèle conceptuel à l'ontologie

Accompagner RiC-CM d'une ontologie est non seulement primordial pour que le modèle soit utilisé, mais cela présente également l'opportunité de tester le modèle ainsi que le souligne Florence Clavaud : « RiC-O a servi de laboratoire pour RiC-CM »[373] avant d'ajouter que les réflexions autour du modèle conceptuel ont également nourri l'ontologie[374]. En ce sens, RiC-O a

---

[369] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 181.
[370] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°3.
[371] Fabien Gandon, *Données liées et Web sémantique : quand le lien fait sens*, actes de la présentation pour les rencontres Inria Industrie « Technologies du Web », 2014, p. 58, [en ligne], disponible sur https://fr.slideshare.net/fabien_gandon/web-semlinkeddata-fabiengandon (consulté le 14 mai 2021).
[372] Jean-Michel Salaün, « La redocumentarisation, un défi pour les sciences de l'information », *art. cit.*, p. 5-6.
[373] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.07.04s.
[374] *Ibid.*

fait office de preuve de concept pour RiC[375]. Mais comment fonctionne-t-elle exactement ? Pour le comprendre, il convient tout d'abord de saisir le cadre normatif du web sémantique. Les images ci-dessous nous permettent d'en avoir une vue d'ensemble :

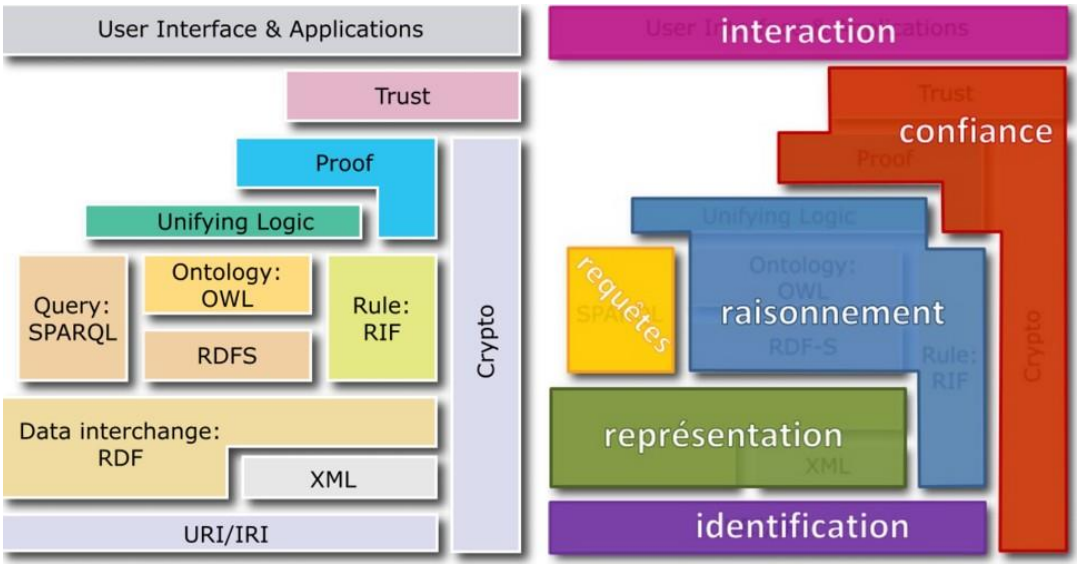

Illustration 2 — Fabien Gandon, *Données liées et Web sémantique : quand le lien fait sens*, *op. cit.*, p. 29 et p. 52.

Ainsi, RiC-O, en tant que transposition en OWL de RiC-CM, repose sur le langage nommé *Resource Description Framework* (RDF). Avec ce dernier, « chaque ressource est décrite par un ensemble de trois balises : le triplet »[376] ; ce triplet se compose d'un sujet, d'un prédicat et d'un objet, nous y reviendrons. Le langage OWL sur lequel est bâtie RiC-O permet de raisonner sur les données représentées. Au sein de ce langage, les entités du modèle conceptuel deviennent des classes[377] ; Florence Clavaud précise par ailleurs que c'est justement parce toutes ces classes d'objets découlent des entités de RiC-CM — et par conséquent qu'elles en partagent les définitions — qu'il est question de transposition[378]. RiC-O présente également des classes ne relevant pas des entités du modèle conceptuel, mais de certaines de ses propriétés pouvant être contenues au sein d'un vocabulaire contrôlé ; une manière d'articuler RiC-O avec des vocabulaires dits *Simple Knowledge Organization System* (SKOS)[379], soit des thésaurus. Parmi ces propriétés figurent également Language, Name et Identifier — les propriétés communes à toutes les entités au sein de RiC-CM — si ces dernières justifient d'être considérées comme des

[375] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°29.
[376] Isabelle Westeel, « Indexer, structurer, échanger : métadonnées et interopérabilité », *Numériser et mettre en ligne*, Villeurbanne, Presses de l'enssib, 2010, p. 119.
[377] Daniel Pitti, Bill Stockting, Florence Clavaud, « An introduction to "Records in Contexts": an archival description draft standard », *art. cit.*, p. 186.
[378] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.10.20s.
[379] Conseil international des Archives, *International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2*, 2021, [en ligne], disponible sur https://www.ica.org/standards/RiC/RiC-O_v0-2.html (consulté le 3 mars 2021).

entités à part entière en sein de RiC-O[380]. RiC-O reconnaît également des classes pouvant apporter une définition plus précise pour certaines entités, comme c'est le cas avec les lieux par exemple : « [Places] exist, are identifiable, have histories, and can be geo-located (over time if necessary). A place has an identity »[381]. Enfin, certaines relations sont également représentées par des classes au sein de RiC-O dans les cas où ces dernières sont datées et disposent d'une description : « you may need to assign different attributes to a relation, e.g. a date, certainty or description, as it is already possible, and quite often done, in a XML/EAC-CPF file »[382]. Les autres relations présentées au cœur de RiC-CM prennent la forme de propriétés dans RiC-O. Elles constituent le prédicat du triplet RDF.

Ainsi que nous l'avons vu, le langage OWL repose sur le langage RDF, qui lui-même repose sur des *Uniform Resource Identifiers* (URI) — pour « identifier, sur le web, ce qui existe »[383] — ou des *Internationalized Resource Identifiers* (IRI) — pour « identifier, sur le web, en toute langue ce qui existe »[384] —. Ces identifiants sont indispensables au web sémantique et permettent de renvoyer au document identifié. Au sein d'un triplet RDF, la balise <subject> porte donc l'URI/l'IRI de la ressource décrite, la balise <predicate> constitue un type de propriété applicable à la ressource identifiée et est également dotée d'une URI/IRI, et la balise <object> représente la valeur prise par cette propriété[385] ; cette valeur peut par ailleurs être une autre ressource identifiée alors par un URI/IRI qui lui est propre. A titre d'exemple pour l'énoncé : « Etoile des limbes est un tapuscrit de Humeau Edmond écrit en 1940 », cela donnerait en RDF trois triplets : Etoile des limbes <subject>, est un <predicate>, tapuscrit <object> ; Etoile des limbes <subject>, a été écrit par <predicate>, Edmond Humeau <object> ; Etoile des limbes <subject>, est daté de <predicate>, 1940 <object>. Formellement sous RiC-O cela donnerait :

```
<rico:record rdf:about="Record/URIdeRecord">
        <rico:title xml:lang="fr">Etoile des limbes</rico:title>
        <rico:hascreator rdf:resource="Agent/0000 0001 2140 896X"/>
        <rico:singledate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1940</rico:singledate>
        <rico:hasdocumentaryformtype rdf:resource="http://data.culture.fr/thesaurus/page/ark:/67717/T3-167"/>
</rico:record>
```

Étant donné qu'il est question d'une pièce, l'entité RiC-E04 a été choisie ; au sein de RiC-O, il s'agit d'une classe. Le titre du tapuscrit est un attribut — RiC-A28 — qui correspond à une classe dans RiC-O. La typologie de la pièce relève de la propriété RiC-A17 qui est en

[380] *Ibid.*
[381] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°33.
[382] Conseil international des Archives, *International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2*, *op. cit.*
[383] Fabien Gandon, *Données liées et Web sémantique : quand le lien fait sens, op. cit.*, p. 14.
[384] *Ibid.*
[385] Isabelle Westeel, « Indexer, structurer, échanger : métadonnées et interopérabilité », *op. cit.,* p. 119.

revanche une classe au sein de l'ontologie. Enfin, la date relève de l'entité RiC-E19 et est par conséquent une classe dans RiC-O.

1.3.2. Les principes de RiC-O

A l'instar des *Ann Arbor Accords* de l'EAD, RiC-O présente également des principes — cinq — sur lesquels elle se fonde. Tout d'abord, RiC-O se présente comme une ontologie de domaine, soit comme un vocabulaire générique avec des règles formelles qui peut créer des ensembles de données RDF — ou les générer à partir de métadonnées d'archives existantes — afin de décrire de manière cohérente tout type d'archives[386]. Un second principe est que l'ontologie se doit d'être immédiatement utilisable ; principe dont Florence Clavaud précise la portée : « Utilisable, ça veut dire qu'on peut l'appliquer aujourd'hui, c'est très important, on n'est pas obligé d'attendre d'avoir opéré des transformations dans ces données »[387]. Cela signifie que même les métadonnées actuelles qui sont peu structurées, voire non structurées comme RiC-CM le recommande, peuvent tout de même trouver leur place au sein de l'ontologie[388]. Par ailleurs, afin d'honorer ce principe, l'EGAD envisage de fournir toute la documentation nécessaire à son appropriation : tutoriels, exemples, ateliers ainsi que d'enrichir encore la documentation déjà présente sur le site du Conseil international des Archives[389]. Indispensable, le principe de l'utilité est étroitement lié au premier ainsi que l'illustre Florence Clavaud, « La fonctionnalité renvoie non seulement au fait que ce soit utilisable, […] mais aussi au fait que c'est intéressant, que c'est utile »[390]. Un principe essentiel puiqu'il sert à démontrer « qu'on gagne quelque chose, parce que faire du travail de transformation, c'est aussi coûteux »[391]. A ces principes s'ajoute la nécessité pour RiC-O d'être flexible, un principe étroitement lié à celui demandant à RiC-O d'être utilisable immédiatement. En effet, RiC-O doit proposer une superpropriété ou une superclasse, plus générale ou générique que ses sous-propriétés ou sous-classes mais aussi des sous-composants plus précis. En ce sens, RiC-O embrasse pleinement l'essence de la description archivistique[392] : « C'est en fonction des moyens, des besoins, des types de documents, que vous allez vous-même définir si vous faites une description très précise, si vous descendez au niveau de la pièce ou si vous vous arrêtez à la série, à la sous-série, au dossier, et si vous avez les moyens de décrire un agent précisément, avec toutes les caractéristiques qui permettent de comprendre notamment quels rôles il a joué, quelles productions documentaires il a pu avoir, quels échanges documentaires il a pu faire avec d'autres »[393]. Enfin, cinquième et dernier

---

[386] Conseil international des Archives, *International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2*, *op. cit.*
[387] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.27.40s.
[388] Conseil international des Archives, *International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2*, *op. cit.*
[389] *Ibid.*
[390] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.27.40s.
[391] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.27.40s.
[392] *Ibid.*, 00.25.35s.
[393] *Ibid.*

principe, l'ontologie se doit d'être extensible afin de rester utilisable même au sein de contextes non strictement archivistiques[394]. Ce principe ouvre la porte à l'interopérabilité, puisqu'il attend de RiC-O de pouvoir être alignée avec d'autres ontologies telles que celle des bibliothèques ou celle des musées par exemple. De plus, ce principe permet à l'ontologie de pouvoir être liée à des thésaurus — encodés en SKOS —.

1.3.3. RiC-O : tout pour plaire ?

Bien que RiC permette « a solid intellectual and technological foundation for collaborations within and outside the community »[395], il existe des critiques à son sujet, et plus particulièrement au regard de l'ontologie. Jean-Séverin Lair — ancien directeur du programme *Valeurs immatérielles transmises aux archives pour mémoire* (VITAM) — nous explique tout d'abord que le web sémantique est une technologie idéalisée à l'avenir incertain[396]. Selon lui, bien qu'il y ait des gens pour décrire leurs données en LOD, peu de personnes les utilisent en tant que telles[397]. Les technologies étant en constante évolution, il prévient des risques inhérents à l'implémentation d'une technologie complexe : « son temps d'adoption fait qu'elle n'est pas intéressante »[398]. Un écueil également soulevé par Edouard Vasseur — ancien responsable fonctionnel de VITAM — qui s'interroge sur les capacités requises pour s'approprier RiC :

> *« Est-ce que c'est un changement possible pour tout le monde ? Parce que c'est quand même très philosophique : c'est lié à des concepts qui séparent des entités qui sont reliées entre elles par des liens, etc. Intellectuellement, ça nécessite d'y réfléchir et d'y penser. »*[399]

L'enjeu repose finalement sur toute la documentation et les ateliers qui vont accompagner la parution de la v.1.0 de RiC afin que le modèle et son ontologie soient adoptés rapidement.

Le temps d'adaptation nécessaire à l'appropriation de RiC-O est également une contrainte au sens où ce temps n'est pas employé à traiter les fonds encore non communiqués car non décrits : « Je préfère que les services d'archives se concentrent sur faire en sorte que les archives soient classées et décrites, parce qu'il y a encore des archives de l'Ancien régime et des archives du 19ème qui ne sont pas décrites »[400]. Aussi selon Edouard Vasseur, la priorité n'est pas de

---

[394] Conseil international des Archives, *International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2*, *op. cit.*
[395] Gretchen Gueguen, Vitor Manoel Marques Da Fonseca, Daniel Pitti, Claire Sibille de Grimoüard, « Toward an International conceptual model for archival description: a preliminary report from the International Council on Archives' experts group on archival description », *art. cit.*, p. 580.
[396] Entretien du 25 mars 2021 avec Jean-Séverin Lair, annexe 1, 00.16.26s.
[397] *Ibid.*, 00.18.45s.
[398] *Ibid.*, 00.20.10s.
[399] Entretien du 26 mars 2021 avec Edouard Vasseur, annexe 1, 00.35.33s.
[400] Entretien du 26 mars 2021 avec Edouard Vasseur, annexe 1, 00.38.53s.

répondre au besoin de visibilité sur le web[401], mais davantage à un besoin de communication ; une problématique à laquelle il a été personnellement confronté en tant que chercheur[402].

Par ailleurs, plus que le temps nécessaire à son appropriation, c'est le fonctionnement même de la technologie sur laquelle repose RiC-O qui est remise en question. En cause : la volumétrie. A ce sujet, Jean-Séverin Lair précise que pour être traitée, l'informatique actuelle répartit cette volumétrie sur plusieurs serveurs ; or, cette solution se prête mal pour la technologie des graphes[403]. Il cite pour exemple le cas de la Bibliothèque nationale de France (BnF) : « la BnF a fait une base complète graphe où les interrogations sont extrêmement rares et où ça peut mettre des heures à trouver un résultat. Et elle a fait une base réduite, avec quelques métadonnées par objets, où là on peut faire des interrogations en secondes »[404]. Jean-Séverin Lair mentionne cependant que les données en LOF sont intéressantes pour des référentiels limités[405] tels que le référentiel des producteurs des AN par exemple. Aussi, l'utilisation d'un modèle mixte lui apparaît plus pertinente au vu du contexte. Il s'agirait d'allier à la fois des référentiels en LOD et des métadonnées d'objets répondant à une logique floue — celle des moteurs de recherche — qui « permet d'arriver relativement rapidement à l'information qui vous intéresse mais par approximation »[406]. L'avantage premier serait donc de pouvoir accéder à l'information recherchée, quand bien même la requête formulée par l'utilisateur final est imparfaite : « pour lui [le modèle sémantique] « Mozart » avec un « t » ou « Mozart » sans « t », ce sont deux nœuds différents du modèle sémantique, et il ne va pas forcément chercher l'un et l'autre en même temps »[407].

Enfin, Jean-Séverin Lair comme Edouard Vasseur regrettent l'absence de collaboration entre les métiers des archives et ceux de l'informatique. Cette association — expérimentée lors de l'élaboration du programme VITAM — a été riche en enseignements pour les deux professions. Edouard Vasseur explique que « C'est la complémentarité qui est intéressante. On est deux professions totalement complémentaires parce qu'on travaille sur la donnée »[408], ce que confirme et précise Jean-Séverin Lair : « La caricature que je faisais mais quelque part elle a du sens, c'est que les archivistes pensent en siècles et les informaticiens en nanosecondes »[409]. L'ancien responsable fonctionnel de Vitam revient également sur l'importance du rôle joué par les informaticiens lorsque ces derniers ont amené les archivistes à se questionner sur leurs besoins métier, à s'interroger sur le « quoi » et non sur le « comment » qui lui relève des compétences et des savoir-faire des premiers[410]. Aussi, Edouard Vasseur nous confie que c'est

---

[401] *Ibid.*, 00.37.31s.
[402] *Ibid.*, 00.38.53s
[403] Entretien du 25 mars 2021 avec Jean-Séverin Lair, annexe 1, 00.22.29s.
[404] *Ibid.*
[405] *Ibid.*, 00.24.14s
[406] *Ibid.*, 00.47.55s
[407] *Ibid.*, 00.48.46s
[408] Entretien du 26 mars 2021 avec Edouard Vasseur, annexe 1, 00.11.55s.
[409] Entretien du 25 mars 2021 avec Jean-Séverin Lair, annexe 1, 00.10.11s.
[410] Entretien du 26 mars 2021 avec Edouard Vasseur, annexe 1, 00.04.44s

là le retour d'expérience qu'il fait auprès de ses élèves de Master 2 : « C'est que oui, on apprend parce qu'on a l'image de l'informaticien, comme de l'archiviste, mais qu'en fait il y a cinquante métiers »[411]. Ainsi, il n'est pas question de dire ici que le web de données ne soit le seul domaine des informaticiens, mais bien de démontrer les avantages que peut présenter une telle association ; face à « l'instabilité du document numérique […] Il s'agit de mettre en place […] une mutualisation de nos réflexions, notamment en mêlant différentes professions et techniques »[412].

[411] *Ibid.*, 00.10.29s.
[412] Olivier Le Deuff, *La documentation dans le numérique*, Villeurbanne, Presses de l'Enssib, 2014, p. 94.

2. De la théorie à la pratique, première expérimentation de RiC à l'échelle nationale

En janvier 2014, le ministère de la Culture dévoile sa feuille de route stratégique intitulée « Métadonnées culturelles et transition Web 3.0 », un document annonçant de nombreux axes de travail à exploiter afin d'inscrire le secteur culturel dans le web sémantique. Le quatrième de ces axes prévoit ainsi d'expérimenter l'interconnexion inter-institutionnelle des données culturelles. Pour ce faire, le Ministère de la Culture suggère l'élaboration d'une preuve de concept afin de sensibiliser les institutions à cet enjeu[413] et en précise la portée : « L'approche "preuve de concept" est une démarche orientée utilisateur final. Elle permet de mettre en œuvre, à titre exploratoire, de nouveaux types de fonctionnalités et d'obtenir un premier retour d'expérience sur le plan des outils, des usages et des pratiques culturelles ». Afin de répondre à cet appel à projets, les AN, le *Service interministériel des Archives de France* (SIAF) et la BnF, accompagnés d'un chercheur canadien — Emmanuel Château-Dutier — du Centre de recherche interuniversitaire en humanités numériques (CRIHN) de l'Université de Montréal, se sont associés dans cette entreprise[414]. Un projet officiellement lancé en 2015 et piloté par les AN ayant pour finalité la création d'une preuve de concept qui puisse représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques[415]. Quels sont les enjeux inhérents à cette preuve de concept ? Quels en ont été les travaux préparatoires ? Quelles sont les résultats de cette expérimentation du RiC à l'échelle nationale ? Enfin, quels sont les projets en matière de métadonnées archivistiques qui ont suivi la preuve de concept établie ?

2.1. Le PIAAF : expérimenter l'interconnexion inter-institutionnelle des données culturelles

2.1.1. Plusieurs acteurs, de multiples enjeux

Le choix des institutions impliquées dans ce projet n'est pas le fruit du hasard. En effet, toutes ont en commun le fait de présenter des enjeux institutionnels propres étroitement liés aux objectifs de la feuille de route du ministère de la Culture. Pour chacune d'entre elles,

> *« Les résultats du projet nourriront leur réflexion sur une possible opération de sémantisation des métadonnées archivistiques, sur l'enrichissement et*

---

[413] Ministère de la Culture, *Feuille de route stratégique, Métadonnées culturelles et transition Web 3.0*, 2014, p. 21, [en ligne], disponible sur https://www.culture.gouv.fr/var/culture/storage/pub/feuille_de_route__metadonnees_culturelles_et_transition_web_3_0_janvier_2014/files/docs/all.pdf (consulté le 18 mai 2021).

[414] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *La Gazette des archives*, n°245, 2017, p. 157-158.

[415] *Ibid.*, p. 158.

CC BY NC ND

l'alignement des données obtenues et sur la construction de services, notamment de visualisation et de navigation, autour de ces données.» [416]

Pour le SIAF tout d'abord — dont le rôle est de veiller à la bonne application des normes scientifiques et techniques qui régissent l'univers de l'archiviste — la preuve de concept est avant tout « un outil à vocation pédagogique dans la diffusion des modèles, méthodes et pratiques utilisés »[417]. Cette expérience a également fourni au SIAF de la matière à penser pour mener à bien un projet mené depuis 2010 avec l'*Association des archivistes français* (AAF). Il s'agit d'un projet de création d'un référentiel national commun de notices d'autorité en XML/EAC-CPF ; ce dernier pourrait alors être converti en RDF, visualisé sous forme de graphe et relié à celui des AN[418]. Cette preuve de concept a pu par ailleurs également suggérer des pistes de réflexions pour l'évolution du portail *FranceArchives.fr*[419].

Pour les AN, l'enjeu d'interconnecter les jeux de métadonnées est d'autant plus nécessaire que l'institution constate — suite à une analyse du référentiel des producteurs menée la même année — les nombreuses relations existantes entre ses notices d'autorité elles-mêmes et entre ses notices d'autorités et ses instruments de recherche[420]. Une conclusion non sans influence sur l'évolution des métadonnées archivistiques : « C'est ce constat qui nous a conduit à envisager pour la première fois le graphe comme modèle de représentation des métadonnées des Archives nationales »[421]. De plus, les AN reconnaissent la nécessité d'attribuer un identifiant ISNI aux producteurs décrits afin d'éviter toute méprise sur leur identité[422]. Ces identifiants sont « des clés de liage importantes pour par exemple aligner nos notices avec celles de la BnF (si une de nos notices stocke un id ISNI et qu'une notice de la BnF stocke le même ISNI, alors les deux notices peuvent a priori être considérées comme décrivant la même entité) »[423]. Des tests d'alignement entre les notices d'autorité des AN et de la BnF montrent des résultats prometteurs[424]. Aussi pour l'institution, cette preuve de concept est l'opportunité de repenser en profondeur son système de métadonnées[425].

---

[416] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit.*, p. 159.
[417] *Ibid.*
[418] *Ibid.*, p. 159-160.
[419] *Ibid.*, p. 160.
[420] Florence Clavaud, « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *Journal of Archival Organization*, vol. 12, n°1-2, 2015, p. 123.
[421] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *La Gazette des archives*, n°254, 2019, p. 69.
[422] Florence Clavaud, « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *art. cit.*, p. 126.
[423] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°12.
[424] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit*, p. 161.
[425] Florence Clavaud, « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *art. cit.*, p. 139.

Du côté de la BnF qui ne dispose pas encore d'outils pour la description archivistique des producteurs[426], cet appel à projet va permettre de mettre en lumière les notices d'autorité des entités productrices d'archives internes réalisées dans le cadre du projet de recherche « Bibliothèque 1368-2015 »[427]. De plus, il va permettre de nourrir la réflexion du projet *Fichier national d'entités* (FNE), un projet impulsé par la BnF et l'*Agence bibliographique de l'enseignement supérieur* (ABES) réalisé dans le cadre du programme national *Transition bibliographique*, qui vise à respecter les principes FAIR — Findable, Accessible, Interoperable, Reusable — pour la production de métadonnées.

Enfin, le rôle d'Emmanuel Château-Dutier est également capital pour comprendre l'apport de la sémantisation des données et de leur visualisation sous forme de graphe du point de vue de l'utilisateur final[428].

2.1.2. Représenter en RDF : identifier ses besoins et s'approprier RiC-O

Le choix d'utiliser RiC-O pour expérimenter l'interconnexion inter-institutionnelle des données culturelles s'est imposé naturellement puisque « Les Archives nationales et le SIAF sont fortement impliqués dans les travaux du groupe EGAD, ce qui permet d'établir facilement le pont entre les deux projets »[429]. Aussi, le PIAAF — réalisé dans le courant de l'année 2017 et rendu public l'année suivante[430] — a nourri la réflexion sur RiC :

> *« Je me suis servie de ce que l'on a fait aux Archives nationales depuis 2016 pour travailler sur l'ontologie. Les problèmes que j'ai rencontrés, les besoins qu'on a eu à traiter nous ont servi pour développer une ontologie répondant à ces besoins. »*[431]

Il est vrai que le projet PIAAF s'est avéré utile pour peaufiner la partie agent de RiC, notamment les relations inhérentes à cette entité[432]. Toute la difficulté du projet réside aussi dans les limites à fixer ainsi que l'illustre Claire Sibille de Grimoüard : « ça peut être vertigineux les relations, on peut faire des relations avec tout. […] D'accord on veut faire des relations mais entre quoi et quoi ? Et quel type de relation ? »[433]. Par ailleurs, afin de répondre aux besoins susmentionnés, seule une partie de l'ontologie a été utilisée ; ainsi n'a pas été utilisée l'entité *Rule* par exemple[434]. En revanche, l'ontologie a été adaptée — nous dirons étendue — aux besoins de la preuve de concept et « La conformité à RiC se mesure[…] au fait qu'on aura étendu correctement

---

[426] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit*, p. 162.
[427] *Ibid.*
[428] *Ibid.*, p. 163.
[429] *Ibid.*, p. 164-165.
[430] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.41.03s.
[431] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.39.21s.
[432] *Ibid.*, 00.40.20s.
[433] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.45.59s.
[434] *Ibid.*, 00.45.08s.

le modèle en définissant des sous-propriétés ou des sous-classes et pas en ajoutant des choses sans les raccrocher au reste »[435].

2.1.3. Les enjeux de la *data visualisation*

La *data visualisation* — ou visualisation des données — est le deuxième point sur lequel devaient se pencher les institutions partenaires. En effet, bien que l'introduction progressive de RiC va permettre de surmonter peu à peu les problèmes liés à la qualité des données produites, il reste encore à proposer un logiciel gratuit et *open source* pour pallier les problématiques actuelles des initiatives — bien que marginales — en matière de *data visualisation* : utilisation de plateformes dont le stockage des données se situe en dehors du territoire français — ce que la note d'information DGP/SIAF/2016/006 interdit formellement — conjuguée aux recours d'adresses email personnelles pour le traitement des données[436] : « Cela signifie, […] que : les archivistes doivent utiliser un compte personnel et que des instruments de recherche, voire des documents, sont hébergés sur des serveurs Google (Alphabet) »[437].

Aussi, l'objectif de la preuve de concept est donc d'offrir « le moyen de parcourir les graphes ou diagrammes depuis plusieurs points de départ différents, ou selon diverses perspectives, sans sacrifier la granularité informationnelle ni la lisibilité »[438]. Un objectif rempli, ainsi que le rapporte Claire Sibille de Grimoüard : « [Le PIAAF présente] un outil de *data visualisation* qui permet de visualiser de manière dynamique les relations »[439].

Une question survient alors : tous les services d'archives sont-ils concernés par la *data visualisation* ? Sont-ils tous amenés à établir une politique institutionnelle en la matière ? A ce titre, il nous semble important de mentionner la nécessité de procéder à l'examen « du type d'archives que l'on conserve, des besoins mesurés ou supposés de leurs usagers, mais aussi des moyens humains et techniques qu'il est possible de mobiliser »[440] avant d'envisager tout projet de *data visualisation*.

2.1.4. PIAAF : organisation et méthodologie

Afin d'élaborer la preuve de concept attendue, les institutions partenaires ont tout d'abord dû procéder à la sélection de jeux de métadonnées sur lesquels appuyer leur travail. Ces derniers

---

[435] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.49.21s.
[436] Julien Pomart, « La visualisation des données archivistiques : les premiers pas, et ensuite ? », *La Gazette des archives*, n°248, 2017, p. 110.
[437] *Ibid.*
[438] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit*, p. 166.
[439] Entretien du 5 mars 2021 avec Claire Sibille de Grimoüard, annexe 1, 00.45.59s.
[440] Julien Pomart, « La visualisation des données archivistiques : les premiers pas, et ensuite ? », *art. cit.*, p. 106.

devaient évidemment s'articuler entre eux et ne pas excéder un certain volume[441]. Il était important que lors de ce processus de sélection et de préparation des métadonnées, chacune des institutions conserve ses outils de travail, applique ses règles et préserve sa propre perspective ; cependant, pour le bien du projet, toutes se sont entendues sur les formats d'encodage — EAD et EAC —, les formes autorisées des noms d'entités ainsi que la typologie des relations à adopter[442]. La description des fonctions a quant à elle conduit à l'élaboration de deux vocabulaires communs, « un pour les domaines fonctionnels ou objets, l'autre pour les activités ou actions »[443] ; tous deux élaborés avec GINCO — *Gestion Informatisée de Nomenclatures Collaboratives et Ouvertes* —, un outil proposé par le ministère de la Culture depuis 2013. Des vocabulaires dont l'élaboration « a conduit à échanger longuement sur les concepts à retenir »[444].

Après analyse de ces jeux de données, le verdict tombe : des chevauchements significatifs se dessinent et une complémentarité se distingue très clairement ; l'hypothèse d'une interconnexion possible est donc confirmée[445]. Aussi, avant de passer à l'étape suivante et convertir ces jeux de données en RDF, le groupe de travail relève des points qui doivent faire l'objet d'une attention particulière. Parmi eux, la représentation du temps pourrait s'avérer problématique :

> *« Par exemple, telle administration décrite dans un référentiel de bibliothèque sera considérée en tant qu'auteur sur la longue durée et décrite très rapidement, tandis qu'un service d'archives sera beaucoup plus sensible à l'évolution de ses fonctions et de son environnement administratif, déterminants pour comprendre sa production documentaire. »*[446]

A cela s'ajoute la nécessité, pour les données RDF produites, de conserver des informations relatives à leur provenance ainsi que le stipule la feuille de route stratégique du ministère[447]. Enfin, l'équipe de travail doit encore définir comment les catégories de collectivités territoriales décrites par le SIAF vont être modélisées au sein de l'ontologie RiC-O[448].

Dernière étape — et non des moindres — : l'élaboration d'un logiciel de *data visualisation*. Ce dernier étant dévolu à un prestataire extérieur, il a par conséquent fait l'objet d'un cahier des charges. Parmi les caractéristiques clés de ce cahier des charges figurent : la volonté d'une forte

---

[441] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit*, p. 167.
[442] *Ibid.*, p. 168.
[443] *Ibid.*
[444] *Ibid.*
[445] *Ibid.*, p. 168-169.
[446] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°7.
[447] Ministère de la Culture, *Feuille de route stratégique, Métadonnées culturelles et transition Web 3.0*, *op. cit.*, p. 21.
[448] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit*, p. 169.

CC BY NC ND

interaction avec le titulaire du marché, la réalisation d'une analyse critique des jeux de données RDF par le prestataire retenu ainsi que la présentation de la méthode qu'il envisage d'employer et la participation des différents publics aux tests du logiciel[449]. Enfin, en plus de présenter une interface qui permette l'alignement des données, leur interrogation et leur exploration intuitive, le logiciel doit pouvoir offrir la possibilité d'ajouter des fonctionnalités supplémentaires à celles initialement prévues par la feuille de route stratégique du ministère[450].

Le résultat de cette collaboration s'avère être une réussite — puisque répondant en tous points aux enjeux énoncés par le ministère de la Culture — et invite les Archives nationales à poursuivre les efforts en la matière.

2.2. Un convertisseur et un composant d'exploration open-source pour les métadonnées archivistiques

2.2.1. RiC-O Converter : un logiciel pour la conversion des métadonnées archivistiques

Le choix de développer un convertisseur de métadonnées archivistiques EAD/EAC en métadonnées RDF conformes à RiC-O résulte du nombre important de notices d'autorité et d'instruments de recherche présents aux AN. Ainsi, *RiC-O Converter* se présente comme « un outil de conversion industrielle » pour faire face aux 15000 notices d'autorité EAC/CPF et aux 28000 instruments de recherche EAD à convertir[451]. En 2018, la société Sparna dirigée par Thomas Francart devient titulaire du marché, et les travaux d'élaboration du convertisseur débutent l'année suivante. Comment ces derniers se déroulent-ils ?

A ce sujet, Thomas Francart — expert en modélisation des connaissances et en gestion des ontologies — nous informe que le travail commence dès la réalisation de la réponse au cahier des charges[452] : « J'avais proposé une architecture logicielle pour la réalisation du convertisseur avec des choix technologiques. Le cahier des charges incitait fortement à utiliser des feuilles de style XSLT parce que c'est une technologie qui est assez répandue dans les archives »[453]. Les feuilles de style XSLT — *eXtensible Stylesheet Language Transformations* — permettent en effet de définir la mise en page d'un document XML, de transformer ce dernier en un autre format. Une fois cela fait, le projet est entré dans une phase de spécification :

---

[449] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit*, p. 169-170.
[450] *Ibid.*, p. 170.
[451] Thomas Francart, Pauline Charbonnier, *RIC-O converter, un logiciel libre de conversion de métadonnées archivistiques (en EAD et EAC-CPF) en jeux de données conformes à RiC-O*, actes de la journée d'étude organisée le 28 janvier 2020 par les Archives nationales à Paris, p. 5, [en ligne], disponible sur https://f.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_4_RiCOConverter.pdf (consulté le 18 janvier 2021).
[452] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.03.41s.
[453] *Ibid.*

« On avait fait un premier tableau de mise en correspondance des éléments EAC avec les éléments de l'ontologie RiC-O. Mais par la suite, […] pour être vraiment sûrs de la qualité de ce que produit le convertisseur […] on s'était dit qu'il fallait passer par la construction de tests unitaires. »[454]

Ces tests consistent à déterminer les résultats du convertisseur sur des petits fichiers EAC et EAD. Thomas Francart précise que c'est de cette manière qu'ont été codées les spécifications très formelles ; cela a permis de s'assurer que le résultat final réponde aux attentes exprimées[455]. Aussi lors de la journée d'étude organisée le 28 janvier 2020 aux AN, le projet comptait environ 90 fichiers de tests, chacun faisant état de plusieurs cas de tests : moitié EAC, moitié EAD[456]. A ce jour, la totalité des notices d'autorité et des instruments de recherche des AN sont convertis en un format RDF conforme à RiC-O. La procédure de conversion elle-même a généré environ 0,7 millions de triplets pour les fichiers EAC et environ 155 millions de triplets pour les fichiers EAD[457] ; autant de triplets qui une fois stockés dans un triplestore — une base de données pour les fichiers RDF — et interrogés par un langage de requête pourront permettre l'émergence d'entités conceptuelles implicites dans les fichiers. En effet, Thomas Francart souligne ici une différence de points de vue adoptés :

« RiC-O nous dit : il y a un seul record, c'est la notice plutôt abstraite, et deux représentations de cette même notice : l'objet physique et son pendant numérique. L'EAD ne dit pas du tout ça. L'EAD dit : j'ai qu'une seule entrée dans mon EAD, qui doit correspondre à la balise <c>, et dans cette entrée j'ai un attribut qui s'appelle <daogrp> qui va pointer vers l'adresse d'une numérisation pour cette notice. »[458]

Par ailleurs, le convertisseur ne présente pas à ce jour la possibilité de convertir des référentiels d'indexation en RDF/SKOS. En effet, outre des raisons budgétaires liées à la conversion d'un format supplémentaire, ces référentiels d'indexation relèvent « [d'] une DTD maison, conforme ni à RiC-O, ni à SKOS, ni à la norme sur les thésaurus ; […] c'est une DTD qui va tomber dans quelque temps, on va avoir besoin de la remplacer »[459]. On comprend alors l'inutilité d'inclure au projet RiC-O Converter une DTD qui n'a pas pour vocation d'être pérenne[460]. Aussi, ces référentiels d'indexation sont entièrement ressaisis pour produire directement du SKOS : « il nous arrive d'utiliser la plateforme Ginco du ministère, pour produire

---

[454] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.03.41s.
[455] *Ibid.*
[456] Thomas Francart, Pauline Charbonnier, *RIC-O converter, un logiciel libre de conversion de métadonnées archivistiques (en EAD et EAC-CPF) en jeux de données conformes à RiC-O*, *op. cit.*, p. 9.
[457] *Ibid.*, p. 14.
[458] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.19.58s.
[459] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 01.03.28s.
[460] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 01.03.28s.

un thésaurus. Nous pouvons ensuite exporter au format SKOS ce thésaurus, et l'enrichir d'autres données. Il reste alors à réimporter dans le SIA les données en question »[461].

Parmi les difficultés rencontrées lors du développement de RiC-O Converter, la plus importante à citer est celle des URI : comment assigner des URI à des entités qui en sont dépourvues ? Ainsi que nous l'avons vu précédemment, le langage RDF repose en effet sur l'assignation d'URI/d'IRI ; c'est là le fondement-même du web sémantique. Deux cas de figure se présentent alors : tout d'abord l'entité d'origine n'a pas d'URI, mais contient tout de même un identifiant — un ID, pour *IDentifier* —, ce qui est le cas des fichiers EAD et EAC des AN[462]. Dans ce cas précis, la solution est de générer une URI en concatenant une racine d'URI à laquelle on ajoute l'ID du fichier à la fin[463]. Dans le deuxième cas, il y a une absence d'ID. C'était le cas aux AN pour les relations entre entités ; aussi pour y remédier, le choix a été fait de faire reposer l'unicité de l'entité — la relation — sur l'ensemble de ses attributs : « il ne peut pas y avoir deux relations du même type à la même date entre les deux mêmes entités. Et donc en prenant ces critères-là, et en les insérant tous dans l'URI qu'on génère, on garantit de générer une URI unique »[464].

Un autre défi à relever réside dans le fait que les URI ainsi générées ne sont à l'heure actuelle pas encore déréférençables — c'est-à-dire qu'elles n'amènent à aucun document, aucune représentation de ces URI — ; aussi, si nous prenons aujourd'hui une URI générée par RiC-O Converter et qu'on l'insère dans notre barre d'adresse, une erreur 404 apparaîtra car le serveur ne connaît l'URI qu'on lui a transmis[465]. A ce sujet, Florence Clavaud précise que :

> *« Globalement les AN doivent encore consolider et étendre leur stratégie d'assignation d'URI ; par ailleurs plus concrètement cela impliquera sans aucun doute de doter l'institution d'un dispositif robuste et flexible, fondé probablement sur ARK, et capable de déréférencer convenablement les URI selon leur contexte d'utilisation comme l'a fait la BnF. »*[466]

Un défi qu'il est indispensable de relever dans le cadre d'une application sémantique visant à publier les jeux de données RDF des AN[467].

2.2.2. Sparnatural : un composant d'exploration basé sur SPARQL

La quête vers le web sémantique se poursuit, avec la création d'un prototype — Sparnatural — dont l'objectif est de permettre la formulation de requête en SPARQL —

---

[461] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°25.
[462] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.09.52s.
[463] *Ibid.*
[464] *Ibid.*
[465] *Ibid.*, 00.13.51s.
[466] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°11.
[467] Thomas Francart, Pauline Charbonnier, *RIC-O converter, un logiciel libre de conversion de métadonnées archivistiques (en EAD et EAC-CPF) en jeux de données conformes à RiC-O*, *op. cit.*, p. 23.

SPARQL Protocol and RDF Query Language — un langage indispensable au web sémantique ainsi que le souligne Tim Berners-Lee : « Trying to use the Semantic Web without SPARQL is like trying to use a relational database without SQL. SPARQL makes it possible to query information from databases and other diverse sources in the wild, across the Web »[468]. Ce composant n'a par ailleurs pas été créé spécifiquement pour l'occasion. En effet, Sparnatural a vu le jour dans le cadre d'un projet — OpenArcheo — mené avec des archéologues[469]. Pour créer ce composant, Thomas Francart s'est inspiré du prototype *ResearchSpace* du British Museum, une plateforme d'interrogation et de visualisation de graphe[470]. Aussi le graphe sur lequel le composant devait se déployer initialement ne répondait pas du tout à RiC-O mais à l'ontologie CIDOC-CRM ; ce n'est qu'après qu'il a été configuré pour naviguer dans une sous-partie d'un graphe RiC-O[471]. Mais comment fonctionne-t-il exactement ? Quels sont les préalables à son utilisation ?

A ce sujet, Thomas Francart nous précise que Sparnatural en lui-même ne nécessite pas d'installation puisqu'il s'agit d'un composant Javascript technique dont la seule chose nécessaire à son fonctionnement est une page HTML combinée à une feuille de style CSS — Cascading Style Sheets — ; cependant, les requêtes formulées par Sparnatural doivent atteindre un triplestore pour qu'apparaissent des résultats[472]. Actuellement, Florence Clavaud utilise le triplestore GraphDB sur lequel elle a chargé les données RiC-O issues du convertisseur[473] et le composant Sparnatural en local[474] afin de poursuivre les tests. Cependant, de tels changements impliquent nécessairement de revoir l'architecture-même du Système d'information archivistique (SIA) des AN :

> *« On a des données RDF en très grandes quantités […] mais on n'a pas implémenté encore dans notre système d'information la brique qu'il faut pour les stocker, les rendre cherchables, et les consulter. Ce n'est pas seulement un problème de visualisation, c'est une infrastructure qu'il faut changer. Il faut y rajouter des modules. »*[475]

Aussi, bien que l'ajout de Sparnatural au SIA des AN constitue à ce jour un objectif de long terme[476], des réflexions relatives à la plateforme de visualisation sont déjà menées[477]. Plusieurs paramètres sont à prendre en compte lors de son élaboration, notamment pour les

---

[468] World Wide Web Consortium, *W3C Opens Data on the Web with SPARQL, Powerful Technology for Querying Distributed and Diverse Data*, W3C, 2008, [en ligne], disponible sur https://www.w3.org/2007/12/sparql-pressrelease.html.en (consulté le 20 mai 2021).
[469] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.53.34s.
[470] *Ibid.*, 00.51.21s.
[471] *Ibid.*, 00.42.53s.
[472] *Ibid.*, 00.43.50s.
[473] *Ibid.*, 00.54.57s.
[474] *Ibid.*, 00.43.50s.
[475] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.32.16s.
[476] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.45.36s.
[477] *Ibid.*, 00.48.20s.

outils de partitionnement de graphe qui vont être proposés à l'utilisateur final. Parmi eux figurent : la capacité du triplestore lui-même, les algorithmes que l'on peut lui ajouter, et les fonctionnalités que l'on souhaite présenter à l'utilisateur final[478]. A ce sujet, Thomas Francart précise que, bien que le choix de ces fonctionnalités appartienne évidemment aux institutions elles-mêmes, il s'agit avant-tout d'un échange entre deux professions :

> *« C'est à la fois un dialogue entre les capacités de la technique d'un côté et les attentes des métiers de l'autre côté. Il faut à la fois être capable de montrer ce qu'est capable de faire la technique et à quels besoins ça peut répondre. D'exprimer des besoins métier et de trouver comment y répondre du côté technique. »*[479]

Des propos qui font écho à ceux d'Edouard Vasseur cités précédemment.

Par ailleurs, il faut également composer avec des difficultés d'ordre technique, telles que le temps de réponse des requêtes effectuées : « comme on va être amené à naviguer dans le graphe de façon assez complexe, de générer des requêtes assez complexes, on peut avoir des questions de temps de réponse derrière »[480] ; un écueil qui n'est pas sans rappeler celui de la gestion de la volumétrie par les technologies actuelles évoqué par Jean-Séverin Lair.

Un modèle conceptuel, des données en RDF, un composant en cours de construction pour les requêter… tous les facteurs semblent réunis pour que les AN prennent le large pour le web 3.0 ; « semblent » car il manque en effet un élément majeur à l'équipage : les référentiels. Quels enjeux ces derniers revêtent-ils ? Comment le passage au web sémantique redéfinit-il leurs rôles ? Quels changements cela implique-t-il de réaliser ?

2.3. Des référentiels en mutation

Les référentiels sont des outils qui doivent permettre à l'utilisateur final d'accéder à des résultats de recherche plus clairs, présentant moins de bruit et moins de silence[481]. Un rôle essentiel donc, mais peu reconnu jusqu'alors :

> *« Un des grands non-dits dans le domaine des archives est celui des référentiels, des données de référence, parce qu'on a besoin en effet de ces données de référence et c'est un sujet dont on ne parle pas assez, dont on ne traite pas assez […] Le référentiel, une fois qu'il existe, on peut le réutiliser partout. Il y a des points de concentration de ces référentiels à faire. »*[482]

---

[478] *Ibid.*, 00.54.57s.
[479] *Ibid.*
[480] *Ibid.*, 00.48.20s.
[481] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°5.
[482] Entretien du 26 mars 2021 avec Edouard Vasseur, annexe 1, 00.40.56s.

Mais comment se présentent les référentiels des AN à ce jour ? Quels sont les projets en cours et à venir les concernant ? Un élément à savoir premièrement est que ces référentiels sont au nombre de dix-sept aux AN et sont de deux types : des notices d'autorité d'un côté et des vocabulaires de l'autre[483]. La volonté d'adapter RiC au futur SIA des AN[484] suscite l'intérêt : comment s'organise la transition de ces référentiels vers le web sémantique ?

2.3.1. Passage d'un référentiel des producteurs à un référentiel des agents

Le déploiement d'un nouveau SIA en 2013 se réalise avec la volonté de créer une base de données dédiée aux producteurs[485]. Un module lui est donc destiné, mais les AN y voient se dessiner trois sous-ensembles de fichiers EAC[486]. En effet, un premier sous-ensemble contient 3430 notices d'autorité issues de la base de données Etanot[487] — la base de données des notaires parisiens —, un second comprend environ 10 000 fichiers EAC provenant de la migration de fichiers EAC issus d'une base de données utilisée par les AN pour la gestion des archives gouvernementales de 1969 à 2009[488], et un dernier fait référence aux 570 notices d'autorité créées depuis le déploiement du SIA[489]. De plus, l'institution constate la présence d'autres notices d'autorité en dehors de celui-ci. Tout d'abord avec les quelque 20 000 instruments de recherche encodés en EAD au sein de la balise <bioghist> de l'EAD, mais également au cœur du Plan d'orientation général (POG) avec la présence de plus de 1200 notices d'autorité[490]. A cela s'ajoutent encore deux listes d'autorité décrivant des personnes physiques et des personnes morales, utilisées pour l'indexation des noms de ces notices d'autorité[491].

Aussi, afin de remédier à cette situation, les AN mettent en place dès fin 2013 une description systématique des producteurs : « Because the module is defined as the knowledge base […] there must be no record on archival creators outside the module, neither in any finding aid nor in any component of the Plan d'orientation général ».[492] Pour mener à bien ce projet, cinq actions sont identifiées : retirer des fichiers EAD et du POG les informations relatives aux producteurs et créer à la place des fichiers EAC au sein du module ; poursuivre le travail d'enrichissement des notices d'autorité issues d'Etanot et de celles provenant de la base de données dédiée à la gestion des archives gouvernementales de 1969 à 2009 ; créer des fichiers EAC pour les collectivités dont les archives sont entrées aux AN avant 1970 ; et enfin tenir à

---

[483] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 01.03.28s.
[484] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *art. cit.*, p. 73.
[485] Florence Clavaud, « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *art. cit.*, p. 125.
[486] *Ibid.*, p. 123.
[487] *Ibid.*, p. 121.
[488] *Ibid.*, p. 122.
[489] *Ibid.*, p. 123.
[490] *Ibid.*, p. 124.
[491] *Ibid.*
[492] *Ibid.*, p. 126.

jour l'ensemble de ces notices d'autorité[493]. Un projet qui s'accélère depuis 2019, grâce à l'implication des agents scientifiques de la direction des fonds, du département de l'archivage électronique et des archives audiovisuelles, ainsi que par l'élaboration de principes et d'une méthodologie en la matière[494].

Un cap est encore franchi début 2018 avec la décision de passer à un référentiel des agents, soit à un référentiel qui conjugue le référentiel des producteurs, et les référentiels des personnes physiques et des personnes morales ; un référentiel « plus riche, plus visible, plus facile à interroger […] qui intègre la description de n'importe quelle entité de type agent (au sens RiC), quelles que soient ses relations avec les documents conservés aux AN »[495]. Ainsi, faciliter l'accès aux métadonnées de description des documents au sein de la salle des inventaires virtuelle (SIV) et permettre l'interopérabilité[496], tels sont les enjeux clés présentés par le référentiel des agents.

2.3.2. Les référentiels d'indexation liés au référentiel des agents : des référentiels de concepts

Les évolutions qui entourent le référentiel des agents ne sont pas sans influer sur les référentiels d'indexation qui gravitent autour de ce dernier. Un constat déjà opéré lors du projet de description automatique des producteurs :

> *« creating an authority record knowledge base for archival creators, in the context and in the temporal coverage of the NAF, will probably imply also developing the functions vocabulary and create rich records for identifying and describing them, creating relationships between them and between them and persons, corporate bodies, and records. »*[497]

a) Le référentiel des fonctions

Aussi, des travaux ont été entrepris, notamment pour le référentiel des fonctions des services d'administration centrale et des organismes divers de l'État. Un référentiel qui envisage de passer d'une liste d'autorité plate et incomplète de 146 descripteurs peu définis à un

---

[493] Florence Clavaud, « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *art. cit.*, p. 127.
[494] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *art. cit.,* p. 84.
[495] Florence Clavaud, Pauline Charbonnier, *Records in Contexts aux Archives nationales : enjeux et premières réalisations*, actes de la journée d'étude organisée le 28 janvier 2020 par les Archives nationales à Paris, p. 31, [en ligne], disponible sur https://f-origin.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_3_RiCauxAN_EnjeuxPremieresRealisations.pdf (consulté le 18 janvier 2021).
[496] Florence Clavaud, *Les référentiels aux Archives nationales*, actes de la rencontre annuelle des correspondants archives des opérateurs et grands corps de l'Etat du 12 octobre 2016, p. 4, [en ligne], disponible sur https://www.francearchives.fr/file/20b2e2aa9a1f62b2b843239beff36a29468db1f3/static_9370.pdf (consulté le 18 janvier 2021).
[497] Florence Clavaud, « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *art. cit.*, p. 129.

thésaurus SKOS et RiC-O — utilisation de la classe ActivityType —[498] ; une évolution qui n'est pas sans rappeler celle du référentiel des types de documents, lui-aussi devenu un thésaurus SKOS et RiC-O — utilisation de la classe DocumentaryFormType —[499].

Dans un second temps sera étudiée l'élaboration d'un même référentiel appliqué aux fonctions des collectivités territoriales[500].

b) D'un référentiel des lieux à un référentiel d'entités géo-historiques

Le référentiel des lieux a également droit au chapitre : fin 2018, il voit débuté un chantier visant à le faire passer de liste d'autorité plate et peu définie à un référentiel d'entités géo-historiques[501] : « Notre référentiel des lieux tel que nous l'avons dans le système, on l'a converti en RiC-O. Les lieux qui étaient définis par de toutes petites notices, sont devenus des instances de la classe place de RiC-O »[502]. Un résultat possible, grâce aux métadonnées RDF de l'INSEE — Institut national de la statistique et des études économiques — et de l'IGN — Institut national de l'information géographique et forestière — que les AN ont importées dans un triplestore et dont « le résultat des requêtes SPARQL a ensuite été intégré dans les données des AN »[503]. Parmi les objectifs futurs figure celui d'assigner des informations de géoréférencement aux 13 000 entrées du référentiel relatives aux voies parisiennes, un projet qui permettra « de projeter le tracé de ces voies sur des plans à des dates différentes »[504].

c) De nouveaux référentiels à venir ?

La voie vers la création de nouveaux référentiels semble inévitable ainsi que le relève Florence Clavaud : « Notons que la logique de construction d'un graphe multidimensionnel est implacable : elle nous conduit à envisager d'autres référentiels »[505]. Des référentiels qui pourront faire émerger de nouvelles entités « comme les textes de référence, à considérer comme des instances de la classe Mandate »[506], ou faire office de référentiels d'indexation pour décrire d'autres entités telles que des types de lieux ou de langues par exemple[507].

---

[498] Florence Clavaud, Pauline Charbonnier, *Records in Contexts aux Archives nationales : enjeux et premières réalisations*, op. cit., p. 28-29.
[499] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 01.06.33s.
[500] Florence Clavaud, Pauline Charbonnier, *Records in Contexts aux Archives nationales : enjeux et premières réalisations*, op. cit., p. 28.
[501] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *art. cit.,* p. 86.
[502] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 01.06.33s.
[503] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *art. cit.,* p. 86.
[504] *Ibid.*, p. 86-87.
[505] *Ibid.*, p. 87.
[506] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *art. cit.,* p. 87.
[507] *Ibid.*

2.3.3. Favoriser l'intéropérabilité des données culturelles

Convertir des métadonnées est une chose, les rendre interopérables en est une autre. Mais pourquoi l'interopérabilité est-elle si importante ? A ce sujet, Jacob Holmblad, ancien vice-président — gestion technique — de l'ISO, nous apporte une réponse imagée : « Le récit biblique de la Tour de Babel, où la multiplication des langues a empêché l'achèvement de la construction qui devait toucher le ciel, montre bien ce qui se produit lorsque, faute de code commun, l'interopérabilité échoue : le progrès ne peut plus avancer »[508]. Si l'interopérabilité est donc la capacité pour des systèmes d'information différents de fonctionner ensemble, d'arriver à se comprendre malgré un « langage » initial différent, comment la mettre en place ? Comment pouvons l'atteindre ? Par quels procédés ?

L'interopérabilité s'opère par l'alignement de concepts d'une institution à une autre. Cela peut se faire manuellement dans le cas de référentiels peu conséquents ou à l'aide d'outils tel qu'OpenRefine, bien que l'intervention humaine soit toujours nécessaire, « rien n'est totalement automatisable »[509]. Mais cet alignement de concepts peut s'entendre à plusieurs niveaux : comment fait-on pour aligner des modèles de données — tels que l'ontologie RiC-O avec l'ontologie FRBR — ? Et comment fait-on pour aligner des référentiels SKOS ?[510] Dans le premier cas, il est question de mettre en relation des personnes en articulant leurs URI par des liens tels que <owl:Sameas> ou <skos:exactMatch> : « Techniquement parlant, ça revient à exprimer un triplet RDF »[511]. Dans le cas des référentiels SKOS, ce serait le même procédé, auquel s'ajoute la possibilité de nuancer la correspondance : « Ça permet de typer un peu le lien de mise en correspondance, ça peut être des liens avec une sémantique un peu plus floue »[512].

Parmi les projets d'alignement des AN figure la possibilité d'échanger des données avec le FNE lorsque celui-ci sera prêt, de même qu'apparaît la volonté de poursuivre les alignements réalisés avec Wikimédias France lors du datathon de mars 2018 :

> *« [Un datathon qui] a permis, en une journée de travail d'une dizaine de collègues et de wikipédiens, d'aligner 665 notices de personnes (sur 2491) et 136 notices de collectivités (sur 683) avec les données qui décrivent les mêmes entités dans la base de données encyclopédique et collaborative Wikidata. »*[513]

Des opérations qui prennent de l'envergure dans le courant de l'année 2018 avec le projet ALEGORIA — *Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge*

---

[508] Jacob Holmblad, « Les normes font tourner le monde », *ISO Focus+*, vol. 1, n°2, 2010, p. 1.
[509] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°29.
[510] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.30.01s.
[511] *Ibid.*, 00.30.43s.
[512] *Ibid.*, 00.31.54s.
[513] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *art. cit.*, p. 85.

— mené par l'IGN ; projet auquel les AN « fournissons depuis plus de 1 an des données RDF »[514]. L'opération a été rendue possible grâce à la réalisation d'un mapping EAD vers RiC-O ; c'est-à-dire d'un procédé de mise en correspondance entre la DTD et l'ontologie, dont Daniel Pitti détaille la réalisation :

> *« Mapping and alignment is an ongoing activity. With respect to methods, this is a matter of analyzing other conceptual models and ontologies and making a judgement with respect to the degree to which a concept in RiC is similar or the same as a concept in another model. »*[515]

Un procédé nécessaire afin d'articuler au mieux les jeux de données RiC-O et les métadonnées techniques relatives aux conditions de prises de vue[516]. Des référentiels qui, une fois alignés, ont permis de passer à l'étape de l'indexation : « l'indexation des lieux ou objets représentés par les photos se fait en utilisant des référentiels de lieux (ceux des AN et ceux de l'IGN) que nous avons alignés »[517].

Ainsi que nous pouvons le constater, RiC — bien qu'actuellement au stade bêta — suscite déjà l'intérêt. Sa première application à l'échelle nationale à travers le PIAAF, attire l'attention d'agrégateurs tel que FranceArchives.fr : « le SIAF pilote la conception et la réalisation du projet de portail national Francearchives.fr et le présent projet [PIAAF] fournira des éléments de réflexion pour les futures évolutions du portail »[518]. Florence Clavaud nous confirme par ailleurs la décision officielle de FranceArchives.fr d'adopter RiC[519]. Pour Edouard Vasseur, c'est là-même l'enjeu premier de ce dernier : « Il faut que ça réponde aux besoins de portails »[520]. Prochaine étape : le portail européen ArchivesPortalEurope.net ? Pour le moment aucune décision est arrêtée, mais la responsable a « appris avec beaucoup d'intérêt la décision de FranceArchives »[521] de passer à RiC-O. Dans ce contexte de portails, et plus spécifiquement dans le cas de FranceArchives.fr, les référentiels semblent être la clé de voûte entre les données des services d'archives départementaux et celles les AN[522]. En ce sens, les référentiels font figure « [d']éléments constitutifs du possible futur graphe de données »[523]. Aussi, bien que les

---

[514] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.36.54s.
[515] Questionnaire adressé à Daniel Pitti le 22 mars 2021, annexe 2, question n°37.
[516] Florence Clavaud, Pauline Charbonnier, *Records in Contexts aux Archives nationales : enjeux et premières réalisations*, op. cit., p. 36.
[517] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°28.
[518] Anila Angjeli, Florence Clavaud, Stéphanie Roussel, « Représenter en RDF, interconnecter et visualiser en graphe des jeux de métadonnées archivistiques de provenances multiples : un projet de prototype. », *art. cit*, p. 160.
[519] Entretien du 26 mars 2021 avec Florence Clavaud, annexe 1, 00.34.30s.
[520] Entretien du 26 mars 2021 avec Edouard Vasseur, annexe 1, 00.37.31s.
[521] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°10.
[522] Florence Clavaud, « Building a knowledge base on archival creators at the National Archives of France: Issues, Methods, and Prospects », *art. cit.*, p. 132.
[523] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations », *art. cit.,* p. 80.

référentiels permettent une meilleure appréhension des fonds et de leur contexte[524], « La compréhension des ressources archivistiques est aussi affaire d'ergonomie et de sciences cognitives, donc d'interfaces »[525]. En effet, la finalité étant de pouvoir représenter les données archivistiques sous forme de graphe, il peut s'avérer intéressant de comprendre les étapes du traitement de l'information visuelle pour orienter l'élaboration de ladite interface de visualisation : comment s'opère-t-elle ? en quelle mesure choisir la représentation des données sous forme de graphe est plus « parlante » pour l'utilisateur final ? En définitive, quels sont les enjeux revêtus par la *data visualisation* pour la valorisation des archives ?

[524] Florence Clavaud, *Les référentiels aux Archives nationales*, *op. cit.*, p. 4.
[525] Questionnaire adressé à Florence Clavaud le 29 mars 2021, annexe 2, question n°3.

CC BY NC ND

3. Représentation sous forme de graphe : un moyen d'optimiser la compréhension contextuelle ?

On entend souvent qu'une image vaut mieux que mille mots, mais comment cela s'explique-t-il d'un point de vue cognitif ? Et est-ce vrai pour la représentation sous forme de graphe ? Répondre à ces questions nous semble intéressant afin de mieux appréhender le travail de *data visualisation* de l'interface et les fonctionnalités que cette dernière devra proposer.

3.1. Graphes et réseaux de neurones, un air de déjà-vu ?

Le travail de *data visualisation* ne peut se faire sans comprendre comment s'opère la perception et le traitement de l'information visuelle. Aussi, comment appréhendons-nous le monde qui nous entoure ? A ce sujet, il convient de revenir sur les étapes cérébrales du traitement visuel, indispensable ainsi que l'illustre Guy Orban, professeur de neurophysiologie à l'université de Louvain et titulaire de la chaire européenne au Collège de France de 2006 à 2007 : « Il n'y a vision que lorsque le signal quitte le système visuel. […] Une vision isolée n'aboutirait qu'à une perception sans vécu »[526]. Pour mieux les comprendre, une brève présentation de notre système nerveux et de l'organisation fonctionnelle du cortex cérébral nous semble pertinente.

Impossible d'aborder le sujet du cerveau sans d'abord comprendre celui du système nerveux dont il est le centre de contrôle[527]. En effet, le système nerveux se compose de trois parties : le système nerveux central (SNC), le système nerveux périphérique (SNP) et le système nerveux autonome. La première partie est constituée de l'encéphale et du canal rachidien[528] ; l'encéphale — dont l'étymologie signifie « qui est dans la tête » — contient les éléments du système nerveux présents à l'intérieur de la boîte crânienne : le cerveau, le tronc cérébral, le cervelet et les nerfs crâniens. Le canal rachidien quant à lui, comprend la moelle épinière et les nerfs spinaux. La seconde partie se compose des nerfs externes au SNC qui assurent le lien entre ce dernier et les différents organes ; et la troisième partie est indépendante du SNC, car responsable des fonctions automatiques telles que la respiration ou le rythme cardiaque par exemple. La partie qui nous intéresse donc est celle concernant le SNC, au sein duquel le cerveau constitue l'objet de notre attention.

Tout d'abord, rappelons que notre cerveau se compose de deux hémisphères dont la surface est appelée « cortex ». Ce dernier se décompose par le biais de scissures — des rainures

---

[526] Guy Orban, *La vision, mission du cerveau : les trois révolutions des neurosciences*, Paris, Fayard : Collège de France, 2007, p. 56-58.
[527] Alfred David, *La vie secrète du cerveau : découvrez ses étonnants pouvoirs !*, Malakoff, Dunod, 2019, p. 14.
[528] Institut du Cerveau, *Comprendre le cerveau, son développement, son fonctionnement, sa plasticité et ses pathologies*, 2020, [en ligne], disponible sur https://institutducerveau-icm.org/fr/actualite/comprendre-le-cerveau-et-son-fonctionnement/ (consulté le 23 mai 2021).

CC BY NC ND

profondes — en quatre parties appelées « lobes » sur lesquelles nous reviendrons, mais il existe également des rainures superficielles nommées « sillons »[529]. Enfin, une fissure longitudinale sépare les deux hémisphères cérébraux. Bien que la structure du cortex des hémisphères est presque symétrique, les hémisphères se distinguent sur le plan fonctionnel[530]. En effet, une spécialisation du cortex de chaque hémisphère se dessine au regard de certaines fonctions cérébrales ; il s'agit de la latéralisation fonctionnelle des hémisphères cérébraux. Ils communiquent cependant ensemble grâce au corps calleux[531].

Enfin, avant de s'intéresser au processus de la perception et du traitement de l'information visuelle, il convient de rappeler que notre cortex se divise en quatre lobes parmi lesquels figurent le lobe frontal, le lobe pariétal, le lobe temporal ainsi que le lobe occipital. Nommer et identifier ces différents lobes est nécessaire pour mieux comprendre comment s'opère le cheminement de l'information perçue. Aussi, au sein du lobe frontal figure l'aire oculomotrice — c'est-à-dire l'aire qui contrôle les mouvements des yeux — tandis que les aires visuelles se situent au niveau du lobe occipital, soit à l'opposé de la première[532]. Mais que comprennent exactement ces aires visuelles ? Deux entités les composent : le cortex visuel primaire (V1) et le cortex visuel extrastrié composé des aires visuelles associatives (V2, V3, V4, V5) ; le premier reçoit les informations visuelles en provenance de la rétine, tandis que le second interprète lesdites informations[533]. Le schéma qu'en donne Guy Orban nous servira à illustrer nos propos :

---

[529] Dominique Hasboun, *Morphologie, Cortex : Développement*, Neuranet, [en ligne], disponible sur http://www.chups.jussieu.fr/ext/neuranat/morphologie/cortex/cortex04.html?fbclid=IwAR3FG88KDNPy6Zck7QNsj0Ox8kt0i3GcF5yLvXtDEfpG8ak_ZohERR3FdhY (consulté le 23 mai 2021).
[530] Institut du Cerveau, *Comprendre le cerveau, son développement, son fonctionnement, sa plasticité et ses pathologies, art. cit.*
[531] Lionel Naccache, Karine Naccache, *Parlez-vous cerveau ?*, Paris, Odile Jacob, 2018, p. 72.
[532] Institut du Cerveau, *Comprendre le cerveau, son développement, son fonctionnement, sa plasticité et ses pathologies, art. cit.*
[533] Guy Orban, *La vision, mission du cerveau : les trois révolutions des neurosciences*, *op. cit.*, p. 26-27.

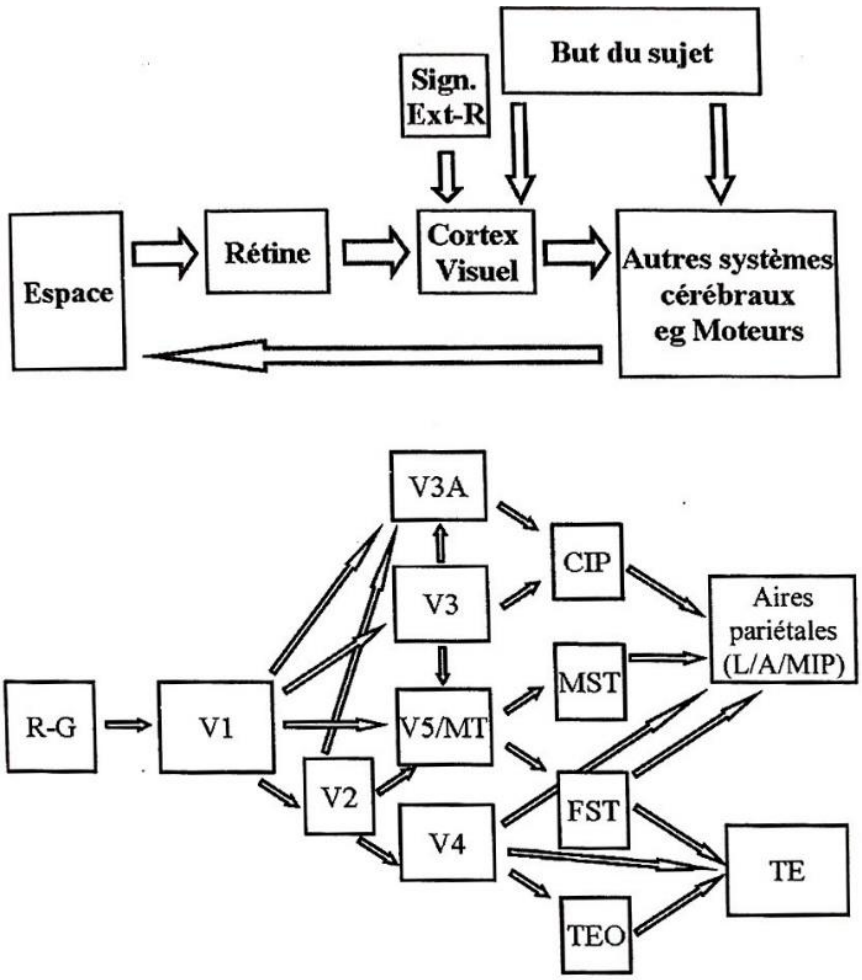

Figure 1 — Schéma de la situation visuel dans le cerveau avec vue détaillée des connexions divergentes et convergentes. Guy Orban, *La vision, mission du cerveau : les trois révolutions des neurosciences*, *op. cit.*, p. 57.

A la lumière de ce schéma, nous comprenons donc que le cheminement de l'information perçue ne se borne pas au seul cortex visuel puisque deux voies sont identifiables : la voie dite « dorsale », l'information rejoint le cortex pariétal, et la voie dite « ventrale », où elle rejoint cette fois le cortex inférotemporal. Alors que la première porte sur la perception du mouvement[534], la seconde identifie la géométrie — la forme bidimensionnelle et tridimensionnelle — ainsi que les propriétés matérielles — la texture et la couleur — de l'objet perçu[535].

Le passage de l'information depuis la rétine au cerveau est ce qui se nomme le « circuit sensitif ». Lorsque l'œil — un photorécepteur — capte un stimulus lumineux, un influx nerveux est engendré. Ce dernier remonte le nerf optique pour rejoindre le cortex visuel et les autres cortex ainsi que nous l'avons vu. La transmission de tout influx nerveux est rendue possible par

[534] Bruno Dubuc, *Les différents cortex visuels*, Le cerveau à tous les niveaux !, [en ligne], disponible sur https://lecerveau.mcgill.ca/flash/a/a_02/a_02_cr/a_02_cr_vis/a_02_cr_vis.html (consulté le 23 mai 2021).
[535] Guy Orban, *La vision, mission du cerveau : les trois révolutions des neurosciences*, *op. cit.*, p. 61.

les neurones qui composent les nerfs de notre corps. Ces derniers sont constitués d'un cytosquelette — le squelette d'une cellule — et de fibres nerveuses de deux types : les dendrites qui conduisent l'influx au sein du cytosquelette, et les axones qui conduisent l'influx vers d'autres neurones[536]. L'influx nerveux parvient alors au cerveau en passant ainsi d'un neurone à un autre. A l'image d'un graphe donc, nos neurones forment un réseau. C'est ce constat qui amène Tony Buzan à créer la carte heuristique — mind mapping — à la fin du siècle dernier :

> *« En étudiant la structure du cerveau, j'ai eu la révélation que je cherchais. […] Chacune de ces cellules [les neurones] est dotées de tentacules qui rayonnent depuis le centre, comme les branches d'un arbre. J'ai compris que je pouvais utiliser ce modèle à la manière d'un diagramme pour créer un outil de réflexion révolutionnaire. »*[537]

Bien qu'il faille se garder de confondre cartes heuristiques et cartes conceptuelles — la première représentant des informations en partant d'une idée centrale et la seconde regroupant un ensemble de concepts sur un sujet[538] — les deux illustrent le fait que notre pensée est radiale. Une idée confirmée par Tony Buzan : « Le cerveau humain ne pense pas en termes de barres d'outils, de menus et de listes ; il pense de façon organique »[539]. Ainsi, en imitant la structure de nos neurones, le graphe semble embrasser le fonctionnement de cette pensée radiale ; en ce sens, RiC apparaît comme le moyen idéal pour permettre une plus grande compréhension contextuelle des fonds. Mais représenter des informations sous forme de graphe est-il suffisant pour saisir pleinement les spécificités des archives ? David Chavalarias nous répond : « Si je vous montre un graphe mis n'importe comment, ça ne va rien vous apporter, mais si le graphe a été travaillé et la disposition des nœuds du graphe suit une certaine logique, vous êtes capable d'attraper cette logique même si vous avez deux cent mille nœuds »[540]. Un graphe « brut », non travaillé, ne présentera donc pas d'intérêt pour l'utilisateur final. Ainsi, le travail de d*ata visualisation* doit prendre en compte le processus de traitement de l'information, afin de proposer une interface adaptée : « We can use these basic attributes [neurons that are tuned to detect particular attributes of the visual information], such as differences in lenght, size, hue, color intensity, angle, texture, shape, and so on, as the building blocks of data visualization »[541]. Ce travail de *data visualisation* ne peut se faire sans prendre en compte le fonctionnement de nos systèmes mnésiques et de nos processus attentionnels. Les premiers constituent une

---

[536] Daniel Richard (dir.), *Mémo visuel de neurosciences : du neurone aux sciences cognitives*, Malakoff, Dunod, 2021, p. 2.
[537] Tony Buzan, *Développez votre intelligence avec le mind mapping*, Paris, Alisio, 2018, p. 14.
[538] Pierre Mongin, Fabienne De Broeck, *Enseigner autrement avec le mind mapping, cartes mentales et conceptuelles*, Malakoff, Dunod, 2016, p. 22.
[539] Tony Buzan, *Développez votre intelligence avec le mind mapping*, *op. cit.*, p. 41.
[540] Entretien du 29 avril 2021 avec David Chavalarias, annexe 1, 00.44.30s.
[541] Stephen Few, « Data visualization for human perception », *The Encyclopedia of Human-Computer interaction*, Interaction Design Foundation, 2014, [en ligne], disponible sur https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception (consulté le 23 mai 2021).

« dimension essentielle des processus cognitifs »[542]. Aussi dans le cadre de notre exposé, il convient de se pencher plus particulièrement sur deux d'entre eux : la mémoire sensorielle et la mémoire de travail. Comment ces deux mémoires fonctionnent-elles ? Quelles sont leurs caractéristiques ? Pourquoi est-il nécessaire de les prendre en compte lors d'un travail de *data visualisation* ? Les seconds sont tout aussi déterminants dans un projet de *data visualisation* ; l'attention étant « mobilisée dans toutes les activités cognitives »[543]. Comment se manifeste-t-elle ? Quel est son rôle ? Autant de questions auxquelles nous nous efforcerons d'apporter des réponses.

3.2. *Data visualisation* : de l'importance d'une ergonomie attractive pour une information qui « saute aux yeux »

La *data visualisation* est un outil précieux pour permettre la compréhension d'un grand nombre d'informations complexes en un seul coup d'oeil. Pour y parvenir, de multiples facteurs sont à prendre en compte tels que la couleur, la police, la taille de cette dernière, les formes, les supports, etc. Tous ces éléments doivent servir la compréhension de l'information représentée. En effet, une *data visualisation* réussie est synonyme de compréhension optimisée. Mais comment pouvons-nous expliquer cela ? Réponse avec l'examen des mémoires éphémères — la mémoire sensorielle et la mémoire de travail — ainsi que des processus attentionnels, et plus spécifiquement ceux de l'attention partagée et de l'attention sélective ; des procédés cognitifs en étroite collaboration.

3.2.1. Systèmes mnésiques et processus attentionnels

Tout d'abord, la mémoire sensorielle permet la conservation de l'information perçue pendant une fraction de secondes à l'état brut, sans opérer à un quelconque traitement[544] : « La mémoire sensorielle permet donc de garder, l'espace d'un instant, quelque trace du stimulus dès que celui-ci n'est plus reçu »[545]. Il en existe une pour chacun de nos sens ; celle qui nous intéresse aujourd'hui est la mémoire sensorielle visuelle, également connue sous le nom de mémoire iconique. Elle est dite éphémère car sa durée de conservation a été mesurée à 550 millisecondes par George Sperling — un psychologue cognitiviste américain — en 1967, soit 0,25 secondes. Aussi, « Cette mémoire retient l'information durant une période de temps tellement brève que le temps d'en rappeler quelques éléments suffit pour que l'information restante soit oubliée »[546]. Une mémoire dont le rôle est ensuite d'alimenter la mémoire de

---

[542] Margaret W. Matlin, *La cognition, une introduction à la psychologie cognitive*, Bruxelles, De Boeck Supérieur, 2001, p. 100.
[543] Patrick Lemaire, *Abrégé de psychologie cognitive*, Bruxelles, De Boeck, 2006, p. 22.
[544] Margaret W. Matlin, *La cognition, une introduction à la psychologie cognitive*, *op. cit.*, p. 150.
[545] *Ibid.*
[546] Claudette Fortin, Robert Rousseau, *Psychologie cognitive : Une approche de traitement de l'information*, Québec, Presses de l'Université du Québec, 2012, p. 37.

travail[547], qui elle va procéder au traitement des informations stockées. Cependant, il est bon de rappeler ici que la quantité d'information qu'il est possible de conserver est limitée ; un constat réalisé par George Miller en 1956 lors de ces travaux sur la mémoire où il met en lumière l'empan mnésique, c'est-à-dire la quantité d'informations qu'il est possible d'appréhender instantanément[548]. Ce dernier est mesuré à « sept plus ou moins deux », un chiffre pouvant varier selon les individus et les stratégies adoptées pour étendre cet empan mnésique, nous y reviendrons ultérieurement. Ainsi, du point de vue de notre mémoire de travail, la représentation sous forme de graphe s'apparente à une extension des capacités cognitives et permet alors de réduire la charge inhérente au traitement cognitif :

> *« Because of the limitations of short-term memory, it is generally difficult to think about a complex issue and at the same time to think about what you are thinking. […] However, a semantic network can serve as an adjunct to or extension of short-term memory, supporting examination of and reflection about one's ideas. »*[549]

De la même manière que nous ne pouvons engranger une quantité illimitée d'informations au sein de nos systèmes mnésiques éphémères, nous ne pouvons porter notre attention sur une quantité illimitée d'informations. Plusieurs processus attentionnels existent : l'attention partagée, l'attention sélective et l'attention soutenue. Dans le cadre de notre exposé, nous nous intéresserons davantage aux deux premiers. En effet, ces derniers sont étroitement liés :

> *« Les problèmes d'attention partagée et d'attention sélective représentent deux facettes d'une même question. Dans le premier cas, l'accent est mis sur le fait qu'il peut être difficile de prêter attention à plusieurs choses en même temps ; dans le second qu'il peut être également difficile de diriger son attention sur une chose en ignorant le reste. Ces deux situations impliquent des stimulations multiples. »*[550]

L'attention partagée permet de traiter en même temps plusieurs stimuli, plusieurs sources d'informations, tandis que l'attention sélective « ignore » les stimuli secondaires pour se concentrer uniquement sur l'information pertinente qu'il faut traiter en profondeur[551]. Aussi, dans un contexte d'interface présentant de multiples entités et leurs innombrables relations, l'utilisateur final sera amené à passer de l'une à l'autre au gré de ses recherches. Afin de rendre son expérience confortable, jouer de l'orientation exogène de l'attention — c'est-à-dire

---

[547] Daniel Richard (dir.), *Mémo visuel de neurosciences : du neurone aux sciences cognitives*, *op. cit.*, p. 121.
[548] Jean-Marc Meunier, *Mémoires, représentations et traitements*, Malakoff, Dunod, 2019, p. 10.
[549] Kathleen M. Fisher, « Semantic networking: the new kid on the block », *Journal of Research in Science Teaching*, vol. 27, n°10, 1990, p. 1009.
[550] Claudette Fortin, Robert Rousseau, *Psychologie cognitive : Une approche de traitement de l'information*, *op. cit.*, p. 64.
[551] Patrick Lemaire, *Abrégé de psychologie cognitive*, *op. cit.*, p. 22.

déclenchée par un stimulus externe en opposition à endogène lorsque provoquée par l'individu même[552] — semble idéal pour une expérience confortable de l'interface. Mais comment parvenir à guider le lecteur dans sa quête d'information ?

3.2.2. Principes de Gestalt, sémiologie graphique : pour un guidage exogène de l'attention

Ainsi que nous l'avons vu, la *data visualisation* repose sur un travail de présentation des données. Il ne s'agit pas en effet de ne travailler que sur le fond, la forme est elle-aussi un outil d'optimisation de la compréhension à ne pas négliger. Un bref examen de nos processus cognitifs nous a appris que nous ne pouvons ni stocker, ni traiter un nombre illimité d'informations instantanément.

> *« Parce qu'il est lui-même borné, l'esprit humain est mal à l'aise devant l'infini. Le monde ne lui devient pensable que dans les limites qu'on lui assigne. […] Comprendre veut dire embrasser par l'esprit, ce qui implique le dis-cernement. »*[553]

La dernière phrase est éclairante, elle nous indique que percevoir signifie découper l'information perçue[554]. Aussi, dans le cadre de notre empan mnésique mesuré à « sept plus ou moins deux », la stratégie de regroupement — dite aussi stratégie de chunking — qui « consiste à ne pas considérer les items à stocker de manière individuelle »[555] pourrait grandement en accroître la capacité de stockage. La suite de chiffres suivante en donne un exemple concret : 092016122019022021. Si nous devions mémoriser ces chiffres de manière individuelle, seuls sept items environ seraient retenus ; or si nous reconnaissons les dates des publications des versions 1.0 et 2.0 de RiC-CM et RiC-O, alors il nous serait possible de conserver chacun de ces chiffres en mémoire. Nous avons donc découpé l'information comme suit : 092016 122019 022021. L'efficacité de cette stratégie de regroupement peut être grandement accrue par la présence d'informations dites « perceptivement saillantes », des informations censées attirer l'attention de manière exogène[556]. Pour ce faire, garder en tête les principes de la Gestalt lors d'un travail de *data visualisation* semble judicieux ; mais qu'est-ce que ces derniers nous apprenent-ils au juste ?

Théorisée en 1920 par Max Wertheimer, Wolfgang Köhler et Kurt Koffka, les principes de la psychologie de la Gestalt traduisent les manières dont notre cerveau organise l'information

---

[552] Patrick Lemaire, *Abrégé de psychologie cognitive*, *op. cit.*, p. 22.
[553] Annette Beguin-Verbrugge, *Images en texte, images du texte : dispositifs graphiques et communication écrite*, Villeneuve d'Ascq, Presses universitaires du Septentrion, 2006, p. 69.
[554] *Ibid.*
[555] Patrick Lemaire, André Didierjean, *Introduction à la psychologie cognitive*, Bruxelles, De Boeck, 2018, p. 82.
[556] Emmanuelle Boloix, « Attention et perception de scènes visuelles », *L'année psychologique*, vol. 107, n°1, 2007, p. 118.

CC BY NC ND

sensorielle et ce de manière systématique[557]. Les connaître peut donc servir un travail de *data visualisation*. Ces principes sont au nombre de six. Sans tous les énumérer, nous allons porter notre attention sur ceux qui, dans notre contexte d'interface pour la représentation des données archivistiques sous forme de graphe, pourraient nous être utiles. Parmi eux figurent le principe de similarité, le principe de proximité ainsi que celui de continuité.

Le premier — le principe de similarité — souligne qu'une ressemblance entre deux objets amène l'esprit à les percevoir comme un groupe. Aussi dans notre cas, il pourrait s'avérer intéressant de distinguer les entités Record Resource (RiC-E02), Instantiation (RiC-E06), Agent (RiC-E07), Activity (RiC-E15) et Mandate (RiC-E17) en leur attribuant une forme propre : un carré pour Record Resource (RiC-E02), un cercle pour Agent (RiC-E07), etc. Ainsi l'utilisateur final saurait, en balayant simplement l'écran du regard, distinguer les producteurs des documents d'archives par exemple : « Quand on crée une visualisation, on sait bien que le choix de certaines couleurs va attirer l'œil plus que d'autres, on peut avoir une forme qui est plus ciselée ou au contraire un peu plus vague dans ses contours, ça va avoir des effets différents »[558]. Ce principe ne repose cependant pas uniquement sur la forme ; la taille, l'intensité, la couleur, la densité de texture et l'orientation sont autant de caractéristiques qui peuvent être attribuées aux entités décrites. Il s'agit-là des « variables visuelles » énoncées par Jacques Bertin dès 1967. Ces variables font alors partie intégrante d'une grammaire graphique édictée par Bertin lui-même. Une grammaire qui fait écho au structuralisme de Ferdinand De Saussure puisque reposant sur l'analyse de l'information à représenter — le signifié — et des moyens du système graphique — le signifiant — ; l'objectif étant de parvenir à « l'efficacité », c'est-à-dire à la limitation du « coût mental » qui se mesure en temps de perception écoulé[559]. Pour y parvenir, toutes ces variables visuelles peuvent être conjuguées entre elles. Ces dernières sont analysées de manière automatique par notre cerveau ainsi que l'illustre David Chavalarias : « Toutes les capacités cognitives de l'œil, du cerveau à détecter des nuances de couleurs, des nuances d'orientation sont automatiques. Par exemple la détection des contours c'est quelque chose qui est automatique, qu'on arrive très bien à faire »[560]. A bien y regarder, cette quête d'efficacité n'est pas si différente des objectifs de la *data visualisation*.

Le second — le principe de proximité — indique que l'esprit tend à imaginer une relation entre deux objets proches spatialement parlant ; selon lui, il s'agit d'un même ensemble. Pour l'élaboration de l'interface de visualisation, ce principe est également digne d'intérêt si l'on souhaite que se « découpent » visuellement les différents fonds d'archives conservés par l'institution par exemple. Aussi ces deux principes — de similarité et de proximité — illustrent notre faculté à regrouper des informations afin d'en faciliter le traitement. David Chavalarias en

---

[557] Patrick Lemaire, André Didierjean, *Introduction à la psychologie cognitive*, *op. cit.*, p. 38-39.
[558] Entretien du 29 avril 2021 avec David Chavalarias, annexe 1, 00.25.27s.
[559] Jacques Bertin, *Sémiologie graphique : les diagrammes, les réseaux, les cartes*, Paris, École des Hautes Études en sciences sociales, 2005, p. 12.
[560] Entretien du 29 avril 2021 avec David Chavalarias, annexe 1, 00.34.34s.

CC BY NC ND

fait également mention : « Nos sens sont précablés pour extraire des corrélations spatiales dans les choses que l'on voit […] on voit qu'il y a des ensembles et c'est en cela que voir un graphe nous enseigne quelque chose, pour autant qu'il soit bien spatialisé »[561]. Cette notion de spatialisation nous amène donc au principe de continuité.

En effet, ce dernier affirme que l'œil se dirige automatiquement d'un objet à un autre dès lors qu'il existe un lien de similarité ou de proximité entre les éléments représentés. En ce sens, il attire notre attention sur la nécessité de bien organiser l'information. Ce principe répond alors au besoin — pour la perception — d'une certaine régularité en termes de composition spatiale ; aussi, en permettant à l'œil d'anticiper le prochain saut oculaire à effectuer pour se positionner sur l'objet suivant, ce principe se porte garant d'un certain confort de visualisation[562], la lecture de l'interface est ainsi fluide. Cependant, le champ d'action de ce principe va bien au-delà de la simple organisation spatiale des objets représentés.

3.3. Explorer le graphe : proposer des outils d'interaction

En effet, le principe de continuité peut se manifester de plusieurs manières au sein d'une interface de visualisation. Au-delà de l'aspect purement visuel, ce principe met l'accent sur l'importance du rôle de l'interactivité dans la création d'une interface de visualisation.

> *« Ce qui est important de mon point de vue dans une visualisation, c'est le côté interraction […] cliquer dessus, tourner le graphe, filtrer des choses… tout cela est très important car c'est ce qui permet justement de tourner autour de l'objet dans l'espace de grande dimension dans lequel il vit. […] et c'est parce que vous êtes capable de tourner autour que vous vous faites une représentation plus complète de la chose. »*[563]

Ce principe sert donc plusieurs objectifs qui indiquent comment articuler *data visualisation* et interactivité[564]. Parmi eux figure la connexité, soit la « considération d'un seul tenant des éléments de l'interface […] pour permettre la continuité visuelle et interactive »[565]. Si nous traduisons cela en matière d'outil devant servir l'interactivité, cela peut être une option de zoom mise à disposition de l'utilisateur du graphe afin de générer diverses vues. Une telle fonctionnalité pourrait pallier les problèmes de compréhension liés à un bruit trop important d'informations ainsi que le mentionne Jean-Séverin Lair : « Il faut que la représentation puisse parler. Regardez un graphe : si vous avez dix nœuds, cent nœuds, mille nœuds, cent mille

---

[561] Entretien du 29 avril 2021 avec David Chavalarias, annexe 1, 00.44.30s.
[562] Jean-Luc Vinot, Stéphane Conversy, *La continuité, une idée-force de design pour l'architecture graphique des systèmes interactifs*, actes de la 27ème conférence francophone sur l'Interaction Homme-Machine organisée à Toulouse en 2015, p. 4.
[563] Entretien du 29 avril 2021 avec David Chavalarias, annexe 1, 00.25.52s.
[564] Jean-Luc Vinot, Stéphane Conversy, *La continuité, une idée-force de design pour l'architecture graphique des systèmes interactifs*, *op. cit.*, p. 3.
[565] *Ibid.*, p. 6.

nœuds, un million de nœuds, ça devient un truc tout noir parce qu'il y a des traits dans tous les sens, vous comprenez rien »[566]. La solution alors envisagée serait la proposition d'une fonctionnalité de type « zoom » :

> *« En fait il y a une différence [entre ce que l'on montre à l'utilisateur final et la vraie structure du graphe], moi je vois ça un peu comme des niveaux de zoom, […] avec le corollaire derrière qui est qu'en fait, ce que l'on montre peut être différent en fonction du profil utilisateur. On peut avoir des vues "utilisateur expert" versus des vues "utilisateur novice". »*[567]

Un procédé qui nous amène cependant inévitablement à faire des choix de représentation ainsi que le souligne David Chavalarias :

> *« Quand vous avez un objet qui est dans un espace de très grande dimension et que vous voulez le représenter de manière à ce que vous avec vos yeux, vous puissiez le voir, […] on va être obligé d'écraser, de compresser l'information, et de ne retenir de cet objet et des informations qu'on a dans cette grande dimension, que certains aspects. »*[568]

La problématique étant alors de savoir quelle information privilégier ; pourquoi conserver tel élément au détriment d'un autre ? Une seule réponse à cette question : l'utilisateur final. En effet, ce sont ses besoins qui doivent orienter ces choix de représentation. Aussi, la possibilité de le laisser « composer » lui-même la vue qui correspond le plus à ses besoins en mettant à sa disposition une liste de filtres qu'il peut cocher ou décocher à sa guise n'est pas à exclure.

La recherche d'une certaine « cohérence des représentations et des comportements interactifs pour garantir la continuité graphique et sémantique »[569] est également l'un des objectifs visés par le principe de la continuité. Cela pourrait se traduire par la création de saillances visuelles — soit des informations perceptivement saillantes[570] — adaptées aux besoins de l'utilisateur final. Initialement, cette expérience a été menée par une psychologue américaine en 1985, Anne Triesman, qui a alors fait la découverte des « caractéristiques pré-attentives », c'est-à-dire pouvant « être perçues par un œil humain très rapidement (< 200 ms), inconsciemment et sans effort »[571]. Aussi, il pourrait s'avérer intéressant que les résultats des requêtes formulées par l'utilisateur final se parent d'une couleur — admettons le rouge — tandis que le reste des entités et des relations du graphe apparaissent en noir, comme mises en

---

[566] Entretien du 25 mars 2021 avec Jean-Séverin Lair, annexe 1, 00.51.40s.
[567] Entretien du 18 mars 2021 avec Thomas Francart, annexe 1, 00.48.20s.
[568] Entretien du 29 avril 2021 avec David Chavalarias, annexe 1, 00.19.07s.
[569] Jean-Luc Vinot, Stéphane Conversy, *La continuité, une idée-force de design pour l'architecture graphique des systèmes interactifs*, *op. cit.*, p. 6.
[570] Emmanuelle Boloix, « Attention et perception de scènes visuelles », *art. cit.*, p. 142.
[571] Jean-Daniel Fekete, Jeremy Boy, « Recherche en visualisation d'information ou dataviz : pourquoi et comment ? », *I2D - Information, données & documents*, vol. 52, n°2, 2015, p. 32.

arrière-plan. Cette fonctionnalité permettrait de détacher un certains nombre d'informations de la masse d'entités et de relations pour en faciliter la visualisation.

Enfin, conjugué aux fonctionnalités de zoom et de saillances visuelles peut être évoquée la possibilité de visualiser le graphe en trois dimensions. Une telle fonctionnalité est indispensable pour une appréhension complète des informations représentées. Il y a bien évidemment une explication cognitive à cela : il s'agit de la cognition située. En effet, cette dernière prend en compte l'influence de l'environnement et des interactions que nous entretenons avec lui dans l'élaboration de nos processus cognitifs[572]. David Chavalarias illustre ce phénomène :

> *« La perception des choses que l'on a est très liée à notre capacité à bouger, à se mouvoir dans le monde et donc à changer les points de vue. Donc dans les représentations, ce qui est important c'est de reconstituer ce rapport aux objets, c'est-à-dire de pouvoir bouger et tourner autour. Et donc à ce moment-là on a une représentation située qui est la manière naturelle de percevoir. »*[573]

Ainsi, de la même manière que le graphe est plus pertinent qu'une représentation linéaire en imitant la structure organique naturelle de nos neurones, la visualisation en trois dimensions permet une compréhension plus grande puisque reproduisant le fonctionnement de notre cognition située.

Conclusion de l'étude de cas

Evaluer l'existant, uniformiser les pratiques, élaborer l'implémentation technique, soumettre à commentaires, recommencer… la confection d'un modèle conceptuel et d'une ontologie est un travail d'orfèvre. A l'échelle internationale, il s'agit également d'une aventure humaine : être patient et bienveillant pour surmonter la barrière de la langue et parvenir à un consensus, faire preuve d'ouverture d'esprit en croisant les pratiques et les expériences, considérer la critique afin d'envisager certains aspects sous un angle nouveau et… recommencer. En ce sens, la construction de RiC est une course de fond ; l'étude de ces deux composants — le modèle conceptuel et l'ontologie — nous a permis de comprendre combien les avancées de l'un pouvaient influencer les progrès de l'autre, et *vice versa*. Par ailleurs, il nous semble important de rappeler que bon nombre de travaux ont inspiré le développement de RiC, qu'il s'agisse de projets issus des communautés professionnelles apparentées ou de projets de description archivistique nationaux. Australie, Nouvelle-Zélande, Espagne, Finlande... RiC ne connaît pas de frontière.

---

[572] Jean-François Dortier (dir.), *Le cerveau et la pensée, le nouvel âge des sciences cognitives*, Auxerre, Sciences humaines, 2013, p. 42.
[573] Entretien du 29 avril 2021 avec David Chavalarias, annexe 1, 00.25.52s.

CC BY NC ND

En tant que voyageur expérimenté, RiC fait preuve d'une grande capacité d'adaptation au sein de son pays d'accueil. Sa mise en œuvre aux Archives nationales à travers le projet PIAAF en est témoin ; RiC est flexible, utile et extensible, c'est là-même ce qui constitue son ADN. Cette expérience a permis de constater qu'une interopérabilité des données culturelles est possible — et souhaitable — et cela a motivé le développement de RiC-O Converter puis de Sparnatural. Par ailleurs, cette intéropérabilité peut se comprendre à plusieurs niveaux qu'il convient tous de traiter : l'interopérabilité entre modèle de données ainsi que l'interopérabilité entre vocabulaires contrôlé. Aussi, la mise en œuvre de RiC nécessite quelques ajustements au regard des référentiels qu'il faut alors enrichir et normaliser.

Enfin, le projet d'une interface qui puisse présenter les données archivistiques des Archives nationales sous forme de graphe promet un outil sans précédent pour la compréhension contextuelle des fonds. En effet, nous avons vu que cette forme de représentation était particulièrement adaptée à notre fonctionnement cognitif, pour peu que l'interface bénéficie d'un travail de *data visualisation*. Il est nécessaire cependant que l'utilisateur final n'aie pas une attitude passive face au graphe qu'il lui sera donné à voir ; il est nécessaire en effet qu'il puisse interagir avec lui afin de pleinement appréhender la complexité qui entoure les documents d'archives.

CONCLUSION GÉNÉRALE

RiC constitue un pas en avant dans l'histoire de la description archivistique, et ce changement est visible à la fois pour le professionnel mais également pour l'utilisateur final.

Tout d'abord, cette norme est née d'un constat : celui de l'impossibilité de prendre part au projet FRBRoo sans d'abord doter la description archivistique d'un modèle conceptuel qui lui est propre. Une fois de plus, l'élément déclencheur est extérieur ; ce sont les normes des bibliothèques et des musées qui motivent l'évolution de notre paysage normatif. Aussi, afin d'être en mesure de dialoguer avec ces communautés professionnelles, ainsi que pour remédier à un cadre normatif quelque peu daté dont les normes s'articulent avec peine, l'EGAD voit le jour. Sans nier le rayonnement culturel des pays issus du monde anglo-saxon — par un nombre de représentants plus important ainsi que par l'usage de l'anglais comme langue de travail — l'EGAD se distingue néanmoins de ces prédécesseurs par le souhait d'atteindre une plus grande représentativité des traditions archivistiques existantes. Cette volonté est traduite à travers la composition du groupe de travail — plus diversifiée — ainsi que sa méthodologie, fondée sur le principe de consensus et l'appel à commentaires.

La norme élaborée par l'EGAD est particulièrement intéressante : la redéfinition des fondamentaux de la description archivistique qu'elle réalise permet de saisir l'essence-même de la discipline. RiC ne constitue cependant pas une rupture avec les normes mère de l'ICA ; beaucoup de leurs caractéristiques se retrouvent au sein du modèle conceptuel. Conjuguées à une plus grande représentativité des pratiques, cela est même indispensable pour s'assurer de l'adoption de RiC ; il ne s'agit pas d'abandonner les pays ne faisant pas état du même niveau d'avancement de normalisation en la matière, le changement doit se faire de manière progressive et en fonction des besoins de son service. Ce dernier élément est par ailleurs essentiel, puisque ce sont ces besoins qui guident la manière de s'approprier le modèle conceptuel et l'ontologie : il ne s'agit pas d'utiliser RiC dans son ensemble, mais de déterminer quel sous-ensemble adopter et transposer pour la description des ressources conservées.

Enfin, la possibilité de voir les données d'une institution sous forme de graphe est également une avancée majeure du point de vue de l'utilisateur. En effet, nous avons constaté que la représentation sous forme de graphe facilitait le traitement de l'information car imitant la structure organique de nos neurones. La cognition est également un élément à ne pas négliger lors du travail de *data visualisation* puisque la seule représentation sous forme de graphe — compte tenu de la volumétrie et de la complexité inhérentes aux archives — ne pourrait suffire à faciliter les recherches de l'utilisateur final. Un travail sur les formes et leur taille, les couleurs et leur intensité, la densité de texture et l'orientation ne doit donc pas être négligé. Enfin, il apparaît essentiel de proposer à l'utilisateur final la possibilité d'interagir avec le graphe, notamment en lui permettant l'exploration de ce dernier grâce à un outil de visualisation en trois dimensions qui réponde aux besoins de sa cognition située.

Pour aller plus loin, il serait intéressant — si ce n'est indispensable — que l'interface future propose l'apposition de filtres de couleur pour que les personnes souffrant de daltonisme puissent également bénéficier du travail de *data visualisation* réalisé. Ainsi l'utilisateur final pourrait activer le filtre correspondant à son type de daltonisme. A défaut de pouvoir proposer cette fonctionnalité dès le lancement de la plateforme de visualisation, tester les contrastes en passant l'interface en niveaux de gris constituerait déjà un début.

ANNEXES

1. Entretiens

1.1. Entretien avec Madame Claire Sibille de Grimoüard

1.1.1. Guide d'entretien

Thème	Sous-thème	Temps
Parcours	Études (diplôme, lieu, dates, stages)	
	En 2018, vous avez été nommée directrice de la bibliothèque et des archives du Conseil d'État. Quels ont été vos emplois précédents ? (instances ?)	
	Candidature ICA (date, motivations, attentes, craintes)	
	Intérêt personnel pour la normalisation	
Normalisation en général	Etapes d'élaboration d'une norme (des travaux préparatoires à l'appel à commentaires)	
	Vérification de la conformité d'une norme	
	Avantages et inconvénients de la normalisation	
Normalisation de la description archivistique au sein de l'ICA, et notamment lors de l'élaboration de RiC	Existe-t-il des conflits d'intérêts palpables au cours du procédé de normalisation ? (membres dont l'objectivité et la neutralité peuvent être remises en cause/volonté de faire-valoir des traditions archivistiques/influencer)	
	Comment êtes-vous parvenu à un consensus par-delà les cultures et la barrière de la langue ? Quelle est la stratégie adoptée ?	
	Paradoxe entre la recherche d'un consensus et la présence de conflit d'intérêt.	
	Le consensus trouvé relève-t-il de l'acceptation d'une tradition archivistique dominante ou d'un subtile mélange de spécificités issues de plusieurs traditions archivistiques ?	
	Consensus et traditions archivistiques : n'y a-t-il pas un risque de perte de spécificités archivistiques ?	

	« *Il reste encore beaucoup à faire, en raison notamment des grandes divergences de concepts d'un pays à un autre* »[574] expliciter.	
	« *Les concepts archivistiques sont difficiles à traduire non seulement d'une langue à une autre, mais également d'une tradition archivistique à une autre* »[575] un exemple ?	
	Faites-vous appel à des traducteurs/interprètes ?	
	« *Le développement des normes de description archivistiques semble être principalement une affaire européo-australo-nord-américaine* »[576] comment y remédier ?	
	Le fait que la langue de travail soit l'anglais constitue-t-il un frein à la participation d'autres pays ?	
Les standards de communication	SQL (technologie de base de données) et XML (technologie de balisage) permettent des approches différentes dans le représentation des données, quel impact du point de vue de la compréhension de ces données et de leur exploitation cela a-t-il permis ?	
	Au sujet du modèle RDF, vous dites dans un article coécrit avec Daniel Pitti qu'il permet de « *révéler de nouvelles connaissances* »[577], expliciter.	
	L'intérêt de RDF pour la recherche d'information (information retrieval). (recherche dans le sens de manière de retrouver des informations provenant d'une ou plusieurs base(s) de données).	
	L'intérêt de RDF pour la recherche (research).	

[574] Claire Sibille-de-Grimoüard, « Normes internationales et influence française », *art. cit.*, p. 29.
[575] Claire Sibille-de-Grimoüard, « Les normes internationales de description archivistique : origines, développements, perspectives », *art. cit.*, p. 88.
[576] *Ibid*. p. 89.
[577] Claire Sibille-de-Grimoüard, « Toward an international conceptual model for archival description: a preliminary report from the International Council on Archives' Experts Group on Archival Description », p. 573.

	(recherche dans le sens d'ensemble d'études et de travaux ayant pour objet de faire progresser la connaissance)	
	Pensez-vous que RDF pourrait favoriser la sérendipité ? (faire des découvertes non anticipées, trouver ce que l'on ne cherchait pas)	
RiC	En quelles mesures, faire la preuve de l'utilité et montrer la faisabilité de RiC serait intéressant pour l'EGAD ?	
	Est-il aussi intéressant pour les archivistes que pour les records managers ?	
	RiC et le cadre normatif du records management : complémentaire ?	
	Piste de développement d'un sous ensemble de RiC-CM et de RiC-O adapté aux ressources conservées par les AN ? (Vers une évolution ou une extension de RiC?)	
	Est-il possible de déconnecter les objectifs archivistiques de la solution technique ?	
	En 2006 vous avez écrit un article au sein de la revue Comma intitulé « Les instruments de recherche électroniques » où vous dites que « *la mise en œuvre des nouveaux standards se heurte souvent aux limites financières et techniques des institutions* », pensez-vous que RiC pourrait connaître le même sort et par conséquent ne pas être adopté par l'ensemble des institutions ?	
Référentiels	Situation actuelle	
	Mise en œuvre de l'enrichissement/fusion	
	Normalisation (normes qui encadrent la production des référentiels)	
	Quel est le processus d'application de la norme ISO 25964 ? (distinction formelle concept/terme préconisée par la norme pour la création et la gestion des thésaurus)	

	Outil pour la transformation des vocabulaires au format SKOS ?	

	Quels sont les résultats de l'étude menée courant 2020 pour évaluer les impacts d'une évolution du SIA des AN vers le support du format SKOS ou d'un format compatible avec ISO 25964 ?[578]	
	Quelles sont les actions menées pour accroître l'utilisation des référentiels (sensibilisation auprès des archivistes ?)	
	Comment utilisez-vous la plateforme GINCO du ministère de la culture dans l'élaboration des vocabulaires contrôlés ?	
	Avantages et inconvénients d'une sémantisation très précise ?	
	Est-ce qu'il existe différents niveaux en matière de sémantisation, en fonction qu'il s'agisse d'un service départemental, d'un portail agrégateur national ou d'un autre niveau ?	
	Est-ce qu'il existe des sémantisations plutôt appropriées à l'échelle d'un service départemental et d'autres à l'échelle d'un service national ?	
Représentation sous forme de graphe	Quelles sont les difficultés d'un passage de l'ensemble des jeux de métadonnées des AN à un modèle de graphe ?	
	Intérêt pour la description archivistique	
	Intérêt en termes de compréhension et d'exploitation des ressources archivistiques	
	Du point de vue de l'utilisateur, comment traiter de grands ensembles d'information ?	
	Proposition de méthodes de partitionnement des graphes pour en faciliter la lecture et l'exploitation ?	

[578] Florence Clavaud, « Transformer les métadonnées des Archives nationales en graphe de donnée : enjeux et premières réalisations », *art. cit.*, p. 81.

1.1.2. Transcription chrono-thématique

00.00.08s Début de l'entretien

Parcours

00.00.46s « Je suis conservateur général du patrimoine, directrice de la bibliothèque et des archives du Conseil d'Etat depuis septembre 2018. J'ai commencé ma carrière comme responsable de la collecte au Service historique de l'armée de terre. J'ai ensuite rejoint le Centre historique des Archives nationales comme adjointe de Christine Nougaret, responsable de la section des archives privées. J'y ai rédigé un Etat des fonds d'archives privées conforme à ISAD(G) qui a fait l'objet des premiers tests d'encodage en XML/EAD aux Archives nationales. J'ai été recrutée à ce moment-là par Catherine Dhérent, responsable du département de l'innovation terminologique et de la normalisation à la direction des Archives de France ; donc j'ai été recrutée sur un poste de chargée de mission de la description archivistique. Quand j'ai rejoint les Archives de France en 2003, c'était la fin de la révision de la norme ISAAR(CPF) donc c'est à ce moment-là que j'ai commencé des travaux de normalisation. Dans la foulée, j'ai ensuite rejoint la Société des archivistes américains où j'ai contribué aux travaux de traduction de la documentation de l'EAD 2002. J'étais chargée du groupe d'experts de l'Afnor chargé de traduire la documentation de l'EAD et de suivre ses évolutions. J'ai également participé au groupe d'experts Afnor chargé de traduire la documentation de l'EAC. Et donc j'ai participé aux instances de normalisation pendant une dizaine d'années. J'ai également participé à l'élaboration des normes du Conseil international des archives, à savoir la norme ISDF sur les fonctions et la norme ISDIAH sur les institutions de conservation. Et j'ai également participé aux débuts des travaux d'élaboration du modèle RiC. »

Origines de l'intérêt pour la description archivistique

00.04.55s « J'ai toujours été très intéressée par les instruments de recherche, comment allait évoluer la description, comment homogénéiser la description. J'étais passionnée par l'indexation aussi. La collecte m'intéressait mais j'ai toujours été passionnée quand même plus par le traitement. »

Etapes de la normalisation au sein de l'ICA

00.05.41s « Alors d'abord il y a la constitution du groupe. La constitution du groupe ça c'est très important parce qu'il faut, en fait, que le plus de traditions archivistiques possibles soient représentées. Et c'est vrai que si on regarde la composition des groupes de normalisation, il y a certains pays qui sont plus actifs que d'autres. Par exemple les pays asiatiques sont très peu représentés. Il y a essentiellement les pays de l'Europe, l'Europe occidentale, d'Amérique, d'Australie. Donc ça c'est un premier point important. Ensuite, il faut bien délimiter le projet, donc par exemple pour la norme sur les fonctions, on a commencé par faire un état des travaux déjà menés, avant que la réflexion soit portée au niveau international, il y avait déjà des pays qui avaient commencé à y réfléchir. Donc on part de cette réflexion pour pouvoir élaborer ensemble des solutions. L'Australie par exemple, avait beaucoup réfléchi sur la description justement des fonctions, également le Canada. Donc nous sommes partis sur des exemples nationaux pour identifier un petit peu les points communs des descriptions. Pour le modèle RiC, le point de départ a été un peu la même chose. C'est-à-dire qu'on a commencé par faire un peu l'état des lieux de la réflexion. Certains pays avaient déjà commencé à travailler, à modéliser la description archivistique : les Australiens, encore eux, mais également l'Espagne. Et puis là ce qui était aussi intéressant dans ce projet c'est qu'on a essayé de comparer avec d'autres professions, en particulier les bibliothèques, qui étaient beaucoup plus en avance sur la modélisation des descriptions. »

00.08.11s « Alors, on essaie de remédier à ça [la vision européo-australo-nord-américaine de la normalisation de la description archivistique] en particulier par les appels à commentaires. Donc le groupe se constitue, après on fait un état des lieux et à partir des travaux sur l'état des lieux, on fait une première version de la norme. Cette première version fait l'objet d'un appel à commentaires et donc à ce moment-là, c'est par le biais de cet appel à commentaires qu'on peut faire intervenir le public pour élargir un peu plus la réflexion. Alors, peut-être aussi quelque chose d'important, tous les travaux se font en anglais. Le fait que les travaux soient en anglais n'est pas neutre. Les concepts fondamentaux sont bien sûr les mêmes, que ce soit dans un pays anglo-saxon ou dans un pays de tradition latine mais il y a quand même des différences. Par exemple, la notion de série. La notion de série, ce n'était pas inné pour les archivistes français, pour les archivistes de tradition latine. Avant ISAD(G), la notion de série renvoyait à la série du cadre de classement. Ou la notion même d'intitulé, ce n'était pas non plus évident. Un archiviste américain ou un canadien est très influencé par la bibliothéconomie et va donc caractériser chaque unité documentaire par un intitulé bref, alors qu'un

archiviste latin, lui va rédiger quelque chose de beaucoup plus détaillé, une analyse archivistique, notamment aux niveaux fins de la description. Même la notion de producteur, la notion de contexte ou la notion de provenance n'ont pas tout à fait la même définition d'un pays à l'autre. Certains archivistes avaient une vision de la provenance un peu statique (un producteur + un fonds). Pour d'autres pays, la provenance c'est au contraire une notion très mouvante, tout peut être créateur de contexte (par exemple les fonctions qui passent d'un producteur à un autre).

L'Anglais, un frein à la participation d'autres pays ?

00.11.21s « Alors, je dirais que pour le CIA, il y a quand même une tradition multiculturelle. On arrive à peu près à s'entendre, c'était beaucoup plus difficile par expérience de travailler dans le groupe de la Société des archivistes américains. Parce qu'il ne faut pas oublier, même si la Société des archivistes américains s'est ouverte à l'international, que l'EAD a d'abord été créée pour des services d'archives américains. Oui, tout à fait [le fait que l'EAD se soit créée au prisme d'une certaine vision américaine, vision qui transparaît à travers la norme]. Et ça, je dois dire que ça a été assez difficile d'arriver à des compromis en particulier pour la France qui aurait souhaité, pour la révision de l'EAD 2002, revoir complètement les concepts sous-tendant le format d'échange. Les services d'archives américains disposent d'un important stock d'instruments de recherche encodés en EAD et donc finalement ils ne voulaient pas que cet encodage soit remis en cause. Ce qui a aussi compliqué les choses, c'est que derrière, il y avait aussi les éditeurs des logiciels et des lobbyings. Il y avait aussi les enjeux de reprise des données ce qui fait que le résultat de la révision n'est au final pas pleinement satisfaisant pour les archivistes français.

00.13.56s « Non, pas du tout [l'ICA ne fait pas appel à des traducteurs ou des interprètes pour la traduction des concepts] alors oui ça aussi c'est très important, c'est que les traductions des normes sont assurées, dans le cas du Conseil international des archives, par les membres du groupe. »

00.14.55s « Si aussi [par rapport aux divergences de concepts d'un pays à un autre] peut-être que certains pays n'ont pas le même niveau d'avancement, et le même niveau de maturité dans la mise en œuvre des normes. Par exemple, des pays qui participent au stage technique international, où je suis intervenue plusieurs fois. C'est vrai, certains pays ont peu de moyens, moins de moyens et donc ils en sont encore à simplement mettre en œuvre ISAD (G). »

00.15.55s « Ce qui est difficile aussi dans la normalisation, c'est que, on ne fait pas une norme pour le plaisir de faire une norme, il faut que ce soit appliqué. Donc si

des pays n'arrivent pas à suivre, c'est un problème. Donc il y a beaucoup à faire pour, déjà, que tous les pays soient au même niveau. Et puis, on n'a peut-être pas assez creusé le fait qu'il y ait des normes qui soient moins utilisées que d'autres. La norme ISAD(G) est la plus utilisée, la norme ISAAR (CPF) l'est moins et les deux autres, pas du tout. Et puis il y a aussi un autre problème, un autre axe de travail, c'est que l'articulation des normes entre elles n'est pas optimale. Les normes ont été élaborées à des époques différentes donc elles sont mal articulées les unes par rapport aux autres, fatalement. Par exemple, dans ISAD (G), vous n'avez pas de zone des relations tandis que dans ISAAR (CPF) vous avez une zone des relations qui permet, si on prend la norme au pied de la lettre, d'établir une relation d'une notice d'autorité vers une description archivistique. Il y a aussi beaucoup de choses qui sont implicites dans les normes et dans leur utilisation comme dans leur manipulation. »

00.18.22s « Alors, si bien sûr il peut y avoir des conflits d'intérêts. Après, les conflits d'intérêts je les ai surtout ressenti, et je reviens à ce que je disais, surtout au sein de la Société des archivistes américains, moins au sein du Conseil international des archives. »

00.19.16s « C'est vrai que l'archivistique dominante c'est quand même l'archivistique anglo-saxonne. Mais je pense qu'on a réussi, notamment avec RiC, à rééquilibrer un peu les choses et à conserver des traditions. Enfin je trouve que la composition du groupe de RiC est plus diverse par rapport aux groupes précédents. Il y a quand même plus de pays de tradition latine, l'Afrique est représentée et il y a aussi des pays d'Europe centrale comme la Roumanie. »

RiC, une preuve de concept ?

00.20.43s « Non, pas du tout [il n'y a pas eu de preuve de concept pour RiC avant la lancée du pilote PIAAF], et ça, ça a été justement un gros débat. Il y a eu des critiques, notamment de la part des experts d'Interpares, lors de l'appel à commentaires, sur la faisabilité de la mise en œuvre de RiC. Enfin il y a des pays qui en sont encore à se poser des questions sur la manipulation d'ISAD(G). Ces pays sont très éloignés d'une réflexion sur la modélisation de la description archivistique et sur l'apport du web de données pour valoriser les inventaires d'archives. »

RiC & records management

00.22.19s « Alors c'est l'ambition. Un des reproches que l'on a fait aux normes de descriptions internationales c'est aussi qu'elles étaient un peu tournées au

niveau des archives définitives et pas du tout adaptées pour les archives courantes et intermédiaires et inadaptées également pour des archives électroniques. »

00.23.05s « La connaissance des fonctions était très utiles pour décrire des archives définitives mais également très importante pour un records manager pour l'évaluation. »

00.23.35s « Alors je sais que dans le panorama des modèles qui ont inspiré RiC, il y a eu des modèles australiens, et justement un modèle orienté records management. Il y a d'ailleurs un représentant australien dans le groupe. »

RiC & les nouvelles technologies

00.25.25s « Oui effectivement, ça peut permettre de prendre du recul [le fait de déconnecter les objectifs archivistiques de la solution technique] par rapport à des descriptions archivistiques existantes. Peut-être les reprendre, de peut-être mieux définir et savoir observer les éléments de description, de réfléchir aux relations aussi entre les différentes normes de description. »

00.26.18s « C'est vrai que c'est comme ISAD (G), finalement ISAD (G) est toujours pertinente parce que ISAD (G) est conceptuelle et déconnectée des technologies donc effectivement si on veut que le modèle évolue et soit pérenne il faut effectivement qu'il soit indépendant des nouvelles technologies. C'est une faiblesse d'ailleurs de l'EAD, parce que l'EAD repose sur le XML et le XML c'est une logique très hiérarchique et donc ça enferme la description archivistique dans un type de représentation avec la description à plusieurs niveaux. »

00.27.21s « En fait, ce que RiC veut démontrer aussi c'est que, certes la description à plusieurs niveaux est tout à fait pertinente, c'est un mode de représentation de l'information tout à fait pertinent, mais il y a d'autres méthodes et d'autres modes de représentation d'un fonds et de ses composantes, en interreliant avec d'autres fonds ou d'autres séries organiques. L'objectif aussi de RiC, par rapport à ISAD (G), c'est de décloisonner les descriptions archivistiques. Avec ISAD (G) en fait, les descriptions archivistiques fonctionnent un peu en silos, donc si vous voulez créer des relations entre deux séries organiques qui sont dans deux fonds différents et dans deux instruments de recherche différents, ça va être compliqué avec ISAD (G). »

Des limites à la mise en œuvre de RiC ?

00.29.06 « Je crois beaucoup à la politique des petits pas. C'est-à-dire qu'en fait pour moi la mise en œuvre de la normalisation se fait par étapes. Donc, le point de

départ pour moi c'est ISAD (G) et puis après, si on a plus de moyens, on va progressivement vers ISAAR ou vers d'autres normes, mais il faut aussi se poser la question : pourquoi on utilise telle norme ? Pourquoi utiliser RiC ? Pourquoi utiliser RiC dans sa complétude ? Pourquoi utiliser une partie de RiC ? Pour quoi faire ? Pour quel public ? Pour quel besoin ? Pour quel type de description ? Si c'est utiliser RiC pour le plaisir d'utiliser RiC, je ne vois pas l'intérêt. Il faut vraiment que ça corresponde à la nature des fonds que l'on a à traiter et aux attentes de son public. Alors oui c'est vrai qu'il y a des limites financières, institutionnelles, mais je pense d'abord qu'il y a une vraie réflexion à avoir sur ce qu'on veut faire et pourquoi on veut le faire. Il faut bien définir le besoin. »

Les standards XML & RDF

00.31.06s « XML est très hiérarchisé. Ça a pesé vraiment sur la représentation de la description archivistique. Regardez par exemple le portail européen, c'est vraiment hyper hiérarchisé. C'est-à-dire que pour arriver à une unité documentaire, il faut descendre je ne sais combien de niveaux. Si vous voulez naviguer d'un fonds à l'autre, vous êtes obligé de remonter toute la hiérarchie. [Du point de vue de l'utilisateur] ce n'est pas forcément ce qu'il y a de plus intuitif. »

00.32.26s « En fait, [au sujet de RDF], c'est donner plus de visibilité sur la toile, sur Internet des descriptions archivistiques et permettre aussi de créer des relations avec d'autres descriptions. Par exemple, le SIA des Archives nationales contient des notices d'autorité producteur et en fait vous avez à côté le fichier d'autorité de la BNF. Et vous pouvez avoir des notices qui décrivent dans les deux fichiers d'autorité les mêmes personnes. Donc ça pourrait être intéressant de faire des rapprochements. »

00.33.30s « Et puis, peut-être que la BNF a des informations que les Archives nationales n'ont pas et *vice versa*. Donc en fait il y a un enrichissement des connaissances de part et d'autre. »

00.35.13s « Donc ça pourrait être intéressant d'avoir quelque chose qui puisse agréger en fait toutes les ressources provenant de ou relatives à Marceau Long, successivement directeur général de l'administration de la fonction publique, président de l'ORTF, secrétaire général du gouvernement et vice-président du Conseil d'Etat. Donc une approche archivistique traditionnelle ne le permet pas. Alors qu'une approche par graphes, au lieu d'avoir un modèle hiérarchisé, avoir un modèle par graphes qui permet d'établir des relations entre les différents types de ressources. Là ça va permettre d'avoir un ensemble plus conséquent

de connaissances illustrant les activités de Marceau Long. »

RiC & la sérendipité

00.36.24s « Tout à fait [le fait que RiC, par l'utilisation du format RDF, favoriserait la sérendipité], je pense que ça peut effectivement faire découvrir aux chercheurs des ressources dont il ignorait l'existence, dont il ne soupçonnait même pas l'existence. Encore une fois, dans le cas de Marceau Long, c'est absolument tentaculaire. »

Normes & référentiels

00.38.49s « ISAD (G), ISAAR (CPF) et même RiC et l'EAD aussi, c'est des coquilles vides. C'est-à-dire que c'est des outils qui servent à structurer l'information et qui ne disent pas forcément comment on doit renseigner le contenu. Donc RiC par exemple, identifie un certain nombres d'entités, il dit à quoi elles correspondent, il définit leurs relations, mais il ne dit pas forcément comment on doit renseigner l'information à l'intérieur. Ça s'explique aussi par le fait que c'est une réflexion au niveau international. Donc c'est déjà très important de s'entendre sur les concepts, sur la manière dont on va échanger l'information, mais après, c'est au niveau de chaque pays, de chaque institution de s'entendre sur les référentiels et sur la normalisation des contenus. Donc en fait la normalisation c'est un peu un cercle vicieux ou vertueux, c'est-à-dire que ça appelle toujours plus de normalisation. Le premier niveau c'est comment on va structurer l'information pour pouvoir l'échanger et puis après quels sont les libellés, quels sont les intitulés. Donc effectivement le rôle des référentiels est très très important. De toute façon, la leçon que je tirerais de la normalisation, c'est que les normes, les outils comme RiC, ce ne sont que des outils donc après ça ne dispense pas d'un travail d'appropriation de ces outils. Ce n'est pas parce que l'on a un outil qu'on l'a pris en mains et on ne se pose pas de questions sur son application, pas du tout. Ça rejoint un peu ce que je disais tout à l'heure : pourquoi on utilise tel outil et comment, après, on va l'utiliser ? RiC effectivement, est-ce qu'on utilise simplement le modèle ou le modèle et l'ontologie ? Si on utilise que le modèle, est-ce qu'on utilise le modèle entier ou est-ce qu'on utilise une partie du modèle ? On ne va peut-être pas avoir la même approche selon que l'on est un records manager ou que l'on est dans un service d'archives définitives. Tout dépend en fait du contexte et du type d'information qu'on a à traiter. »

Sémantisation

00.42.18s « Oui [au sujet que le niveau de la sémantisation varie en fonction du type de service où l'on se trouve], je pense que ça ne va pas forcément être la même chose. Et puis je pense aussi qu'il faut réfléchir plus globalement, parce que les Archives nationales et les archives départementales ont des fonds qui sont complémentaires. Donc je pense qu'il y a un rôle au niveau de l'institution mais après aussi un rôle au niveau de comment on s'insère dans un écosystème. »

00.43.16s « [Au sujet des avantages et inconvénients d'une sémantisation très précise] Je pense qu'il ne faut pas être trop ambitieux dès le départ, qu'il faut y aller progressivement. Moi si demain on me dit : "vous avez le modèle RiC, comment vous allez l'implémenter aux archives du Conseil d'Etat ?" qui sont des archives intermédiaires, je ne vais pas l'implémenter tout de suite en bloc. Je vais peut-être faire un test d'abord, peut-être sur une partie des fonds qu'on a traité. Donc pour moi, on ne peut pas aller tout de suite vers une sémantisation très précise, ça peut se faire que progressivement. Sinon le risque c'est de se décourager. C'est de se décourager mais c'est aussi de décourager les équipes. Parce que ce n'est pas le tout de décider d'implémenter telle norme, il faut aussi que l'équipe suive, donc il faut derrière aussi les former et conduire le changement. Ça aussi c'est une dimension qui est très importante. Il ne faut pas que ça soit qu'une affaire de spécialistes. »

Formation à RiC

00.44.33s « Je pense qu'il faudra effectivement penser derrière à des ateliers oui. C'est un peu ce qui s'est passé pour l'EAD, parce qu'au départ on était même pas une dizaine de personnes formées puis peu à peu on a mis en place des formations. On a mis en place des formations de formateurs pour passer le relai, que ce soit pas toujours les mêmes personnes qui forment. Je pense que pour RiC il va se passer la même chose. »

Représentation sous forme de graphe

00.45.59s « [Au sujet des difficultés potentielles lors d'un passage de l'ensemble des jeux de métadonnées des Archives nationales à un modèle de graphe] Alors, je suis le projet de loin alors ce que je vais vous dire c'est des souvenirs que j'ai du projet PIAAF. Donc en fait, on avait travaillé à partir d'un corpus de métadonnées en EAD et en EAC. Donc après il y a eu un gros travail de qualité des données. Enfin quel que soit le format que l'on utilise, il faut des données de qualité. Il y a un travail d'homogénéisation, même si le format de départ c'est l'EAD. Et puis après effectivement, pour passer à une logique de graphe,

il faut une politique d'identifiants pérennes. Parce qu'en fait, les graphes reposent sur des triplets et donc en fait pour établir les relations entre les triplets, il faut que les ressources soient caractérisées par des identifiants. Donc je pense qu'il faut réfléchir à la qualité des données, réfléchir à une politique d'identifiants et puis également réfléchir aux relations. Le gros problème aussi aux Archives nationales, c'est qu'ils ont plusieurs milliers d'instruments de recherche, à côté ils ont des notices d'autorité producteurs, et en fait, derrière il y a énormément de relations qui sont implicites. Donc quel type de relation on veut créer ? Ça peut aller très loin, ça peut être vertigineux les relations, on peut faire des relations avec tout. Donc il y a aussi, je pense, une limite peut-être à se poser. D'accord on veut faire des relations mais entre quoi et quoi ? Et quel type de relation ? Et puis après il faut pouvoir représenter les relations. Dans PIAAF il y a eu ce travail de préparation des données mais il y a derrière un outil qui permet justement de visualiser, un outil de *data visualisation* qui permet de visualiser de manière dynamique les relations. »

00.49.13s « [L'intérêt en termes de compréhension des ressources archivistiques] C'est pour un individu donné, la possibilité par exemple de visualiser sous forme de graphe toutes les ressources qui le concernent, qui émanent de lui. Ou visualiser sous forme de graphe ses relations avec d'autres individus ou des institutions, des collectivités. Documenter le contexte. Avoir, en fait, des multiplicités de points de vue sur le contexte. Même les points d'accès, même les termes d'indexation peuvent être des éléments de contexte. »

00.50.38s « Il y aura une vue d'ensemble plus grande, une plus grande facilité quand même d'accès à l'information. Une plus grande richesse effectivement de la description contextuelle. Et puis peut-être que ça va donner aussi aux chercheurs des idées de rapprochement. Le fait de rapprocher des ressources ça peut donner des idées de rapprochements à d'autres sources. Et puis ça peut peut-être aussi permettre de décloisonner les archives, peut-être de rapprocher les archives d'autres types de sources. Donc, ne plus avoir simplement des travaux qui s'appuient uniquement sur des ressources d'archives mais sur d'autres types de sources. »

00.52.02s « [Sur la manière d'exploiter de grands graphes] Je pense que les technologies de graphes, enfin je dis peut-être une bêtise, mais je pense qu'elles ont leurs limites dans la masse de données qu'elles peuvent gérer. Manuellement, s'il [l'utilisateur] le consulte comme ça, il ne va pas pouvoir l'exploiter, donc peut-être qu'il va falloir que lui aussi derrière il ait des outils qui permettent d'exploiter les données en graphes. Et peut-être qu'il va falloir qu'il ait

justement des outils qui lui permettent, par exemple, de faire des requêtes en langage SPARQL. En fait s'il veut exploiter les données ce n'est pas uniquement les visualiser en graphes, il faut aussi qu'il puisse les interroger, qu'il puisse les manipuler, qu'il puisse les retraiter, et puis peut-être qu'il puisse aussi les rapprocher, créer à son tour d'autres relations. Donc je pense que le chercheur de demain, il devra être familier des outils informatiques, des nouvelles technologies pour exploiter pleinement ce qu'on lui présente. »

00.54.10s « [Au sujet du fait que la plateforme de visualisation des graphes puisse fournir des outils pour leur exploitation] Alors là il faudrait effectivement réfléchir à une évolution peut-être des systèmes d'information archivistique. C'est sûr que le SIA des Archives nationales tel qu'il se présente aujourd'hui, ne rend pas du tout compte de toutes ces relations. Et même un outil comme le PIAAF, c'est un prototype mais effectivement, il faut aller plus loin. »

00.56.08s Fin de l'entretien.

1.2. Entretien avec Monsieur Thomas Francart

1.2.1. Guide d'entretien

Thème	**Sous-thème**	**Temps**
Parcours	Études (diplômes, lieux, dates, stages)	
	Vous êtes actuellement à la tête de la société Sparna. Emplois précédents ?	
	Comment s'est mise en place l'association avec l'EGAD pour le développement de l'ontologie ?	
RiC-O	Quand on élabore une ontologie, il y a nécessairement un modèle conceptuel de réalisé en amont sur lequel on s'appuie ?	
	Comment avez-vous traduit le modèle conceptuel en ontologie ? Quelles sont les méthodes et les outils nécessaires à son élaboration ?	
	Est-ce que RiC-O respecte les critères propres aux ontologies qui sont : la clarté, la cohérence, l'extensibilité, un engagement ontologie minimal et l'évitement de toute déformation d'encodage minimale ? Pouvez-vous détailler pour chacun de ces critères la manière dont RiC-O les respecte ?	
	Pourquoi avoir choisi de considérer les lieux comme des entités géo-historiques ?	
	Pourquoi avoir choisi de considérer certaines relations comme un objet à part entière (avec une date et une description) ? Quel intérêt se cache derrière ce choix ?	
	Est-ce que toutes les relations sont considérées comme un objet à part entière ?	
	Quel est l'intérêt des relations binaires ? (propre à RiC-O ou toutes les ontologies les proposent ?)	
	Comment faites-vous pour aligner les concepts archivistiques avec les concepts des bibliothèques/des musées ? (quel procédé, quels dispositifs/solutions techniques nécessaires)	
	Comment fonctionne SKOS ?	

CC BY NC ND

	J'ai lu que vous utilisiez le logiciel Protégé pour réalisez des tests. Autres logiciels ?	
	Comment testez-vous l'ontologie pour savoir si elle est fonctionnelle ? (Tests unitaires, en quoi ils consistent)	
	Est-ce que vous avez recours à la spécification SHACL de W3C pour valider RiC-O ? Si oui, de quelle manière ? Comment est-ce que ça fonctionne ? Quel est le jeu de contraintes définit que l'on souhaite tester sur les données ?	
	Dans un article intitulé « SHACL (RDF Shapes Constraint Language : enfin la possibilité de valider des données RDF » qui est sur votre blog, il est écrit que *« SHACL permettra de spécifier la « forme » à laquelle on souhaite qu'un graphe RDF réponde »*. Pouvez-vous expliciter svp ?	
	Dans ce même article, vous écrivez également que grâce à cette spécification *« il ne sera plus obligatoire de passer par une étape où l'on doit expliquer ce qu'est une ontologie »*, pouvez-vous expliciter svp ?	
	La seconde version bêta de RiC-O vient de paraître. Quand la version 1 sera disponible, comment un service d'archives intéressé par RiC pourra l'installer dans son SIA ?	
	Travaillez-vous sur une piste de développement d'un sous ensemble de RiC-CM et de RiC-O adapté aux ressources conservées par les AN ? (Vers une évolution ou une extension de RiC ?)	
Représentation sous forme de graphe	Avec RiC on va pouvoir avoir des jeux de métadonnées archivistiques en RDF. Mais qu'en est-il de la création d'outils d'interrogation et de visualisation de ces données ?	
	L'objectif est-il de créer une plateforme de visualisation des graphes indépendante de toute SIV ou d'intégrer cette fonctionnalité au SIA de chaque institution ?	
	SI CREATION PLATEFORME : Quelles sont les difficultés rencontrées lors de sa création ?	
	Vous inspirez-vous de l'interface développée par le projet de recherche ALEGORIA ou de celle de la BnF ?	

	Est-ce qu'il y a des prérequis à la réalisation de requêtes SPARQL ? (logiciel à installer)	
	Qu'est-ce que SPARNATURAL ? Comment on l'installe ?	
	Comment ce composant d'exploration interagit avec la SIV ? (indépendant, extension à ajouter au SIA) ?	
	Intérêt de la représentation sous forme de graphe pour la description archivistique ?	
	Lors de la journée d'étude du 28 janvier 2020, il a été dit que RiC-O permet de représenter certaines informations de plusieurs manières. Exemple : propriété « rico:history » + texte ou propriété « rico:rulefollowed » + texte. Qu'est-ce que vous entendez par « texte » ? Celui ajouté par l'utilisateur lors de la recherche (où il pourra choisir un point de vue dans le graphe) ou bien celui ajouté par l'archiviste qui décrira le fonds selon un certain point de vue ?	
	Quel est l'intérêt selon vous de pouvoir représenter certaines informations de plusieurs manières (pdv utilisateur) ?	
	Intérêt en termes de compréhension des ressources archivistiques ?	
	Intérêt en termes d'exploitation de ces dites ressources ?	
	Du point de vue de l'utilisateur, comment traiter de grands ensembles d'information ?	
	Est-ce que la plateforme de visualisation proposera des outils/méthodes de partitionnement de graphes ? Si oui, lesquels(les) ? (bipartition de graphes, méthode hiérarchique, algorithmes de *clustering*, etc.)	
RiC aux AN, PIAAF	Vous contribuez à l'élaboration de RiC-O. Avez-vous également contribué à l'élaboration du pilote PIAAF aux Archives nationales ?	
	Est-ce que PIAAF utilise toutes les fonctionnalités de RiC ? PIAAF a-t-il été pensé comme une évolution ou une extension de RiC ?	
RiC-O Converter	Quelles ont été les étapes de son élaboration ? Comment le testez-vous pour savoir s'il est fonctionnel ?	

	RiC-O a pour objectif de convertir des notices d'autorité et des instruments de recherche en RDF/OWL mais peut-il convertir des référentiels d'indexation en RDF/SKOS ?	
	Comment fonctionne la conversion exactement (que ce soit en langage OWL ou SKOS) ?	
	Le langage RDF exige l'utilisation d'URI, comment assigner des URIs à des entités qui n'en ont pas ? Quelle est la politique d'URI pour les AN (stratégie de construction et de gestion d'URIs) ? Dans la présentation de RiC-O le 28 janvier 2020, il est dit que *« des choix provisoires ont dû être faits »*, pouvez-vous expliciter ?	
	Comment parvenir à des URIs déréférençables ?	
	Comment faire émerger des entités conceptuelles qui sont implicites dans les fichiers ? Dans quel but ?	
	Dans la présentation de RiC-O du 28 janvier 2020, il est dit que *« [RiC-O Converter] ne s'occupe pas d'autres formats de description, donc en particulier des manifestes SEDA utilisés par les SAE comme VITAM »*, pourquoi & comment surmonter ce problème ?	
Les standards de communication	L'intérêt de RDF pour la recherche d'information (information retrieval).	
	L'intérêt de RDF pour la recherche (research).	
	Pensez-vous que RDF pourrait favoriser la sérendipité ?	
Référentiels	Quel est l'intérêt d'une structure hiérarchique (thésaurus) pour la consultation et l'exploitation des ressources archivistiques ?	
	Quelles sont les étapes à suivre pour fusionner deux référentiels ? (méthodes pour son enrichissement, la fusion des référentiels, etc.)	
	Quel est le processus d'application de la norme ISO 25964 ?	
	SI CE N'EST PAS RiC-O CONVERTER : Quels outils pour la transformation des vocabulaires au format SKOS ?	

	Avantages et inconvénients d'une sémantisation très précise ?	
	Du point de vue de l'utilisateur, en quelle mesure la consultation directe des référentiels améliorerait-il la recherche (information retrieval + research) ?	

1.2.2. Transcription chrono-thématique

00.00.05 Début de l'entretien

Parcours

00.00.31s « Je suis consultant indépendant, j'ai un salarié, à Sparna on est que deux. Je travaille dans le monde du web de données et du web sémantique depuis 20 ans, depuis que je ne suis plus à l'université. J'ai commencé par être développeur et responsable technique d'une société qui s'appelle Mondeca, qui existe toujours, de laquelle je suis parti en 2012 pour continuer à valoriser cette expertise sur le web de données, les graphes et les ontologies mais plutôt sous l'angle des services alors que Mondeca est un éditeur logiciel. Ce qui m'amène à faire à la fois du conseil, essentiellement pour des acteurs publics qui ont des missions de diffusion de données comme les archives, comme les musées, comme les bibliothèques, comme les labos de recherche. Je fais du conseil sur tous les aspects des projets qui touchent au web de données, donc la mise en place de graphes de connaissances, la diffusion de données, la modélisation. Je fais un peu de développement aussi donc des briques logiciel, et je fais un peu de formation. »

RiC-O Converter

00.02.16s « Fin d'année 2019, j'ai été contacté une première fois pour réaliser le prototype PIAAF mais en fait on n'avait pas donné suite au cahier des charges. Sur PIAAF à l'époque on avait décidé de ne pas y répondre. Tel qu'il était réalisé, le cahier des charges semblait trop ambitieux. J'ai été contacté une deuxième fois sur le développement du convertisseur *Records in Contexts* fin 2019. »

00.03.41s « On commence par réfléchir à la réponse au cahier des charges. Le travail commence dès la rédaction de la réponse sur comment on imagine le système technique à produire, le travail intellectuel commence là. J'avais proposé une architecture logicielle pour la réalisation du convertisseur avec des choix technologiques. Le cahier des charges incitait fortement à utiliser des feuilles de style XSLT pare que c'est une technologie qui est assez répandue dans les archives et donc les Archives nationales étaient susceptibles de bien s'approprier donc j'avais suivi cette recommandation. Donc il y a une étape de proposition d'architecture au moment de la rédaction du cahier des charges et puis ensuite, qui vient s'ajuster au moment du démarrage effectif du projet.

Après si je me rappelle bien, on avait eu au début du projet plutôt une phase de spécification. On avait fait un premier tableau de mise en correspondance des éléments EAC avec les éléments de l'ontologie RiC-O. Mais par la suite, on s'est assez vite dit que pour être vraiment sûrs de la qualité de ce que produit le convertisseur sur des cas qui sont quand même assez pointus parce qu'il y a pas mal de cas à prendre en compte, notamment par rapport à certaines spécificités que peuvent avoir les Archives nationales dans la façon dont les elles implémentent de l'EAC EAD, on s'était dit qu'il fallait passer par la construction de tests unitaires. C'est-à-dire la spécification par les Archives nationales, sur des petits morceaux de fichiers EAC et de fichiers EAD, de ce que ça peut donner en RiC-O en sortie. En fait les spécifications très précises et formelles, ont étaient codées, spécifiées de cette façon-là, c'est-à-dire en donnant le petit fichier d'entrée et ce qu'on attend sous forme de petit fichier de sortie. Ca permet de n'avoir aucune ambiguïté sur ce que doit faire le convertisseur. »

00.07.21s « Non, [sur le fait que RiC-O converter puisse convertir des référentiels d'indexation en RDF/SKOS], c'est hors du périmètre. Dans le cadre des Archives nationales, ils ne les ont pas converti, ils les ont ressaisi pour produire directement du SKOS, dans un outil de gestion qui s'appelle GINCO. »

Comment assigner des URI à des entités qui en sont dépourvues ?

00.09.52s « Il y a deux cas. C'est soit votre entité à l'origine, n'a pas d'URI mais elle a quand même un ID. Par exemple, dans le cas des Archives nationales, toutes les notices EAC ont un ID et par exemple, tous les éléments <c> dans les instruments de recherche EAD ont aussi des ID. Ce qui n'est pas une obligation en EAD mais aux Archives nationales, ça tombe bien, il y en a. Donc à partir de cet ID, on vient générer une URI en concaténant une racine d'URI et puis on vient coller l'ID technique à la fin. Donc ça c'est le premier cas où c'est facile. Le deuxième cas où c'est moins facile, c'est quand il n'y a pas d'identifiant unique, quand il n'y a pas de clé dans la donnée source et donc là, encore deux cas : Si on n'a pas un ID technique clair, on peut néanmoins se donner un ensemble d'attributs qui vont garantir l'unicité de l'entité. De mémoire dans RiC-O Converter, c'est ce qu'on fait pour identifier de façon unique ou pour générer des URI pour les relations entre entités. Les relations elles-mêmes n'ont pas d'identifiant, n'ont pas d'ID, il n'y a pas un seul champs d'ID, cependant, si on considère l'entité source de la relation, l'entité cible de la relation, la date de début de la relation, la date de fin de la relation, et probablement le type de la relation, cet ensemble de critères nous garantit l'unicité. C'est-à-dire qu'il ne peut pas y avoir deux relations du même type à

la même date entre les deux mêmes entités. Et donc en prenant ces critères-là, et en les insérant tous dans l'URI qu'on génère, on garantit de générer une URI unique. »

Sur la politique des URI

00.12.12s « [Sur le fait que des choix provisoires ont dû être faits] Les choix dont il est question ce sont les choix de génération d'URI qu'on a fait dans RiC-O Converter par rapport à une politique d'URI générale qui n'est pas encore mise en place aux Archives nationales. C'est-à-dire que les Archives nationales n'ont pas encore décidé à ce jour quelle devrait être la structure des URI qui seront publiées un jour par les Archives nationales pour identifier leurs entités. »

00.13.00s « [Au sujet de la politique des URI des Archives nationales] Ils sont en train de refaire un schéma directeur de leur système d'information sur plusieurs années et en fait, la réflexion se fera dans le cadre de ce schéma directeur. »

00.13.11s « La décision principale va se faire plutôt sur l'espace de nom, c'est-à-dire le début des URI : est-ce que ça va être du data *archivesnationales.fr* ou autre chose. Le choix principal est là. Et il est aussi bien sûr sur ce que l'on met après les URI, comment elles sont structurées dans leur suite. »

Sur les URI déréférençables

00.13.51s « Pour que l'URI soit déréférençable, il faut que l'URI qu'on a généré, lorsqu'on le met dans la barre d'adresse du navigateur en haut, nous ramène à un document. Il faut qu'il nous ramène à une représentation de cet URI. Aujourd'hui, les URI qu'on a généré sur RiC-O Converter ne correspondent à rien par rapport au serveur des Archives nationales. Si on prend un des URI qui a été généré dans RiC-O Converter, qu'on la met dans notre barre d'adresse, on aura une erreur 404 qui nous dit qu'il n'y a pas de document : le serveur ne connaît pas cet URI, il n'y a pas de représentation de cet URI. Donc pour arriver à avoir des URI déréférençables, il y a des choses à mettre en place sur les serveurs de l'institution pour être capable de dire que quand il y a un URI comme ça qui m'arrive, qu'est-ce que je dois faire, à quoi on l'associe. Ca peut aller de règles très simples à des règles plus compliquées mais typiquement il peut y avoir déjà des pages web, des URL, qui présentent dans des sites, des portails d'archives existants, la fiche d'identité avec peut-être un ID qui doit se retrouver quelque part dans l'URL de cette page. Donc par exemple, l'URI qu'on a généré pour cette entité pourrait rediriger tout simplement vers la page existante sans que l'on ait besoin de faire plus que ça. Donc ça peut aller de ça jusqu'à des règles plus compliquées où pour un

URI, il est normalement de bon ton d'être à la fois capable de renvoyer à la fois une représentation pour l'humain donc une page HTML ou une représentation pour la machine, c'est-à-dire les données RDF qui décrivent cette entité. Et donc là il y a des règles plus compliquées à mettre en place : il faut que le serveur soit capable de détecter si le client qui lui parle est plutôt un navigateur qui attend du HTML ou si c'est plutôt une application qui aimerait bien avoir du RDF. Et puis il faut qu'il soit capable de produire, d'avoir sous la main la représentation RDF, de la même façon qu'il peut avoir sous la main la représentation HTML dans un portail, il faut qu'il soit capable de produire ou de générer la représentation RDF de cette entité pour la machine. Si on arrive à faire ça, on a coché toutes les cases pour faire de la bonne exposition des données dans le web de données. »

Au sujet de l'instanciation

00.16.41s « Le modèle, ce n'est qu'un modèle, il faut l'instancier. C'est-à-dire que chaque institution doit créer les entités correspondantes à ce modèle, créer des instances de *records resources*, de *records agents*, de choses comme ça. Les instances sont des entités concrètes, des classes qui sont définies dans la coquille vide qu'est le modèle. La coquille vide nous dit : dans RiC-O, on définit des types d'entité qui s'appellent *records resource*, qui s'appellent *records*, qui s'appellent *agents*, qui s'appellent *relations*. Le modèle définit cet ensemble de types d'objet et les propriétés sur ces objets et chaque institution doit dire : ok à partir de ce modèle, chez moi, voilà la liste des *records resources*, voilà la liste des *agents* ou voilà les *relations* pour produire des données qui sont organisées suivant le modèle que définit RiC-O. C'est ce que fait RiC-O Converter pour dire : à partir de fichiers EAC et EAD, on va produire les instances de *records*, d'*agents*, etc. »

Comment RiC-O Converter fait-il ressortir des fichiers implicites ?

00.18.04s « C'est parce qu'il y a des différences de points de vue. D'abord avant le comment, le pourquoi : c'est parce qu'il y a des différences conceptuelles de points de vue entre la façon dont est organisée le fichier EAD et le point de vue qu'adopte RiC-O dans son modèle conceptuel. Enfin en l'occurrence c'est plutôt RiC-CM, le modèle conceptuel, qui a reposé toutes les bases conceptuelles de ce que sont les notices et comment sont organisées toutes ces entités que les archivistes manipulent. Il dit qu'on va considérer qu'on manipule des *records*, donc des notices, et ces notices, qui sont les niveaux intellectuels, sont liées à des instanciations ; instanciation ça va être le niveau physique. Donc je vais avoir un *record* qui peut avoir par exemple typiquement deux instanciations :

une instanciation physique et une instanciation numérisée. »

00.19.58s « Et donc RiC-O nous dit : il y a un seul record, c'est la notice plutôt abstraite, et deux représentations de cette même notice : l'objet physique et son pendant numérique. L'EAD ne dit pas du tout ça. L'EAD dit : j'ai qu'une seule entrée dans mon EAD, qui doit correspondre à la balise <c>, et dans cette entrée j'ai un attribut qui s'appelle <daogrp> [groupe d'objets archivistiques numériques] qui va pointer vers l'adresse d'une numérisation pour cette notice. Donc l'EAD dit : j'ai une seule notice et voilà l'identifiant du fichier numérisé. Donc l'EAD ne dit pas du tout que j'ai un record et deux instanciations. D'un côté dans RiC-CM on a trois entités : le record et deux instanciations. Dans l'EAD on a une seule notice qui pointe vers un truc qui dit : voilà le pendant numérisé, mais la notice contient toutes les caractéristiques plutôt de l'objet physique de départ. Et donc on est obligé, à partir de cette structure de départ qui est dans l'EAD de faire émerger des entités qui correspondent aux instanciations qui n'existent pas en tant que telles dans l'EAD. L'EAD n'a pas du tout cette distinction entre le niveau intellectuel et le niveau physique alors qu'on l'a dans RiC-CM. Donc ça nous amène à faire émerger ces entités qui correspondent aux instanciations qui ne sont pas présentent dans l'EAD. On est dans des logiques de comment on va les identifier, quels URI on va leur donner à ces instanciations et quelle est la règle qui va dire qu'à partir de la structure de l'EAD je vais créer des instanciations – une plutôt physique et une plutôt numérique – et comment je vais ventiler les attributs qui sont au départ appliqués dans la notice de l'EAD. Typiquement, on a peut-être des attributs qui sont sur la notice EAD qui sont plutôt liés à l'objet initial qu'est l'objet physique que l'on va venir redescendre sur l'instanciation physique dans le graphe RiC-O. Donc à la fois on a besoin de faire émerger ces nouvelles entités, et à la fois on a besoin de réorganiser un peu la façon dont sont exprimées les choses dans l'EAD pour venir dispatcher ça de façon un peu différente dans le graphe. Les informations ne sont pas structurées pareil parce que RiC-CM a reposé toutes les bases conceptuelles des entités qui sont manipulées par les archivistes. »

Sur le fait que RiC-O ne traite pas d'autres formats de description tels que les manifestes SEDA utilisés par les SAE comme Vitam

00.24.46s « Pour avoir échangé avec Florence pas plus tard que la semaine dernière, ils veulent le faire. »

L'intérêt de RDF pour la recherche d'information

00.25.30s « On est sur des structures qui sont fondamentalement différentes [celle de RDF comparée à celles de l'EAD/l'EAC]. Dans le monde RDF on va être sur une structure graphe qui est moins prégnante dans l'EAC et l'EAD parce qu'on est sur des structures plutôt hiérarchiques. Déjà il y a un intérêt qui est lié au nouveau modèle conceptuel, à RiC-CM, qui va reposer les bases conceptuelles de comment tout ça c'est organisé. Ça veut dire que les données qu'on va exprimer en RiC-O vont faire moins de mélanges sémantiques et ça va être un petit peu plus clair au niveau de la structuration. Et après la structure RDF permet de prendre le graphe par n'importe quel bout. C'est-à-dire que du fait que j'ai un graphe, je peux rentrer en disant : je cherche une organisation qui a publié tel fonds d'archives entre telle date et telle date, ou je peux rentrer par un document d'archives. C'est-à-dire que je cherche un document d'archives qui a été publié par une organisation, ou je cherche un agent, etc. donc en fait on peut entrer par le graphe suivant plein d'angles, alors que si on regarde les fichiers EAD et EAC, la structure est plus rigide. »

L'intérêt de RDF pour le recherche

00.27.16s « Ils [les chercheurs] peuvent rentrer par le graphe suivant l'axe qui les intéresse eux, et pas des axes qui seraient précablés par un système de recherche. »

00.27.38s « Je pense qu'on va explorer ça avec les Archives nationales [les pratiques des chercheurs en archives] de mettre en regard des capacités d'exploration du graphe avec des besoins métiers en termes de recherche d'informations. »

Est-ce que RDF pourrait favoriser la sérendipité ?

00.28.10s « Oui car on est sur la capacité de faire des liens dans tous les sens. A la fois on va faire des liens dans tous les sens dans le graphe de données qui va être produit par l'institution, donc ça veut dire que l'institution va produire un graphe très riche avec beaucoup de liens entre les entités que l'on va être capable de traverser un peu comme on veut, donc à l'intérieur des données de l'institution ça va permettre une richesse d'exploration, de navigation et de recherche, mais aussi parce que, comme on est dans le monde du web de données, notre graphe on va pouvoir aussi l'articuler avec des données d'autres institutions, d'autres sources de données en général sur le web, ce qui va permettre de rebondir vers d'autres données ailleurs. »

Comment aligner des concepts archivistiques avec des concepts d'autres institutions ?

00.30.01s « La question peut s'entendre à plusieurs niveaux. Ça peut vouloir dire : comment on fait pour aligner des modèles de données, donc par exemple comment on ferait pour aligner l'ontologie RiC-O avec l'ontologie FRBR par exemple. Elle peut s'entendre au niveau des référentiels SKOS : comment on fait pour mettre en correspondance des référentiels SKOS. Elle peut s'entendre plutôt au niveau des instances, donc plutôt au niveau des entités comme des personnes. Les problématiques sont légèrement différentes en fonction des niveaux qu'on considère. »

00.30.43s « Le niveau le plus simple c'est mettre en relation des instances, c'est-à-dire que la personne dans mes données est la même qu'une personne qui serait décrite ailleurs. Techniquement parlant, ça revient à exprimer un triplet RDF, une flèche dans le graphe, d'un type qui peut varier, il y a des choix à faire sur le type mais ça peut être des liens qui s'appellent <owl:sameAs>, de même que ça peut être des liens qui s'appellent <skos:exactMatch>, ça dépend de la sémantique que l'on veut se donner, et on va simplement mettre en relation par le biais d'une de ces propriétés-là, l'URI qui identifie la personne hors données avec l'URI qui identifie la même personne dans une autre ressource de données. On va venir en fait articuler les deux URI par une propriété tout simplement. »

00.31.54s « [Au niveau des référentiels SKOS] Ce serait fait de la même façon, donc on utiliserait plutôt <skos:exactMatch> mais en SKOS on est pas obligé de dire que des correspondances exactes. On peut dire : mon concept-là est exactement le même que celui-là. Mais on peut dire aussi : mon concept-là est un peu plus générique que celui-là, ou il est un peu plus spécifique que celui-là, ou c'est à peu près le même que celui-là. Ça permet de typer un peu le lien de mise en correspondance, ça peut être des liens avec une sémantique un peu plus floue. »

Tester RiC-O Converter

00.33.07s « Pour tester le convertisseur, à la fois on l'a testé par le biais des tests unitaires qui servent à valider le fait que le convertisseur produise bien la sortie attendue pour ces minis tests et on l'a testé aussi par des passes de validations humaines sur des fichiers plus importants, des fichiers complets, à deux niveaux : c'est-à-dire que les Archives nationales s'étaient données quelques notices, pour dire, voilà on va voir ce que ça donne sur ces notices-là. Et la deuxième passe c'est, on convertit l'ensemble du fonds, et après on vient regarder au hasard : je regarde la sortie et je regarde si ça me paraît correcte ou pas correcte. »

Avez-vous utilisé SHACL pour tester RiC-O Converter ?

00.34.16s « Non, pas formellement. C'était dans les tuyaux, on a gratté un tout petit peu cette piste mais c'était aussi hors du périmètre de ce qui était prévu. C'est une question de coût, une question de temps, une question aussi un petit peu, à l'époque, de niveau de maturité des outils. Aujourd'hui, un an et demi/deux ans après, j'aurais plus d'outils et plus de facilité à proposer des choses un peu plus industrielles entre guillemets là-dessus. Mais c'est effectivement quelque chose qui peut être mis en place pour vérifier la cohérence du graphe produit par le convertisseur. »

00.35.11s « [SHACL] va chercher ce qu'on lui demande de chercher. C'est-à-dire qu'il faut lui donner des règles de vérification, des règles typiquement qui disent, on va prendre un exemple bête : les personnes ne doivent avoir au maximum qu'une seule propriété de date de naissance. Si je trouve une personne avec deux dates de naissance, ça me sort une erreur. En fait, on définit la structure du graphe qu'on attend beaucoup en termes de cardinalité comme ça, et il n'y a qu'une propriété maximum ou minimum pour les propriétés obligatoires. C'est-à-dire qu'il y a un travail de définition de ces règles de vérification en amont, un travail de définition qui peut s'automatiser d'une certaine façon en dérivant un certain nombre de règles de vérification justement à partir de l'ontologie. A partir de RiC-O, je n'ai pas fait l'exercice mais, on devrait être capable de dériver des règles de vérification qui seraient interprétables par un validateur SHACL. »

00.36.48s « Pour prolonger l'explication de ces jeux de contraintes, ils sont intéressants parce qu'ils peuvent à la fois vérifier la structure du graphe avec des règles structurelles comme l'exemple que j'ai cité, mais on peut venir aussi encoder en SHACL des règles beaucoup plus métier, c'est-à-dire qu'elles sont liées à la spécificité du jeu de données qu'on vient produire. Des choses du style : toutes les notices dans ce fonds-là doivent avoir une date de versement supérieure à telle date. »

00.37.50s « [Sur le fait que SHACL puisse permettre de spécifier la forme à laquelle on souhaite qu'un graphe RDF réponde, sur le fait que ces règles modulent le graphe] C'est les deux, c'est à la fois le niveau structurel des règles, le niveau 1 entre guillemets, avec des règles assez simples de cardinalité : où est-ce que la propriété est mise et vers quel objet elle pointe. Et puis des règles on va dire de niveau 2, des règles plutôt métier qui vont parcourir le graphe à plusieurs niveaux ou qui vont permettre de donner des règles très spécifiques sur un sous-ensemble d'objets dans le graphe. C'est l'ensemble de ces

choses-là qui font qu'on va définir la structure du graphe qu'on attend. »

00.38.59s « [Sur le fait que grâce à SHACL, on ne sera plus obligé de passer par une définition de ce qu'est une ontologie] En fait, SHACL a un positionnement un peu différent par rapport à OWL et aux ontologies. Les ontologies sont là pour modéliser un modèle de connaissances. Mais cette modélisation ne prend pas du tout en compte la façon dont les données sont effectivement produites, générées dans le système. On peut faire qu'une seule ontologie qui peut modéliser qu'un seul domaine de connaissances mais en réalité dans le système il va y avoir des morceaux qui vont être produits à différents endroits. Par exemple dans RiC-O Converter, on a une seule ontologie RiC-O mais on va produire des *agents*, qui sont les personnes des organisations, à partir de l'EAC. Et on va produire des *records* à partir de l'EAD. Donc on a deux flux. Au départ ce sont deux sources, donc ce sont deux flux de données. SHACL lui adopte un point de vue différent parce qu'il sert à vérifier la cohérence d'un graphe, d'un flux. En SHACL on dit : voilà les contraintes sur le flux des *agents* RiC-O, voilà les contraintes – dans un autre fichier SHACL – sur les flux des *records*. Donc c'est un peu moins conceptuel de présenter comme ça pour les gens en disant : vous avez des flux différents, on définit la structure de ce flux-là, la structure de ce flux-là, la structure de ce flux-là, et c'est un peu plus compréhensible pour des clients que d'expliquer le niveau très conceptuel de ce que c'est qu'une ontologie de domaine. Et les deux s'articulent : on ne peut pas avoir de contraintes SHACL sans avoir spécifié dans une ontologie les classes, les propriétés dont il est question dans notre modèle. »

00.41.44s « Quand SHACL n'existait pas, parce que SHACL c'est récent, on ne faisait qu'une ontologie puisqu'on avait que ça, et il y avait une attente implicite des clients qui disait : ok vous m'avez défini une ontologie donc vous êtes capable de valider que les données que l'on produit sont conformes. Sauf que pas du tout. L'ontologie ne sert qu'à spécifier l'ensemble des classes et des propriétés d'un domaine mais ne permet pas du tout de valider la conformité d'un fichier de données, ce que permet justement SHACL. »

Un outil de visualisation : Sparnatural

00.42.53s « Oui, c'est ce qui avait été présenté à l'évènement du 28 janvier. C'était un prototype, c'est un outil qui s'appelle Sparnatural qui permet de naviguer visuellement dans un graphe de données RDF. On est sur un composant graphique fontaine pour l'utilisateur final qui permet de construire des requêtes de navigation dans un graphe. Dans un graphe RDF de façon générale et il a été configuré pour le prototype pour les Archives nationales, il a été configuré

spécifiquement pour naviguer dans une sous-partie d'une graphe RiC-O. »

00.43.50s « [Sur comment installer Sparnatural] Il faut lui fournir une configuration qui va lui spécifier quelle est la liste des types d'objet qu'il doit afficher, quelle est la liste des relations entre les types d'objets. Ensuite il n'y a pas d'installation à proprement parler dans le sens où ça n'est qu'un composant Javascript technique donc la seule chose que l'on a à avoir sur la machine c'est une page HTML en local avec la configuration Sparnatural à côté ou dedans. Sparnatural n'est qu'un composant de requêtes et d'interrogations mais il doit évidemment envoyer ces requêtes SPARQL à une base de données. C'est là où il peut y avoir une nécessité d'installation, il faut que l'on ait ces données quelque part, c'est là où il peut y avoir nécessité d'installer la base de données sur sa machine également. C'est ce qu'à fait Florence aux Archives Nationales : elle a un environnement Sparnatural en local sur sa machine, en ayant installé une base de données RDF dans laquelle elle a chargé ses données RiC-O, le résultat du convertisseur, et elle a cette page HTML avec la configuration associée qui va interroger sa base. »

00.45.36s « [Sur le fait que Sparnatural devienne une brique logicielle associée au SIA des Archives nationales] Dans trois ans ce sera ça mais là il faut d'abord que les Archives nationales aient l'infrastructure, qu'elles aient déployé leur base RDF sur leurs serveurs. On est sur une autre temporalité, mais l'objectif long terme c'est ça. En tout cas la philosophie de cet outil de navigation des graphes est bien d'être capable de proposer sur les sites web des institutions des interfaces de navigation « grand public » sur un graphe de données. »

00.46.24s « [Sur le fait que Sparnatural apparaisse comme une brique à ajouter aux SIA actuels ou s'il faudrait totalement les SIA pour parvenir à son utilisation] Ça vient plutôt comme un ajout, mais c'est un ajout qui se conçoit aussi comme une évolution du système d'information puisque c'est un ajout qui demande un passage vers le monde des graphes, vers le monde RDF et c'est pour ça que le converter a été développé : ça demande une transition des formats classiques EAD, EAC vers une structure en graphe. Et ça n'est qu'une fois que l'on a cette structure en graphe que l'on est capable d'avancer l'interface de navigation correspondante, donc il y a une transition à faire. Alors les deux peuvent coexister : on peut continuer à avoir des systèmes avec des outils qui fonctionnent sur l'EAD et l'EAC et puis avoir une synchronisation des outils graphe à côté, ça c'est possible. Mais ça demande quand même un cheminement de transition vers ces technologies-là. »

00.47.23s « Sparnatural c'est le nom du composant technique qui génère l'interface. Sparnatural est le nom de la brique logicielle. »

00.48.20s « [Sur les difficultés possibles lors de l'élaboration d'une plateforme de visualisation] Il y a des difficultés d'ordre technique sur par exemple le temps de réponse des requêtes : comme on va être amené à naviguer dans le graphe de façon assez complexe, de générer des requêtes assez complexes, on peut avoir des questions de temps de réponse derrière. Il y a une autre question qui est liée aux paramétrages de ce que l'on veut montrer à l'utilisateur donc ça on ne l'a pas rencontré sur le prototype avec RiC-O parce qu'on a pas été assez loin mais on le rencontre sur d'autres mises en œuvre de ce même composant Sparnatural. C'est-à-dire qu'il peut y avoir une différence entre la structure du graphe sous-jacente donc la structure d'un graphe RiC-O par exemple, et ce qu'on veut montrer à l'utilisateur final. Et en général, ce que l'on veut montrer à l'utilisateur final, c'est plus simple que la structure du graphe sous-jacente. Le graphe est fait pour être très précis alors que si on rentre dans cette précision pour un utilisateur final, ça peut être trop compliqué. En fait il y a une différence, moi je vois ça un peu comme des niveaux de zoom, c'est-à-dire que le graphe de départ est très précis et en fait on veut dézoomer un peu pour prendre un point de vue un peu plus haut niveau pour présenter ça à un utilisateur. Donc il y a une différence entre ce qu'on montre et la vraie structure du graphe sous-jacente, avec le corollaire derrière qui est qu'en fait ce que l'on montre peut être différent en fonction du profil utilisateur. On peut avoir des vues « utilisateur expert » versus des vues « utilisateur novice ». Ce sont des réflexions qu'on a en ce moment avec les Archives nationales et d'autres. Le graphe sous-jacent est toujours le même, mais on en donne des points de vue légèrement différents. Cela peut être à la fois en termes de niveaux de détails ou en termes des parties de ce graphe que l'on veut montrer, en termes de périmètre aussi de ce que l'on veut présenter. »

00.50.44s « On peut imaginer, mais ça ce sont des réflexions qui ne sont même pas à l'état embryonnaire, mais on peut imaginer de proposer à des utilisateurs plusieurs points d'entrée en disant : si vous voulez rechercher sur le graphe des producteurs d'archives, cliquez-là. Si vous voulez rechercher dans les fonds d'archives, cliquez-là. Avec, du coup, des interfaces de navigation qui seront délimitées sur leur périmètre. »

00.51.21s « [Sur les plateformes de visualisation qui ont été sources d'inspiration] Le point de départ de Sparnatural c'est un prototype du British Museum qui s'appelle *ResearchSpace* [une plateforme d'interrogation et de visualisation]. Ils ont un composant qui permet aussi de décrire des requêtes en se baladant dans le graphe qui est moins bien que ce que l'on a fait dans Sparnatural mais on a repris cette philosophie d'écriture de requêtes avec le point de départ, la

relation, le point d'arrivée et l'imbrication comme ça des critères de navigation. En termes visuel, je pense qu'on a été un cran plus loin que ce qu'ils ont proposé mais la philosophie générale est la même. *ResearchSpace* a plus de choses notamment sur la présentation des résultats de recherche. Nous dans Sparnatural on a vraiment que le composant d'écriture de la requête et on ne fait rien de particulier sur la présentation du tableau de résultats qui est un listing assez basique en fait. »

00.53.04s « On pourrait imaginer des affichages sous forme de petites cartes avec un visuel un peu plus sexy mais aussi de la *data visualisation*. Nous on a juste un tableau : si tu cherches des organisations qui ont déposé des archives entre telle date et telle date : une liste d'organisations avec leur nom. »

00.53.34s « Au départ ce composant [Sparnatural] a été développé dans le cadre d'un projet avec des archéologues qui s'appelle *OpenArcheo*, des archéologues qui sont à Tours, et ça a été déployé la première fois sur un graphe de données d'archéologie, donc un graphe qui n'est pas du tout structuré sur le modèle RiC-O, qui est structuré sur le modèle conceptuel des données patrimoniales qui s'appelle le CIDOC-CRM et il a été déployé sur quelques autres, plutôt sous forme de prototype pour le moment, mais quelques autres structures de graphe. Comme il est configurable un peu à façon, c'est toujours la même brique logicielle mais avec une configuration différente qu'on lui donne en fonction de la structure du graphe qu'on interroge. »

00.54.57s « [Sur les outils de partitionnement de graphe] Ca dépend à la fois des capacités des outils et des algorithmes qui sont plutôt du côté de la base de graphe, l'outil qui va stocker les données, qui est en l'occurrence dans le cas des Archives nationales et dans beaucoup de projets qu'on implémente, on se sert d'un outil qui s'appelle *GraphDB*, c'est la base de graphes RDF. Donc ça dépend des capacités à la fois de cet outil intrinsèquement et de ce qu'on peut mettre autour de cette base en termes d'algos, et ça dépend aussi des fonctionnalités que l'on veut présenter à l'utilisateur final en termes de navigation ou de visualisation : qu'est-ce qu'on veut lui montrer ? A quoi on veut que ça serve ? Ca c'est aux institutions de le définir. C'est à la fois un dialogue entre les capacités de la technique d'un côté et les attentes des métiers de l'autre côté. Il faut à la fois être capable de montrer ce qu'est capable de faire la technique et à quels besoins ça peut répondre. D'exprimer des besoins métier et de trouver comment y répondre du côté technique. »

00.56.35s « [Les institutions] ne sont pas en capacité de définir les algorithmes, elles peuvent exprimer des besoins. Ceci dit, elles peuvent exprimer des besoins sur la base de quelque chose qu'on leur a montré en termes d'algos, c'est-à-dire qu'on nous dit : "Regardez, on peut aussi faire ça, on pense que ce serait bien

pour telle fonction." C'est un dialogue entre les deux de toute façon dans la réalité. »

Sur les référentiels

00.57.21s « [Sur les intérêts de la structure hiérarchique des thésaurus] Il y a plein d'intérêts. Donc d'abord on va être sur des vocabulaires contrôlés et ça va dépendre aussi de la façon dont ces référentiels sont présentés et ce qu'ils permettent de faire en termes de fonctionnalités utilisateur. Cela peut permettre à la fois d'avoir un accès à l'information par cette hiérarchie : j'ai une hiérarchie de sujet, je peux la naviguer comme un arbre et cliquer et trouver tous les documents qui sont relatifs à ce sujet. Ca va permettre de m'autoriser des champs de recherche intelligents par exemple avec de l'auto-complétion : je commence à taper quelques lettres et on me propose un terme qui vient du vocabulaire contrôlé, donc ça guide les champs de la recherche plein texte. Dans ce guidage, ça permet aussi de bien montrer à l'utilisateur qu'il a cherché un terme, mais que ce terme est un synonyme d'un terme préférentiel. Tout ça c'est lié à SKOS : à la fois cette hiérarchie, à la fois ce renvoi vers le synonyme. Ca permet en plus de documenter ces notions, c'est-à-dire de dire : dans mon vocabulaire contrôlé, cette notion-là, ce sujet-là, je lui attache une définition, je le définis comme ça. Ca permet aussi de générer les listings de termes, c'est-à-dire d'avoir des entrées par listing aussi. Et puis derrière des systèmes un peu plus compliqués : ces référentiels peuvent aussi venir alimenter des outils d'analyse du langage naturel qui vont être capables de détecter les termes de mon référentiel dans le contenu texte de mes documents. »

Avantages et inconvénients d'une sémantisation très précise ?

01.01.24s « Les objectifs entre la sémantisation très précise qui se base sur un modèle conceptuel et la sémantisation pas précise qui va se baser plutôt sur des modèles peut-être plus partagés sur le web mais moins précis, il y a une différence en termes d'approche et d'objectif. C'est-à-dire que la sémantisation très précise a pour objectif de décrire le plus finement possible et le plus fidèlement possible les données qui sont dans le périmètre du système d'information de l'institution. Donc l'objectif c'est que l'institution se donne un graphe de données qui est très précis et très fidèle aux données qu'elle veut décrire. L'avantage c'est qu'on a un truc qui est hyper bien décrit, il n'y a pas d'ambiguïté, on a les différences entre niveau intellectuel et niveau physique. Donc on a un truc hyper pointu et qui répond bien aux objectifs de gestion de l'institution. Et la sémantisation moins précise type *Dublin Core* ou un autre

schéma de métadonnées, est là plutôt pour des enjeux d'interopérabilité : je veux donner une vue *Dublin Core* de mes données, c'est parce que je sais qu'il y a peut-être d'autres organisations qui comprennent le *Dublin Core*, d'autres outils qui comprennent le *Dublin Core* mais qui ne comprennent pas ma sémantique métier très précise. Dans les archives on va y venir avec des agrégateurs comme *FranceArchives* mais l'exemple dans les musées est un peu plus parlant parce qu'il y a des agrégateurs comme *Europeana*, un agrégateur de métadonnées pour les musées européens. Donc les agrégateurs comprennent un niveau de sémantique qui est, on va dire, pas précis. Alors que les musées ou les institutions patrimoniales de façon plus générale, pour décrire fidèlement leurs objets, ils ont besoin d'une sémantique beaucoup plus précise que ça. Donc les musées ont besoin du CIDOC-CRM en interne pour décrire leurs ressources, et ils ont besoin d'être compatibles avec le modèle d'*Europeana* qui s'appelle EDM [*Europeana Data Model*] quand ils envoient leurs données à EDM. Ils seront toujours capable de générer une vue plus simple sur leurs données à partir de la vue plus détaillée. Donc il y a un intérêt à passer par une vue précise et détaillée pour bien décrire fidèlement les données à gérer par l'institution et pour être capable derrière de regénérer des vues, des sémantiques plus simplifiées pour des questions d'interopérabilité. »

01.05.00s « On peut les diffuser aussi [les données] dans les deux modèles : à la fois en RiC-O parce que j'ai d'autres services d'archives qui comprennent RiC-O, qui sont plutôt du même métier donc eux vont prendre la vue plus précise, et puis je peux fournir des vues plus simples pour différents canaux d'utilisateurs. »

01.05.20s « Si on a la vue très précise dès le départ, on peut tout faire. Si on a fait que la vue simplifiée, on ne peut pas faire plus que la vue simplifiée. Donc l'avantage est là. C'est surtout dans le monde de CIDOC-CRM pour les musées que l'avantage est vraiment décrit comme ça et les gens disent : nous a une description hyper précise de nos objets patrimoniaux, et après on est capable à partir de ça de regénérer ce que l'on veut. La question se pose beaucoup par rapport au vocabulaire qui s'appelle *Schema.org* qui est fait pour envoyer les données aux moteurs de recherche pour rendre visible des données structurées pour les moteurs de recherche type Google. *Schema.org* c'est le vocabulaire qui est compris par Google pour le dire simplement. Donc il y a des institutions, par exemple la BnF le fait, qui publient leurs données en respectant ce vocabulaire *Schema.org* pour faire plaisir à Google pour le dire vite, mais elles publient ces données en *Schema.org* à partir de données plus précises exprimées, dans le cas de la BnF, en FRBR. »

01.08.03s Les relations binaires dans la structure des graphes

« Il y a des façons de modélisation qu'exploite d'ailleurs très bien RiC-O là-dessus. C'est-à-dire que le lien est toujours binaire mais on peut vouloir dire des choses sur une relation. Par exemple : il y avait une relation entre les deux *agents* dans RiC-O, mais cette relation a été valable entre telle date et telle date ; telle personne a été patron de telle société entre telle date et telle date. Donc là on a besoin d'exprimer des informations sur la relation. Mais donc dans ce cas-là, comme on a nécessairement une structure de graphe avec des liens binaires, on vient introduire un nœud intermédiaire qui représente la relation elle-même ; donc on a l'*agent* de départ, l'*agent* d'arrivée et au milieu une autre entité qui représente la relation. Et donc on a un lien entre la personne de départ et la relation, un lien entre la relation et l'organisation d'arrivée, et des informations qui sont portées par la relation. On fait de la relation un objet, on fait de la relation une entité. Et RiC-O exploite ce pattern-là à fond, toutes les relations se font comme ça. »

01.09.32s Fin de l'entretien.

1.3. Entretien avec Monsieur Jean-Séverin Lair

1.3.1. Guide d'entretien

Thème	**Sous-thème**	**Temps**
Parcours	Études (diplômes, lieux, dates, stages)	
	Vous êtes actuellement DSI de l'INSEE. Emplois précédents ?	
	Comment avez-vous rejoint le programme VITAM ?	
	Pourquoi avoir rejoint le programme VITAM ?	
	Vous avez une longue expérience dans des postes de direction IT dans le secteur public et dans le privé. Qu'avez-vous appris de ces expériences pour l'élaboration de système d'information ?	
Programme VITAM – Equipe & méthodologie	Vos missions en tant que directeur du programme ?	
	<u>Constitution de l'équipe</u> : Comment s'est-elle constituée ? En avez-vous choisi les membres ? Selon quels critères ?	
	<u>Différences et points communs entre archivistes et informaticiens</u> : Dans un article de 2016 intitulé « Table ronde Vitam au congrès de l'ICA à Séoul » écrit par Madame Mélanie Rebours dans la revue *Archivistes !*, il est indiqué au sujet des deux professions (archivistes et informaticiens) que : *« elles ont des points communs, mais aussi des distinctions qui compliquent parfois les travaux en commun »*. Quelles sont ces distinctions et comment les traitez-vous ?	
	<u>Collaboration programme VITAM</u> : Il existe plusieurs collaborations au sein du programme VITAM avec la collaboration interministérielle, la collaboration entre les deux professions et la collaboration interne à chaque partenaire. Comment s'est opérée la coordination entre les différentes équipes ?	

	Qu'est-ce que la comitologie mise en œuvre dans le cadre du programme VITAM sur laquelle repose la collaboration interministérielle ?	
	Une des spécificités de VITAM est d'avoir été développé à la fois par des archivistes, des records managers et des informaticiens. Pour quelles raisons ?	
	Comment s'opère la collaboration entre les deux professions ? Plus particulièrement entre le responsable fonctionnel incarné par Monsieur Edouard Vasseur et le responsable technique par Monsieur Frédéric Brégier ?	
	Toujours dans l'article de Mélanie Rebours, il est écrit au sujet de cette collaboration que : *« De façon plus inattendue, elle permet aux archivistes de réseaux différents de discuter , de confronter leurs pratiques, parfois divergentes ».* Quels bénéfices en sont ressortis ? Un exemple ?	
	La collaboration interne à chaque partenaire est censée accompagner la transition des services d'archives vers le numérique et assurer la conduite du changement. Méthodes employées ? Sensibilisation ? Applicables/Transposables pour RiC ?	
	<u>Méthode</u> : Qu'est-ce que la méthode Agile ? Qu'est-ce que Scrum ?	
	Comment l'appliquez-vous ?	
	Pour quelles raisons l'appliquez-vous ?	
Programme VITAM – la solution logicielle	Que dit la feuille de route initiale ?	
	Avec quelles contraintes avez-vous dû composer ?	
	Comment avez-vous étudié la faisabilité du projet ?	
	Comment avez-vous recueilli les besoins des utilisateurs ?	
	Avez-vous réalisé un cahier de recettes/tests ?	

	Norme OAIS (Système ouvert d'archivage d'information) : comment avez-vous utilisé cette norme ? Comment vous l'êtes-vous approprié ?	
	Pourquoi avoir choisi d'utiliser ce modèle conceptuel ?	
	Vous êtes-vous inspiré d'autres modèles conceptuels ?	
	Quelles fonctionnalités ont émergé des ateliers pour constituer la feuille de route 2020-2024 ?	
	Quelles sont les interfaces à développer pour la gestion technique et fonctionnelle d'une plateforme d'archivage ?	
Records in Contexts	RiC, « une fausse bonne idée » ?	
	Liens RDF : Toujours dans la revue *Archivistes !*, vous parlez dans un article paru ce mois-ci des liens RDF en disant qu'il y a un certain nombre de problèmes et notamment : *« l'incertitude des choix sur les liens faits par l'homme »*. Pouvez-vous expliciter ?	
	Pouvons-nous parler de subjectivité sous-jacente donc de la part de l'archiviste au moment de la description ?	
	Traitement des grandes volumétries par les SI : Vous donnez l'exemple de la BnF et du système SPAR qui gère son graphe sur un serveur de calcul et qui pourtant ne peut faire de requêtes sur la base complète des métadonnées sans que cela nécessite beaucoup de temps. Pensez-vous que le même sort attend Records-in-Contexts ?	
	Selon vous, quels sont les avantages et les inconvénients de proposer une base réduite de métadonnées avec des champs limités ?	
	Vous dites qu' : *« il faut savoir mixer les modèles et sans doute avoir des référentiels sémantisés, à côté de grands gisements de données plus traditionnels (et efficaces !). »*, pouvez-vous expliciter svp ? Méthodes ? Moyens ?	
	Sur les exigences primordiales des bases SQL : lorsque vous dites *« que pour atteindre de très grandes volumétries, on est obligé de limiter certaines exigences, pourtant jugées primordiales sur les bases SQL »*, à quelles exigences pensez-vous ?	

	Pour VITAM, quelles ont été les solutions de base étudiées pour les très grandes volumétries ?	
	Pourquoi avoir finalement fait le choix de solutions noSQL ?	
	Craintes vis-à-vis de RiC : Au sujet du développement de RiC, vous dites avoir peur que la démarche soit portée par des hyperspécialistes qui ne répondent plus à des objectifs concrets et sont loin du terrain. Quelles sont vos craintes vis-à-vis de ce projet exactement ?	
Les standards de communication	Sur XML dans la profession archivistique : Vous dites que sa pleine et totale adoption *« a gêné les dernières évolutions dans le domaine de l'archivage numérique »,* que le fait que les systèmes d'information récents ne font plus de XML mais du JSON *« a compliqué les développements de Vitam »*. Pouvez-vous m'en dire plus ?	
	L'intérêt de RDF pour la recherche d'information (information retrieval).	
	L'intérêt de RDF pour la recherche (research).	
	Pensez-vous que RDF pourrait favoriser la sérendipité ?	
Référentiels	Du point de vue de l'utilisateur, en quelle mesure la consultation directe des référentiels améliorerait-il la recherche (information retrieval + research) ?	
	Avantages et inconvénients d'une sémantisation très précise ?	
Représentation sous forme de graphe	Intérêt en termes de compréhension des ressources archivistiques ?	
	Intérêt en termes d'exploitation de ces dites ressources ?	
	Du point de vue de l'utilisateur, comment traiter de grands ensembles d'information ?	

1.3.2. Transcription chrono-thématique

00.00.08s Début de l'entretien.

Parcours

00.00.33s « Ma formation c'est polytechnicien, école des Télécoms à Paris. Ensuite, j'ai une carrière très interministérielle dans la Défense, la Poste, les services du premier ministre, Bercy, etc. Je suis arrivé DSI du Ministère de la Culture en 2008. En 2011, on a commencé à réfléchir à des possibilités de faire de l'archivage numérique et ça je le détaillerai un peu plus, j'ai trouvé ça intéressant. Et en 2015, je suis devenu le directeur du programme Vitam. »

Vos missions en tant que directeur du programme Vitam ?

00.01.30s « En tant que directeur du programme, j'étais missionné doublement. En fait ce programme réunissait trois ministères : les Affaires étrangères, la Culture et la Défense, et ce qui est intéressant c'est que dès 2011, les trois ont commencé à travailler ensemble et en plus sur une intelligence de mettre les archivistes et les informaticiens ensemble, donc on était à 6 pour monter progressivement les choses. Il y avait deux figures tutélaires, deux sponsors : la DISIC à l'époque qui est devenue la DINUM donc la Direction interministérielle du numérique et des systèmes d'information et de communication, et le CIAF, le Comité interministériel aux Archives de France, qui étaient tous les deux consultés sur ce programme. Les rôles définis de l'équipe l'ont été par une convention entre tous ces acteurs-là. Si je devais revenir à comment était défini mon rôle, j'avais deux prescriptions : une par la DISIC, une par le CIAF. Par la DISIC c'était de faire en sorte que ce programme interministériel soit un exemple de réussite d'un programme en plus fait en Agile. »

00.03.46s « On prend une équipe avec une forte composante interne, on la responsabilise, on éclate les prestataires pour qu'il n'y ait pas un seul prestataire dominant, et en plus on le fait en Agile. »

Sur la coordination entre les différentes collaborations au sein de Vitam

00.04.42s « La concertation avec les archivistes a été beaucoup animée grâce à la communauté des archivistes qui est déjà très fortement cohésive. Quand je dis qu'on a voulu une forte compétence interne : il y avait quatre archivistes de très bon niveau et il y avait deux ou trois architectes techniques. Donc on avait

des gens très forts qui pouvaient fédérer une communauté autour d'eux parce qu'ils étaient reconnus aussi. Et donc typiquement les archivistes ça s'est beaucoup construit grâce au relationnel d'Edouard Vasseur et de Mélanie Rebours. Après est venue s'ajouter Nathalie Morin aussi qui était connue. Et puis de la même façon du côté des informaticiens, moi j'avais été DSI donc ça facilitait l'entrée à haut niveau et du côté expertise technique, on avait un directeur technique de très haut vol et donc qui est arrivé très facilement à être crédible face à tous les informaticiens. »

Comment s'est passée la collaboration entre archivistes et informaticiens ?

00.06.24s « Excellement bien. Mais ça c'est une harmonie de personnes. Ce qu'il faut voir c'est que l'Agile c'est vraiment une question de personnes. Donc il faut arriver à construire une équipe qui fonctionne bien et qui roule bien. Ce qu'il faut voir c'est qu'on avait mis en place un *triumvirat* : il y avait Edouard et Frédéric chacun sur sa partie et moi en troisième qui éventuellement contrebalançait ou arbitrait sur certains points. Et donc vraiment la direction c'était le *triumvirat*. »

00.08.53s « Pourquoi avoir pris tout ce monde-là ? Parce qu'on voulait que tous les choix soient fait par l'administration et le service public. Donc il y avait une vraie volonté de contrôle et de maîtrise. »

00.09.38s « On a très vite constaté que la chose à fuir absolument c'est ce qu'on appelait l'archivaticien. C'est-à-dire que l'on avait beaucoup plus à gagner à s'appuyer les uns sur les autres pour nos compétences et nos forces, qu'à essayer de remplacer l'autre et à essayer de singer l'autre. »

00.10.11s « Pour compléter ce que disait Mélanie, on est deux professions qui traitons de l'information, donc on manipule des concepts similaires. Mais oui, juste seulement similaires et pas avec les mêmes contraintes du tout. La caricature que je faisais mais quelque part elle a du sens, c'est que les archivistes pensent en siècles et les informaticiens en nanosecondes. Et ça donne une vision fabuleusement différente de la façon de gérer l'information. L'informaticien va essayer de la gérer la plus efficacement possible dans l'instantané. L'archiviste est prêt à perdre du temps pour que ce soit pérenne. Et puis après les concepts de métadonnées… il y a toute une série de concepts où on est approchant mais on n'entend pas forcément la même chose ni les mêmes finalités. »

00.11.23s « Ce qu'il fallait, c'était se rendre compte sur là où on ne se comprenait pas. Après ça allait bien parce qu'on savait où étaient les zones de flou donc on se méfiait. »

00.11.56s « [Sur les bénéfices tirés par les archivistes] Je pense qu'il y en avait un énorme puisqu'on avait des ateliers où on faisait venir 20/30 archivistes mais simplement en leur disant : "l'atelier a lieu", on ne les harcelait pas pour qu'ils viennent. Donc il y avait vraiment un intérêt d'échange de pratiques. Typiquement, je ne sais pas où ils en sont actuellement, mais quand je suis parti, la communauté Vitam active sur Jalios était de plus de quatre cents personnes je crois, dont une très grande majorité d'archivistes. Donc il y avait vraiment un échange sur les pratiques et surtout sur les pratiques naissantes. Typiquement : qu'est-ce qu'on prend dans les messages ? Qu'est-ce qu'on ne prend pas ? Qu'est-ce qu'on prend dans la bureautique ? Qu'est-ce qu'on ne prend pas ? Est-ce qu'on va préserver ? Est-ce qu'on veut tel format ? Est-ce qu'on va imposer des formats ? Etc. Il y a eu tout un foisonnement sur les nouvelles pratiques, et sur la transposition des anciennes dans le numérique. Ce qui est bien aussi, c'est qu'ils mettaient en perspective leurs procédures et leurs pratiques papier pour dire : est-ce qu'on veut vraiment faire plus en numérique ? Est-ce qu'on aura le temps de faire plus en numérique ? Il y avait un vrai échange sur leurs pratiques actuelles et comment ils allaient les faire évoluer. »

00.15.05s « On avait un rôle de coordination globale au niveau du programme et on a très vite pris un rôle justement de mobilisation des archivistes pour les emmener dans un même mouvement. Nous on y avait intérêt, parce qu'il fallait que notre logiciel s'adapte le mieux aux pratiques qu'ils mettaient en place. Et donc on avait intérêt à faire converger les gens vers une même pratique. Donc c'est pour ça qu'on a invité beaucoup d'archivistes en partant de sujets qui n'étaient pas informatique justement sur leurs pratiques : comment ils faisaient ? Est-ce qu'ils faisaient tous pareil ? Etc. Et comment nous on envisageait de le faire dans le numérique. En général il y avait une ou deux personnes dans un service qui était sensibilisée au fait mais qui pouvaient venir avec un autre archiviste qui pouvait témoigner des pratiques papier, et du coup on pensait à comment ça allait évoluer, et comme ça, ça permettait aussi dans leurs propres services de faire venir d'autres personnes sur ces sujets-là. »

Records in Contexts

00.16.26s « Je pense que Records in Contexts est quelque chose de beaucoup trop conceptuelle, trop complexe, et un objet technologique qui ne va pas forcément survivre dans les années à venir. Records in Contexts parie sur une technologie qui n'arrive pas à prendre. La technologie sémantique, c'est un rêve qu'on a depuis plus de dix ans, et qui ne prend pas. Dans le monde internet actuel, une technologie qui ne prend pas en dix ans, pour moi, elle est morte. »

Le standard XML

00.18.23s « XML c'est un moyen. Et ce moyen a été remplacé dans tous les systèmes modernes par JSON. L'XML n'existe quasiment plus dans les systèmes modernes. »

00.18.45s « Alors JSON, il y a de toutes petites nuances qui font qu'il est beaucoup plus facilement interprétable et utilisable par des systèmes d'information, et d'autre part c'est un format de description standard des données d'un langage qui a pris beaucoup d'importance qu'est Javascript. Donc le langage utilisé par les navigateurs. Vu que c'était plus facile à implémenter que le XML, que c'est beaucoup moins verbeux que le XML, que ça se lit plus facilement, que c'est plus léger à lire, ça a été adopté. Et du coup l'XML est tombé un peu en désuétude, mais je ne dis pas que JSON il faut investir dessus parce que dans vingt ans ce sera encore là. Les technologies bougent beaucoup. Et le RDF, le web sémantique, etc. c'est une technologie qu'on a essayé de lancer à un moment, ça ne prend pas dans les faits parce que certes il y a des gens qui décrivent leurs données en *linked data*, mais quasiment personne ne les utilise en *linked data*. »

00.20.10s « Le XML, le JSON ou le web sémantique ce sont des formats de description des données, pas d'utilisation programmatique. C'est-à-dire que quand c'est dans la machine en base ou autre, ça n'a plus cette forme-là. Typiquement aussi, exemple : Les archivistes quand le XML était l'alpha et l'oméga, on fait un pari d'avoir des bases XML. Ça s'est retrouvé être une catastrophe absolue, parce que les bases relationnelles existantes étaient nettement plus utilisées, étaient hyper optimisées, alors que les bases XML sont restées un tout petit marché de niches, avec des logiciels quasiment expérimentaux qui allaient de guingois. Dans le monde des bases relationnelles on ne répare plus les bases, ça s'autoentretient sans aucun problème. Donc il faut faire aussi attention au choix des technologies, d'investir sur une technologie en se disant que c'est l'alpha et l'oméga, ça n'est jamais le cas. A notre époque, les technologies durent dix ans et elles sont remplacées. C'est pour ça aussi qu'une technologie trop complexe, son temps d'adoption fait qu'elle n'est pas intéressante. »

00.22.29s « Typiquement ce qu'il faut voir c'est par exemple ce qu'il se passe à la Bibliothèque nationale, c'est pour moi complétement emblématique. Ils ont investi sur du graphe, c'est une catastrophe. C'est-à-dire que quand ils sont à un million d'objets, ils sont saturés. Un million d'objets d'archives, c'est rien. Vitam pour donner une comparaison atteint les milliards d'archives. C'est impossible de faire des milliards d'archives en graphe. Typiquement, un des

problèmes du graphe pour être très simple, c'est que ça ne se découpe pas en petites machines. Toute l'informatique des grosses volumétries actuelles tient par le fait qu'on arrive à découper le problème et à le répartir sur plein de serveurs. Un graphe hyper connecté, on ne peut pas le découper, parce qu'il n'y a pas de frontière où l'on peut dire : ça c'est d'un côté, ça c'est de l'autre. Et donc le répartir sur une puissance de calcul fragmentée comme ça pose énormément de problèmes et pour la moindre interrogation, on mobilise tous les serveurs. C'est pour ça par exemple que la BnF a fait une base complète graphe où les interrogations sont extrêmement rares et où ça peut mettre des heures à trouver un résultat. Et elle a fait une base réduite, avec quelques métadonnées par objets, où là on peut faire des interrogations en secondes. »

00.24.14s « La donnée sémantisée a un intérêt sur des référentiels limités, justement pour avoir à la fois la souplesse et ne pas avoir de problème de calcul. Ce que j'appelle des référentiels limités ça va être des noms d'œuvres, des noms d'auteurs, des périodes historiques, des choses comme ça qui peuvent être toutes liées et même dans le domaine des archives on pourrait avoir tout ce qui est grandes autorités, les entités administratives, tout ça, là-dessus, on peut avoir sans doute un référentiel sémantisé qui permet toutes sortes d'interrogations. Par contre dès que l'on va atteindre l'objet et la métadonnée de l'objet, on va être sur des volumétries qui empêchent le graphe d'être pertinent en termes de capacité technique derrière. C'est-à-dire que pour moi, du lien sur de la métadonnée d'entités, d'entités nommées, de structures, de dates, de périodes, etc. oui. Le lien sur de la métadonnée d'objet, je n'y crois pas. »

00.25.46s « [Sur l'incertitude des liens faits par l'Homme] Ça aussi c'est un des gros problèmes du lien sémantique. Comment on génère ce lien sémantique ? C'est une information utile en effet mais, en général, qui est donnée par l'Homme. Il y a une part de subjectivité : plus on va manipuler des concepts, plus c'est la représentation du concept qui va faire que. C'est tout un savoir-faire pour les archivistes : le décortiquer et le mettre en place dans des relations sémantiques, ça va sans doute dépendre de l'archiviste dont on va extraire l'information. »

00.27.32s « [Sur les exigences primordiales des bases SQL]. C'est ce qu'on appelle les propriétés ACID. Les propriétés ACID, ce sont l'atomicité, la cohérence, l'isolation et la durabilité. Et globalement c'est être sûr que l'information est toujours purement exacte, cohérente, etc. Dans les bases NoSQL, alors NoSQL ça ne veut justement pas dire que ce n'est pas un langage SQL, ça veut dire qu'elles n'ont pas ces propriétés ACID. Ce sont des bases massives où par

exemple, si on enregistre quelque chose et qu'on interroge tout de suite derrière sur cette chose, selon le serveur que l'on va attraper, on va avoir l'info ou on ne va pas l'avoir, et on considère que ce n'est pas grave, car on sait que une heure plus tard, on est sûr de l'avoir. Alors que dans le monde SQL, on acceptera pas de répondre à une question tant que l'on n'aura pas fait les écritures pour être sûr que l'on puisse répondre. En relaxant ce genre de contrainte, c'est comme ça que l'on peut faire de la donnée massive, et qu'on peut la répartir sur des tas de serveurs parce que l'on n'est pas obligé de toujours en permanence les synchroniser à la seconde près. »

00.29.42s « C'est-à-dire que, on va se donner un certain temps pour être sûr que l'information soit bien partout la même. Mais dans les archives, ce n'est pas grave si on a une heure avant de pouvoir interroger ce qu'on a rentré. Il [L'utilisateur] a une information qui est vieille d'une heure. »

00.30.29s « Les bases SQL, dès qu'on dépasse les millions, ça devient très compliqué voire impossible. Pourquoi ? Car une base SQL, pour garantir justement cette cohérence, elle a plutôt tendance à faire grossir le serveur, donc à avoir un moyen de calcul qui reste le même et qui est très puissant, alors que les bases NoSQL, ont beaucoup plus tendance à faire plein de petits serveurs. Et donc faire mille serveurs c'est facile, faire grossir mille fois un serveur c'est très compliqué, c'est même impossible. »

00.31.16s « Si on veut appliquer le RiC aux métadonnées d'objets, ça ne marchera pas, ça je peux vous le garantir, c'est une certitude technique actuelle. Ou alors il faudra faire des macros objets d'archives, c'est-à-dire que ce sera des métadonnées de cartons qu'il faudra gérer, mais certainement pas des métadonnées de documents. »

Records in Contexts

00.38.23s « Typiquement, ce que j'en dis et ce que j'en pense, c'est une réflexion sur la technologie sémantique plus que sur RiC en particulier. »

00.40.03s « Je pense qu'il faut que les archivistes pensent archivistique et qu'ils fassent appel à des informaticiens pour faire ce qui leur est vraiment utile. Et puis après qu'ils prennent en compte aussi les limitations parce qu'être dans le rêve qu'on va pouvoir faire du web sémantique sur tout, pour moi c'est vraiment un rêve. »

Sur un front-office Vitam

00.41.00s « Ca aussi ça avait été un parti pris de Vitam : c'est de dire on arrivera jamais à faire converger sur le front-office, ça va être trop compliqué. Donc on va s'occuper d'abord d'avoir un back-office qui tienne la route, qui soit vraiment une brique essentielle, que les gens construisent autour, et après, s'ils ont vraiment besoin qu'on les aide, ils viendront vers nous et c'est ce qui s'est passé. Et c'est comme ça qu'a été mis en route le projet Vas, donc Vitam as a service, enfin Vitam accessible en service. Et là il y a la mise en place de tout un front-office, et même carrément d'une offre de service. »

00.42.06s « Il n'y a pas de graphe de métadonnées dans Vitam donc ça m'étonnerait qu'il y ait une représentation graphe, par contre il y a des graphes sur les objets. Il y a des relations très libres dans Vitam entre les objets, c'est-à-dire qu'un objet peut appartenir à plusieurs hiérarchies. Il ne peut pas y avoir de boucle dans le graphe, par contre on peut croiser des hiérarchies, on peut avoir une hiérarchie fonctionnelle ou une hiérarchie par entité, etc. »

00.43.01s « On est sur une technologie *ElasticSearch* qui est la plus reconnue actuellement des moteurs de recherche et qui permet de faire des croisement quand même très intéressant.

XML, un obstacle pour les dernières évolutions en termes d'archivage numérique ?

00.43.41s « Le SEDA a été une aide et une complication. L'aide ça a été les métadonnées et leur structuration. La complication c'est qu'on a figé un modèle technique en XML. C'est exactement toujours la même idée, c'est-à-dire que le fait que le métier archiviste réfléchisse à : quelles sont mes métadonnées essentielles ? Ça oui. Mais qu'ensuite on fasse un modèle XSD [*XML Schema Definition* en anglais] pour dire : c'est comme ça ; c'est là que ça nous a cassé les pieds. Il y avait des façons plus efficaces de le faire. »

Un modèle mixte ?

00.45.04s « Pour des référentiels limités, ça peut être très efficace, par contre il ne faut pas rêver de l'avoir à l'objet, c'est pour ça que je vous parle du modèle mixte. C'est-à-dire que sur un certain nombre de métadonnées, on peut rester en milliers d'objets, ou en dizaine de milliers mais : des entités, des auteurs, des périodes historiques, des domaines d'intervention, etc. et croiser tout ça et être en capacité de faire des interrogations complexes, pourquoi pas. Je pense que ça a une réelle utilité. »

JSON

00.46.15s « Ça permet de décrire pour l'emmener ailleurs de l'information qui est dans une base de données ou documentaire ou relationnelle. »

Moteur de recherche et logique floue

00.47.55s « C'est là aussi qu'il faut faire attention au mythe de : on va avoir une information super structurée, parce que ça nécessite un travail humain énorme de structurer les choses. Ce qu'il faut bien voir c'est que tout ce qui est données massives actuellement dans le monde de l'internet, c'est de la logique floue. Ce que vous avez sur votre moteur de recherche Google, ça vous permet d'arriver relativement rapidement à l'information qui vous intéresse mais par approximation. Imaginez qu'on se dise : pour faire des recherches sur le web, il faut que l'on ait de la description sémantique de chaque site ; ce qui était le rêve d'origine je le rappelle. D'une part on aurait rien actuellement, et d'autre part il faudrait que chaque personne sur son site s'amuse à faire de la description sémantique de toutes ses pages. »

00.48.46s « Sur les archives, c'est pareil, si on veut avoir de la métadonnée sémantique, ça veut dire qu'il va y avoir un humain – alors pas toujours parce qu'il y a des choses qui sont sans doute automatisables – mais il y aura besoin quand même d'interventions humaines pour mettre de l'information sur, et donc on revient à ce que je vous disais sur la charge et sur le potentiel d'erreur de s'appuyer sur le moindre truc par des humains. Alors que sur une logique de moteur de recherche classique, on va prendre toute la métadonnée et on va pouvoir faire des recherches approximatives, parce qu'un modèle sémantique c'est très mal fait : pour lui « Mozart » avec un « t » ou « Mozart » sans « t », ce sont deux nœuds différents du modèle sémantique, et il ne va pas forcément chercher l'un et l'autre en même temps. »

Avantages de la sémantisation précise ?

00.50.12s « C'est la capacité de faire des requêtes conceptuelles complexes, encore faut-il savoir les exprimer aussi, ce qui n'est pas toujours simple, en tout cas, ce n'est pas en faisant du SPARQL, ce n'est pas Monsieur tout le monde qui va faire du SPARQL en tout cas. »

00.50.51s « Si on est sur de l'information plus structurée, non pas graphe, il faudra avoir prévu la clé de croisement pour pouvoir faire la recherche. »

Intérêt de la représentation sous forme de graphe d'un point de vue cognitif

00.51.40s « La représentation sous forme de graphe est très séduisante, mais elle ne tient que sur les petits corpus. Comment fait-on pour avoir une représentation en graphe pertinente sur un très gros corpus ? Il faut que la représentation puisse parler. Regardez un graphe : si vous avez dix nœuds, cent nœuds, mille nœuds, cent mille nœuds, un million de nœuds, ça devient un truc tout noir parce qu'il y a des traits dans tous les sens, vous comprenez rien. Ca veut dire que si on veut utiliser la représentation graphique pour qu'elle soit humainement pertinente, ça veut dire qu'il faut se mettre à un niveau suffisamment élevé de volumétrie pour avoir quelques dizaines de nœuds tout au plus qui puissent donner du sens à tout ça. Donc premier problème : comment je trie le bon niveau de pertinence d'information ? »

00.52.58s « En effet le mindmap typiquement c'est quelque chose de tout à fait intéressant et pertinent mais qui contrairement à ce qu'on pense est très souvent utilisé en modèle hiérarchique, parce que même s'il est rayonnant, le mindmap a un sens. »

00.53.21s « Comprendre un graphe qui est vraiment complètement connecté c'est assez complexe aussi, parce qu'on ne sait pas par quel bout le prendre justement. »

00.53.51s « L'approche graphique des choses peut amener des apports que l'on imagine pas. Je vais vous donner l'analogie : les archivistes qui ont fait ces représentations graphiques d'arborescences bureautiques, etc. se sont rendus compte qu'à l'œil elles arrivaient à attraper des choses par le graphisme, sans rien d'autre. Et peut-être que dans des graphes aussi on arriverait à trouver des concentrations de nœuds autour de certains concepts ou des choses comme ça, simplement par observation de représentations graphiques. »

Les outils de partitionnement de graphes

00.55.20s « Il y a tout ce que fait l'IA dans le domaine, il y a pas mal de choses en intelligence artificielle qui doivent exister. Utiliser l'IA pour faire du tri là-dedans est une bonne idée à la base. »

00.55.55s Fin de l'entretien.

1.4. Entretien avec Madame Florence Clavaud

1.4.1. Guide d'entretien

Thème	**Sous-thème**	**Temps**
Parcours	Études (diplômes, lieux, dates, stages)	
	Vous êtes actuellement membre de l'EGAD, également responsable de la mission référentiels et de projets numériques aux AN. Emplois précédents ?	
	Comment avez-vous rejoint l'EGAD ?	
	Intérêt personnel pour la description archivistique ?	
RiC-O	Comment avez-vous traduit le modèle conceptuel en ontologie ? (par quoi avez-vous commencé, méthodes employées)	
	Comment faites-vous pour aligner les concepts archivistiques avec les concepts des bibliothèques/des musées ? (quel procédé/outil/méthode)	
	Comment testez-vous l'ontologie pour savoir si elle est fonctionnelle ?	
	Comment le prototype du British Museum « Research Space » vous a inspiré pour RiC-O ? Quels points en particulier ?	
Représentation sous forme de graphe	Intérêt - compréhension des ressources archivistiques ?	
	Intérêt - exploitation de ces dites ressources ?	
	Du pdv de l'utilisateur, comment traiter de grands ensembles d'information ? Est-ce que des outils de partitionnement de graphes seront proposés ? Si oui, lesquels(les) ? (bipartition de graphes, méthode hiérarchique, algorithmes de *clustering*, etc.)	
RiC aux AN, PIAAF	Est-ce que PIAAF utilise toutes les fonctionnalités de RiC ? PIAAF a-t-il été pensé comme une évolution ou une extension de RiC ?	

	Au sujet de l'expérience PIAAF, vous dites avoir beaucoup appris de la pratique et du point de vue des institutions partenaires et *« pu mesurer combien ceux-ci déterminent les métadonnées créées »*[579], pouvez-vous expliciter ?	
	Piste de développement d'un sous ensemble de RiC-CM et de RiC-O adapté aux ressources conservées par les AN ? (Vers une évolution ou une extension de RiC ?)	
	Comment les AN utilisent les premières versions de RiC actuellement ?	
	Qu'est-ce que RiC a changé aux AN (SIA, habitudes de travail, pdv utilisateurs) ?	
	Comment fonctionne RiC avec des portails web de type FranceArchives ou le portail européen des archives ?	
	RiC et le portail FranceArchives : quel intérêt ? Est-il déjà utilisé ? Si oui, de quelle manière ? Mêmes questions pour portail européen des archives.	
RiC-O Converter	RiC-O Converter a pour objectif de convertir des notices d'autorité et des instruments de recherche en RDF/OWL. Comment faites-vous en revanche pour convertir des référentiels d'indexation en RDF/SKOS ?	
	Quelle est la politique d'URI pour les AN (stratégie de construction et de gestion d'URIs) ? Les AN doivent échanger des données sur les agents avec la base ISNI. Mais qu'est-ce qu'être membre du consortium ISNI apporte aux AN du point de vue de sa politique d'URI ?	
	Dans la présentation de RiC-O le 28 janvier 2020, il est dit au sujet des URI que *« des choix provisoires ont dû être faits »*, pouvez-vous expliciter ?	
	Dans la présentation de RiC-O du 28 janvier 2020, il est dit que « [RiC-O Converter] ne s'occupe pas d'autres formats de description, donc en particulier des manifestes	

[579] Florence Clavaud, Pauline Charbonnier, *Records in Contexts aux Archives nationales : enjeux et premières réalisations*, *op. cit.*, p. 22.

	SEDA utilisés par les SAE comme VITAM », pourquoi & comment surmonter ce problème ?	
Référentiels des AN	Situation actuelle	
	Est-ce que tous les référentiels (ou du moins les principaux) ont une structure hiérarchique (thésaurus) ?	
	Comment parvenir à une structure hiérarchique (par quel procédé/outil/méthode) ?	
	Quel est l'intérêt d'une structure hiérarchique (thésaurus) pour la consultation et l'exploitation des ressources archivistiques ?	
	Où en êtes-vous de la conception d'un référentiel des agents unique ? Quelles sont les étapes à suivre pour y parvenir ? (méthodes pour son enrichissement, la fusion des référentiels, etc.)	
	Au sujet du référentiel agents, vous avez dit lors de la journée d'étude du 28 janvier 2020 qu'il y avait une *« nécessité de faire évoluer les fonctionnalités, l'interface et le modèle de données »*, pouvez-vous expliciter ?	
	Où en est l'idée de développer un référentiel national couvrant les fonctions des collectivités territoriales ? (méthodes suivies pour sa mise en place ?)	
	Lors de la rencontre annuelle des correspondants archives en 2016, vous aviez dit que *« l'interopérabilité se construira aux Archives nationales quasiment exclusivement sur les référentiels »*. Pouvez-vous expliciter svp ?	
	Normalisation (normes sur lesquelles les AN s'appuient pour la production des référentiels)	
	Quel est le processus d'application de la norme ISO 25964 ?	
	Quels sont les résultats de l'étude menée courant 2020 pour évaluer les impacts d'une évolution du SIA des AN	

	vers le support du format SKOS ou d'un format compatible avec ISO 25964 ?[580]	
	Comment utilisez-vous la plateforme GINCO du ministère de la culture dans l'élaboration des vocabulaires contrôlés ?	
	Comment appliquez-vous la feuille de route sémantique du ministère de la Culture ?	
	Du point de vue de l'utilisateur, en quelle mesure la consultation directe des référentiels améliorerait-il la recherche (information retrieval + research) ?	
Projet ALEGORIA	L'enjeu de ce projet est de parvenir à lier des données RDF avec d'autres. Comment assurez-vous l'articulation entre les métadonnées des AN et celle de l'IGN (L'Institut national de l'information géographique et forestière) ? (méthode/outil/procédé)	
	Comment s'opère l'alignement des référentiels d'indexation dans le cadre d'ALEGORIA ? (étapes à suivre) Et pour lier les données des AN à celles de la BnF/Wikidata, etc ?	

[580] Florence Clavaud, « Transformer les métadonnées des AN en graphe de données : enjeux et premières réalisations », *art. cit.*, p. 81.

1.4.2. Transcription chrono-thématique

00.00.01s Début de l'entretien.

Parcours

00.00.23s « En quelques mots, je suis archiviste paléographe, j'ai fait mes études à l'Ecole nationale des Chartes et j'en suis sortie, après avoir fait le choix des archives, conservateur du patrimoine en 1990. Je suis aujourd'hui conservateur général. En 1990, mon premier poste a été celui d'adjoint au responsable du CARAN, le Centre d'accueil et de recherche des Archives nationales, où je suis restée jusqu'en 1999. »

00.01.14s « A partir des années 1995, comme le ministère de la Culture voulait expérimenter la numérisation des documents, j'ai commencé à m'intéresser de très près à la numérisation en montant les premiers projets de numérisation des Archives nationales. A partir de 1997/98, à m'intéresser à la description archivistique et aux métadonnées qui sont devenues un de mes domaines d'expertise. Donc j'ai travaillé à implémenter, promouvoir l'EAD aux Archives nationales : premiers appels d'offre de numérisation, d'encodage d'instruments de recherche, développement d'applications (dont un site web de recherche et de consultation d'instruments de recherche en EAD avec le logiciel PLEADE), développement d'une application sur l'EAC-CPF (projet ETANOT), etc. A partir de fin 1999 jusqu'à juin 2004, j'étais responsable informatique aux Archives nationales, je me suis occupée notamment de métadonnées, de bases de données, de fichiers EAD, de mises en ligne des premiers sites web des Archives nationales donc ETANOT, ARCHIM. »

00.02.19s « J'étais dès ce moment-là dans le groupe de travail de l'AFNOR sur l'EAD qui s'est occupé donc de traduire ce dictionnaire des balises. Puis également dans le groupe de travail de l'AFNOR sur les données d'autorité qui s'est occupé de la norme EAC . J 'ai donc commencé à faire du développement informatique. J'ai notamment aidé à développer le logiciel Pléade, un logiciel libre de publication des instruments de recherche en EAD. »

00.02.49s « Je suis partie des Archives nationales en 2004 pour passer dans le secteur privé pendant trois ans en disponibilité pour avoir une autre expérience. Donc j'ai été employée pendant trois ans, comme consultante (notamment en AMOA), également chargée de projets de normalisation de métadonnées archivistiques et de développement d'applications web, et formateur de la société Anaphore, qui est éditeur de logiciel de description archivistique

français. »

00.03.21s « En 2007, j'ai quitté la société Anaphore pour revenir en tant que fonctionnaire, comme j'étais en disponibilité dans le secteur public, mais cette fois-ci dans l'enseignement supérieur et la recherche puisque de 2007 à 2013 j'ai été d'abord responsable numérique à l'Ecole des Chartes, ce qui m'a conduite à m'occuper de plusieurs projets d'humanités numériques (comme les éditions des minutes notariales de P. Christofle ou des carnets de prison de Delescluze) ; certains d'entre eux sont encore en ours aujourd'hui, et je m'en occupe toujours en tant que membre de l'équipe de recherche de l'école (édition des chartes de l'abbaye de Saint-Denis ; édition du Petit Thalamus de Montpellier) . »

00.03.48s « J'ai été à partir de 2007 responsable du Master 2 *Technologies numériques appliquées à l'histoire* de l'Ecole des Chartes (que j'ai monté, dont j'ai assuré une moitié des cours et l'encadrement pédagogique complet, stages inclus, des étudiants pour 6 promotions). »

00.04.16s « J'ai quitté l'Ecole des Chartes en 2013 après cette aventure qui continue avec d'autres responsables, pour revenir aux Archives nationales comme responsable de la mission référentiels où je m'occupe donc aujourd'hui des données d'autorité et des vocabulaires des Archives nationales. »

00.04.42s « Donc ma tâche quotidienne, avec mon adjointe, est la gestion de la qualité (*quality management)* de ces référentiels (sans parler de support auprès de leurs auteurs, d'élaboration des méthodologies et règles de description, de l'évolution fonctinnelle du module informatique qui les gère). Je m'occupe aussi de tous les projets liés à la transition des métadonnées vers des graphes d'entités liées et à la trajectoire sémantique de l'établissement. »

00.05.02s « Depuis les années 2000, je suis donc membre de groupes de travail de normalisation à l'échelle nationale. Je suis également membre du groupe de travail de la *Society of American Archivists* sur les standards archivistiques encodés (TS-EAS) . Et depuis 2012, membre exécutif du groupe EGAD. »

00.05.24s « Pour résumer, ma trajectoire s'est construite beaucoup autour des questions de numérique, de l'application des technologies numériques aux archives, qu'il s'agisse de métadonnées, de modélisation, mais aussi d'édition critiques de documents, de développement web. »

00.05.48s « Je connais très bien les technologies XML, je suis développeur, et je suis aussi maintenant – pour y avoir passé énormément de temps depuis plusieurs années – spécialisée dans les ontologies et l'ingénierie des connaissances. »

Records in Contexts

00.06.43s « *Records in Contexts* c'est un modèle conceptuel et une ontologie, c'est les deux. Plus tard, ce sera un manuel d'application. Pour l'instant les manuels, on ne les a pas encore réalisés puisqu'on est toujours en train de faire les deux premières parties. »

00.07.04s « Quand le groupe de travail s'est formé en 2013, j'y suis entrée parce qu'on est venu me chercher au titre de mon expertise, comme on est venu chercher les autres membres. Donc quand ce groupe s'est monté, il a assez vite décidé de faire ces deux choses : un modèle conceptuel et une ontologie, parce qu'il a constaté que non seulement il manquait un modèle conceptuel pour la description des archives, mais aussi qu'il manquait, pour le web de données, un bon modèle de représentation des métadonnées archivistiques alors que d'autres communautés en avaient, ce qui faisait qu'on avait un vide, quasiment un vide sidéral, dans le web de données sur les métadonnées archivistiques, donc on s'est dit : on va faire les deux. Donc la décision de faire les deux a été prise très tôt dans l'histoire du groupe EGAD, dès l'automne 2013, et dès ce moment-là, on a commencé à élaborer les deux parallèlement. On n'a pas traduit en OWL le modèle RiC-CM, on a construit les deux selon une boucle d'itération. Et je dirais même que RiC-O a servi de laboratoire pour RiC-CM, parce que quand vous élaborez une ontologie, vous devez définir très précisément les objets dont vous considérez qu'ils font partie du monde que vous décrivez. Ça vous oblige à conceptualiser, à définir leur positionnement les uns par rapport aux autres, à réfléchir formellement aussi aux notions de sous-catégories d'objets, de propriétés. Et donc on a commencé à travailler sur l'ontologie dès le tournant de l'année 2014, alors que parallèlement on travaillait aussi sur le modèle conceptuel, et les discussions que l'on a eues sur la création de l'ontologie ont nourri le modèle conceptuel, et vice versa. »

00.10.20s « L'ontologie c'est - au départ – pour une très grosse part moi qui l'ai élaborée. Je contribue aussi beaucoup au modèle conceptuel. Le président du groupe EGAD (Daniel Pitti) m'a demandé de prendre la responsabilité du développement de RiC-O ; un petit groupe de membres d'EGAD, ayant quelques compétences techniques, s'est donc constitué pour s'en occuper, que j'ai piloté ; j'ai concrètement fait l'essentiel des versions actuelles de RiC-O, en concertation étroite avec les membres de ce groupe. J'ai notamment beaucoup travaillé sur les relations puisque dans une ontologie vous devez définir des propriétés qui sont des relations, et ce que j'avais posé sur la table, ce qu'on a ajouté à l'ontologie, a beaucoup servi ensuite pour la réflexion sur les relations dans le modèle conceptuel. Bien sûr, au fur et à mesure que le modèle conceptuel mûrissait, on a également ramené dans l'ontologie, par exemple, des classes d'objets qui étaient définies dans le modèle conceptuel. Donc par

exemple, dans l'ontologie, toutes les classes d'objets qui ont leur correspondant dans RiC-CM ont leur définition textuelle identique à celle du modèle conceptuel. Donc c'est pour ça que l'on parle d'une transposition. »

00.11.35s « L'ontologie est beaucoup plus précise et riche que le modèle, ce qui est très classique, ce n'est pas du tout une particularité de RiC-O. Quand vous faites une implémentation - une transposition technique d'un modèle conceptuel - vous êtes amené à être plus précis. »

00.11.59s « On pourrait retourner la définition en disant que ce n'est pas une transposition du modèle conceptuel, mais on pourrait même carrément aller jusqu'à dire que le modèle conceptuel est un sous-ensemble de RiC-O, on peut inverser la logique, d'un certain point de vue, ça dépend comment vous vous positionnez. Mais en tout cas, la construction des deux s'est faite de façon extrêmement articulée. »

00.13.00s « Le groupe EGAD, sa première réunion a été en novembre 2013 à Bruxelles, le temps de monter le groupe, donc la chose qu'on a faite c'est définir ce que l'on voulait faire. Si vous voulez quand vous venez, vous êtes vingt personnes qui viennent du monde entier, qui ne se connaissent pas, qui sont réunies pour faire ce genre de travail, il faut commencer par trouver des points de consensus pour savoir ce qu'on veut faire, donc une grosse partie des journées que l'on a passé ensemble a été consacrée à faire connaissance les uns avec les autres et à poser nos objectifs. Donc c'est là que l'on a décidé – parce que le CIA ne nous avait pas au départ demandé de construire une ontologie – de construire une ontologie, ce qui est devenu aujourd'hui l'un des mandats officiels de l'ICA. On nous avait donné un mandat très large : réarticuler les quatre normes existantes, définir un nouveau modèle. Nous on a considéré qu'un modèle conceptuel ne pouvait pas suffire, et qu'il fallait aller jusqu'à combler le vide existant ; autant, proposer une transposition technique du modèle conceptuel immédiatement, sinon un modèle conceptuel risque de ne jamais être appliqué, ça c'est l'autre problème. »

Au sujet de l'organisation de l'EGAD

00.14.49s « On a une liste de discussion qui nous permet de discuter tous les jours ou presque sur certains sujets. On a des conférences en visio tous les quinze jours. Et puis on a quand on peut – c'est vrai que l'année dernière on n'a pas pu à cause du Covid – on a une réunion annuelle autour d'une table de trois à quatre jours pleins. »

00.16.14s « Absolument toutes ces discussions – écrites, orales – se font en anglais, toute notre production est en anglais, parce que c'est la seule langue compréhensible

par tout le monde dans le groupe. Donc ce sont des discussions et des réunions qui demandent un effort conceptuel, un effort intellectuel, doublé d'un effort linguistique pour beaucoup d'entre nous, et sur un sujet compliqué. C'est toute une aventure avec des personnes qui ont des points de vue, des expertises, des expériences différentes. Il faut arriver à monter un modèle générique qui puisse être appliqué dans n'importe quel pays, dans n'importe quelle culture, pour n'importe quel type de document. »

Comment testez-vous l'ontologie pour savoir si elle est fonctionnelle ?

00.17.44s « C'est un travail de conceptualisation qui est lié à nos compétences et à notre expertise – je parle du groupe –, on est tous engagé dans le métier d'archiviste depuis des années voire des décennies, donc on s'appuie sur notre expérience, sur notre formation, sur ce que l'on fait dans nos institutions, sur ce que l'on sait de l'archivistique, sur avec une littérature sur la description archivistique qui est très abondante, avec des écoles d'archivistiques différentes dans le monde : vous avez l'école anglo-saxonne, vous avez l'école australienne, vous avez l'école européenne. »

00.18.46s « Après, pour faire des tests il faut implémenter. Les réflexions théoriques sont intéressantes mais il faut faire des tests, c'est pour ça que l'on a eu besoin de RiC-O. »

00.19.05s « Pour tester - alors je vais vous faire une réponse à l'échelle des Archives nationales françaises - nous on a bien évidemment testé RiC-CM, puisqu'on a testé RiC-O : on a converti à l'heure actuelle l'ensemble de nos métadonnées, donc 29 000 instruments de recherche en EAD, 15 000 notices d'autorité, plus tous nos vocabulaires en RDF conformément à RiC-O ; comme tout ça fonctionne très bien, c'est une preuve d'applicabilité puisque l'on est à l'échelle de l'institution ; ce n'est pas un petit projet, ce n'est pas sur un sous-ensemble qu'on aurait choisi. »

00.21.22s « Le caractère fonctionnel d'un modèle, ce n'est pas seulement être capable de l'utiliser pour des métadonnées dans une institution, c'est aussi voir ce que ça apporte, en quoi il est intéressant fondamentalement. Il est intéressant parce qu'il permet d'être plus précis, plus nuancé, d'avancer vers la multidimensionnalité, de sortir du carcan des instruments de recherche où on a une bidimensionnalité essentiellement : une hiérarchie et des séquences, pour faire quelque chose de beaucoup plus contextualisé, dans lequel on peut entrer par un nombre beaucoup plus important de portes, en posant des questions que l'on ne peut pas poser dans des systèmes d'information classiques. »

00.23.55s « Avec RiC, on passe à la multidimensionnalité, donc vous avez toutes les entités de contexte qui sont définies comme des entités à part entière : des agents, des fonctions, des lieux, des évènements, des activités. C'est pour cela que l'on parle de graphes. Donc ça permet fondamentalement à l'utilisateur ou à l'archiviste, à n'importe quel utilisateur final de ces métadonnées d'entrer – c'est une image – mais d'entrer par beaucoup plus de portes. Et ça permettra aussi, en construisant des graphes d'entités, d'interroger les relations, de parcourir le graphe *via* ces relations, chose que l'on ne peut absolument pas faire dans un système d'archivistique classique, tel que celui des Archives nationales par exemple. Dans le SI des AN, à la limite, vous voyez une toute petite partie du graphe, mais vous ne pouvez pas l'interroger, vous ne pouvez jamais voir l'ensemble des entités qui entourent un document et encore moins interroger les relations. »

00.25.35s « Ça crée les conditions de la multidimensionnalité et ça permet aussi d'être beaucoup plus précis et nuancé puisque l'on a un niveau de précision qui est plus important que dans les précédentes normes et qui est accru dans l'ontologie, avec une flexibilité aussi qui est conservée, c'est-à-dire que rien ne vous empêche malgré tout de décrire les documents de façon très rapide et les entités de contexte de façon très sommaire, mais si vous le souhaitez, vous pouvez les décrire de façon extrêmement précise. C'est toute la nature d'un modèle ontologique que de permettre ça : la flexibilité, ce qui va très bien avec la description archivistique d'ailleurs puisque par essence, elle est flexible, le niveau de granularité n'est jamais prédéfini. C'est en fonction des moyens, des besoins, des types de documents, que vous allez vous-même définir si vous faites une description très précise, si vous descendez au niveau de la pièce ou si vous vous arrêtez à la série, à la sous-série, au dossier, et si vous avez les moyens de décrire un agent précisément, avec toutes les caractéristiques qui permettent de comprendre notamment quels rôles il a joué, quelles productions documentaires il a pu avoir, quels échanges documentaires il a pu faire avec d'autres. C'est ça l'essence de la description des agents en archivistique, c'est de comprendre comment les agrégats documentaires se sont formés, comment ils se sont transmis. »

00.27.40s « La fonctionnalité renvoie non seulement au fait que ce soit utilisable, donc c'est ce que je vous ai dit tout à l'heure, on peut prouver que RiC est utilisable, aujourd'hui on commence à le prouver clairement, mais aussi au fait que c'est intéressant, que c'est utile. D'ailleurs, c'est dans les principes de l'ontologie. Si vous lisez les *design principles* de RiC-O (https://www.ica.org/standards/RiC/RiC-O_v0-2.html#design-principles) vous avez : une ontologie de domaine utilisable et fonctionnelle. Utilisable, ça veut

dire qu'on peut l'appliquer aujourd'hui, c'est très important, on n'est pas obligé d'attendre d'avoir opéré des transformations dans ces données. »

00.28.35s « On peut prouver que c'est utilisable mais aussi utile, qu'on gagne quelque chose, parce que faire du travail de transformation, c'est aussi coûteux. Et, par ailleurs, bien sûr c'est immédiatement utilisable mais ce sera encore plus intéressant si on fait subir aux métadonnées existantes un certain nombre de traitements d'amélioration de la qualité. »

00.29.42s « Donc tout ça ce sont des cadres de travail, et c'est à chaque institution qui le souhaite de s'en emparer. C'est pour cela qu'il y a nécessité d'avoir des manuels d'application génériques, mais c'est aussi à chaque institution d'en faire ce dont elle a besoin, en fonction de son corpus de métadonnées, en fonction de ses moyens et surtout de ses objectifs. Le pari c'est que notamment RiC-CM peut aider à repenser entièrement sa manière de voir les métadonnées. C'est un modèle très inclusif, global, dont le périmètre couvre autre chose que les instruments de recherche classiques. Il couvre aussi par exemple la description des reproductions numériques des documents, etc. Ce qui peut aider à décloisonner les silos de métadonnées que l'on a tous dans nos institutions ; où on a plein d'instruments de recherche certes mais où on a aussi plein de bases de données, on a une bibliothèque numérique, on a un système d'archivage électronique, tout ça, ça vit quasiment en vase clos. Il y a très peu d'interactions entre ces choses-là, or RiC-CM en propose. Il indique par exemple que les images numériques des documents sont des instantiations. On peut les considérer comme des instances de la classe de l'entité instanciation, des entités autonomes, à part entière, qui ont leur statut, et qu'on peut relier avec les documents. Alors qu'aujourd'hui vous avez très souvent dans les bibliothèques numériques des collections d'images auxquelles on n'a pas affecté de statut. C'est un des intérêts pour la profession, pour le métier, après vous avez les intérêts pour les utilisateurs finaux. »

Au sujet de la plateforme d'interrogation et de visualisation

00.32.16s « Pour l'instant, on n'a pas d'application de visualisation. On a des données RDF en très grandes quantités puisque l'on a 155 millions de triplets, ce qui est massif, mais on n'a pas implémenté encore dans notre système d'information la brique qu'il faut pour les stocker, les rendre cherchables, et les consulter. Ce n'est pas seulement un problème de visualisation, c'est une infrastructure qu'il faut changer. Il faut y rajouter des modules. Donc ça suppose d'implémenter une base RDF, ce qu'on appelle un *triplestore*, avec par-dessus le protocole de recherche de SPARQL et ensuite une interface web de recherche et de consultation. Donc ça, ce n'est pas fait, on est en train de construire des petits

morceaux de ça. On a fait une preuve de concept, on a fait un prototype avec PIAAF. »

00.33.28s « On va avoir, probablement à la fin de l'année, un démonstrateur de taille beaucoup plus significative. C'est-à-dire qu'on pourra y faire des recherches avec SPARQL sur les jeux de données RDF des Archives nationales, mais pas tous les jeux de données, une partie conséquente mais pas complète, ce sera juste un démonstrateur. Et à terme il y aura effectivement, à tout le moins je l'espère et je travaille dans ce sens, un module qui exposera l'ensemble de nos métadonnées en RDF RiC-O. »

00.34.17s « [Au sujet de *data.bnf.fr*] C'est une application entièrement en RDF donc nous, c'est un peu ce que l'on voudrait faire pour les Archives nationales et avec le modèle RiC-O. »

Le portail FranceArchives

00.34.40s « Le portail FranceArchives qui est l'agrégateur de métadonnées archivistiques pour le réseau français d'archives publiques, a choisi de passer au sémantique et à RiC-O. Et donc là on sera en face d'une application de taille conséquente, puisqu'elle va permettre d'interroger des métadonnées RDF de l'ensemble des services d'archives français. Donc de taille encore plus conséquente que notre propre application mais elle n'est pas faite, il va leur falloir deux à trois ans pour la monter. »

00.36.24s « FranceArchives sera dans deux/trois ans un bel exemple d'utilisation de jeux de métadonnées massifs de RiC-O. »

Projets externes aux Archives nationales

00.36.54s « Nous aux Archives nationales, nous avons transformé toutes nos métadonnées EAD, EAC-CPF et tous nos vocabulaires en RiC-O, mais nous n'avons pas d'application pour les montrer. Par contre, nous commençons a également fournir du RDF à des projets de recherche externes. Nous travaillons avec des équipes sur des projets de recherche scientifiques, dans le cadre notamment du projet ALEGORIA, où nous fournissons depuis plus de 1 an des données RDF. Et d'ailleurs l'IGN – l'Institut géographique national – qui est le pilote de ce projet – ALEGORIA – utilise aussi RiC-O pour ses métadonnées dans le cadre du projet. Donc nous n'avons pas d'application pour les Archives nationales mais nous commençons à fournir du RDF RiC-O à des projets externes. »

00.38.04s « On est dans une situation où, même si les résultats à très grande échelle, à échelle réelle, ne sont pas encore visibles pour le grand public, pour certains projets de recherche, ils le seront prochainement. »

PIAAF

00.38.45s « Nous [les Archives nationales] sommes partis un peu les premiers. Le prototype PIAAF qui a été publié en février 2018, n'a pas d'équivalent encore dans le monde. Il était à l'époque le premier au monde. On est partis parmi les premiers, tout simplement – c'est un peu un hasard - parce qu'il se trouve que je suis membre du groupe EGAD, principale auteur de l'ontologie, et également à un poste qui s'occupe de qualité des métadonnées aux Archives nationales. »

00.39.21s « Je reviens aux tests : j'ai absolument voulu utiliser les projets des Archives nationales également pour tester RiC. Je n'ai pas fait ça uniquement que pour les Archives nationales, je l'ai aussi fait dans l'idée que ça pouvait nourrir la réflexion sur RiC, et d'ailleurs, ça a été le cas. Je me suis servie de ce que l'on a fait aux Archives nationales depuis 2016 pour travailler sur l'ontologie. Les problèmes que j'ai rencontrés, les besoins qu'on a eu à traiter nous ont servi pour développer une ontologie répondant à ces besoins. »

00.40.20s « Ça nous a servi [le projet PIAAF pour améliorer RiC-O] pour établir une liste des relations dont on avait besoin entre les agents par exemple, même si aux Archives nationales on avait déjà une bonne réflexion bien avant. Ce n'est pas PIAAF qui nous a fait entrer là-dedans. On a 15 000 notices d'autorité qui sont reliées entre elles donc on avait déjà pas mal réfléchi à ces sujets. Mais le projet PIAAF, qui inclut quand même en majorité des notices d'autorité, nous a permis de travailler pas mal sur la partie agent du modèle et de l'ontologie, agent avec toutes les relations. »

00.41.03s « PIAAF c'est un prototype qui a été publié en février 2018, donc on l'a réalisé en 2017, mais à cette époque on avait une ontologie qui était à un stade beaucoup moins avancé. Donc depuis elle a énormément évolué, ce qui fait que PIAAF n'est pas conforme à la dernière version en date de RiC-O (RiC-O à.2), mais il était conforme à RiC-O à l'époque. »

00.41.55s « Une ontologie c'est un système poly-hiérarchique de classes et de propriétés, et donc vous avez une logique d'héritage. Si vous avez par exemple une classe *agent*, toutes les propriétés de cette classe vont être héritées par ses sous-classes. Donc dans l'ontologie, vous avez la classe *Agent* et vous avez une sous-classe *Person*, qui est la transposition de l'entité *Person* de RiC-CM. Dans le modèle on dit que l'entité *Person* est une sous-catégorie de l'entité *Agent*, dans l'ontologie l'entité *Person* est une classe, sous-classe de *Agent*. Et les propriétés que l'on a associé dans l'ontologie à la classe *Agent*, sont également utilisables sur la classe *Person*, qui par ailleurs a d'autres propriétés. Donc ce que je veux dire, c'est que quand vous mettez en œuvre une ontologie, soit elle vous convient parfaitement, c'est-à-dire que rien ne manque, soit vous

avez des besoins beaucoup plus spécifiques que ce que vous dit une ontologie générique et internationale, et donc vous devez faire ce qu'on appelle : étendre l'ontologie. Mais pour l'étendre, vous devez avoir une ontologie bien faite, qui vous propose des classes de haut niveau. »

00.44.50s « Si l'ontologie est bien faite, on pourra définir des sous-classes et des sous-propriétés parce que c'est aussi valable pour les propriétés. Si vous avez une relation qui est définie dans l'ontologie et qu'elle ne vous suffit pas, qu'elle n'est pas assez précise, vous pouvez définir une sous-propriété de cette relation. Donc ça, ça s'appelle : étendre. »

00.45.08s « Dans le projet PIAAF, on a étendu l'ontologie, mais on n'a aussi utilisé qu'une partie de l'ontologie, parce qu'elle est riche et que le modèle RiC-CM est également riche. Donc si vous voulez, par exemple, on n'a pas utilisé la classe *Rule*, et aujourd'hui il y a des classes que l'on n'utilise pas dans nos données RiC-O. Donc quand on utilise des modèles comme ça, en général on prend un sous-ensemble mais parfois on est en même temps amené à l'étendre. »

00.45.46s « Je reviens à ces normes internationales, c'est très classique. Je veux dire que quand vous utilisez ISAD(G), en général, ce n'est pas ISAD(G) qui va vous aider dans le détail à savoir comment vous allez décrire vos cartes et plans. Il va bien falloir que vous descendiez un peu plus dans le détail. C'est exactement le même problème. Nous notre ambition, ce n'était pas de faire une ontologie qui s'applique aux besoins des archivistes français, ou des Archives nationales ou des archives américains ou du projet SNAC [Social Networks and Archival Context], ou que sais-je, c'était de faire un modèle ou une ontologie qui s'applique à n'importe quel service d'archives, à n'importe quel document. Donc on est à un niveau d'abstraction relativement élevé. C'est aussi ce qui va faire l'avenir de ce modèle et de cette ontologie : plus on aura d'utilisateurs, plus ces gens-là vont nous dire : là il nous manque quelque chose. Ou bien on peut aussi avoir des gens qui disent : oui mais cette relation-là, finalement nous on n'est pas d'accord, vous ne l'avez pas défini comme il faut, on préfèrerait qu'elle soit définie comme ça, concrètement dans le vie réelle on aurait plutôt besoin de ça que de ça. Ou alors encore une fois : on a besoin de plus. A ce moment-là - c'est ce que l'on espère - c'est que le feedback permette d'améliorer le modèle et l'ontologie, soit en ajustant ce qui existe déjà, soit en ajoutant des choses. Mais si ce dont vous avez besoin d'avoir n'est pas dans l'ontologie, elle est censée vous permettre de l'ajouter en créant votre ontologie. »

00.47.43s « Pour l'instant aux Archives nationales, on n'a pas eu besoin de définir de sous-classe et de sous-propriété. On en aura peut-être besoin quand on va travailler beaucoup plus précisément à la description de certains types de

documents, et je pense aux documents d'archives nativement numériques. Mais on n'aura pas forcément besoin de beaucoup d'ajouts. »

00.49.21s « La conformité à RiC se mesurera au fait qu'on aura étendu correctement le modèle en définissant des sous-propriétés ou des sous-classes et pas en ajoutant des choses sans les raccrocher au reste. Si vous définissez une sous-classe d'une classe RiC-O, comme les entités en question seront également des sous-entités de la classe RiC-O, vous restez *RiC-O compliant*. »

Du brouillon à la version officielle

00.50.06s « Donc tout l'enjeu aussi – pour revenir au groupe EGAD – c'est (parce que nous avons diffusé la version 0.2 de RiC-O en février dernier (https://www.ica.org/standards/RiC/ontology) et nous nous apprêtons à diffuser la version 0.2 de RiC-CM au mois d'avril (ou de mai), donc les deux sont conformes l'un à l'autre) ,une fois que l'on aura fait ça, de faire un appel à commentaires pour recueillir – on l'a déjà fait une fois – les avis des uns et des autres, de prendre en compte ces remarques pour fabriquer une version 1.0. »

00.51.20s « L'ontologie c'est un brouillon, et le modèle conceptuel c'est encore un brouillon. Mais on veut passer à une version 1.0, c'est-à-dire une version officielle, une recommandation normative, qui ne soit plus un brouillon, à l'automne, et entre les deux, on va essayer notamment de prendre en compte les retours qui vont nous être faits. »

00.51.54s « Ce n'est pas contradictoire de dire : on va utiliser un sous-ensemble et on va l'étendre. C'est complétement normal dans cette logique d'appropriation. Mais ce n'est possible que si les modèles sont bien faits. Donc le fait que ce soit possible valide aussi si vous voulez d'une certaine manière le modèle, parce que si on disait : je n'arrive pas à étendre l'ontologie, ça ne marche pas, c'est mal fait et je ne peux pas raccrocher mon besoin à ce qui est déjà décrit dans l'ontologie, ça serait assez problématique pour RiC. »

00.52.43s « Cette année notamment on va travailler avec les collègues des Archives nationales qui s'occupent de la description des archives numériques, et on va approfondir le sujet de la mise en œuvre de RiC, et il en sortira peut-être des extensions de RiC-O. »

Description des archives numériques

00.53.21s « Il y a aussi le dictionnaire de données PREMIS, qui est un petit peu un modèle de référence dans l'élaboration du SEDA. »

00.53.44s « Il y a un format d'échange qui a été défini par la communauté française qui s'appelle le SEDA, qui est formalisé sous la forme d'un schéma XML ; ce n'est

pas une ontologie. D'ailleurs il a été transposé en norme ISO ce format. Il sert à définir comment on échange des informations entre le service versant, le SAE – le système d'archivage électronique -, les producteurs, les gestionnaires, etc. ets il embarque des métadonnées de description des archives. La partie métadonnées de description et la partie métadonnées de conservation s'appuient un peu sur le modèle de référence PREMIS, qui est un modèle pour les métadonnées de préservation. Donc c'est un modèle qui se situe dans un périmètre un peu différent des métadonnées de description dont s'occupe RiC. Donc ce que je voulais dire, c'est que l'on va travailler avec les personnes qui travaillent sur les archives numériques aux Archives nationales, à voir comment les métadonnées de nos fichiers SEDA peuvent être formulées conformément à RiC, et à RiC-O de préférence. Si on trouve que dans ces métadonnées SEDA il y a des choses qui sont conformes au modèle PREMIS, on n'ajoutera pas dans RiC-O des choses qui doubleraient PREMIS. C'est ce qu'il me fait dire que, *a priori*, il n'y aura pas tant d'évolution que ça dans RiC-O sur la partie archives numériques. C'est un modèle pour la description, pas pour les métadonnées de préservation dont s'occupe PREMIS. »

00.55.56s « A la base, RiC c'est un modèle qui s'applique à tous types d'archives, y compris nativement numériques, mais qui se concentre sur leur description, et pas sur les métadonnées techniques ou de préservation, qui sont le propos de PREMIS. Et dans le SEDA vous avez à la fois des métadonnées de description, des métadonnées de préservation, des métadonnées techniques, etc. donc on va regarder ce que l'on peut faire avec les fichiers SEDA et RiC. »

RiC-O Converter

00.56.27s « D'ailleurs je ne sais pas si Thomas vous en a parlé mais ce qu'on envisage – mais ce n'est qu'une idée parce que pour l'instant il faut trouver les moyens financiers, il faut aussi trouver une communauté d'intérêts, il faut définir un projet – on envisage d'ajouter à RiC-O Converter de quoi transformer les fichiers SEDA. »

00.57.17s « Tous les archivistes français utilisent l'EAD d'une manière ou d'une autre, soit dans des bases de données qui font des exports, soit nativement. Donc c'était vital d'avoir un outil de conversion de fichiers EAD. On convertit aussi des fichiers EAC-CPF. Donc ça c'était le point le plus important, la priorité. »

00.58.06s « L'autre raison pour laquelle on n'a pas travaillé sur le SEDA, c'est que à l'époque, quand on a lancé le projet, le système d'archivage électronique venait à peine d'être déployé donc nous n'avions pas accès aux fichiers SEDA des Archives nationales. »

00.58.34s « RiC-O Converter est un logiciel libre. Il peut être adapté par n'importe qui ;

on l'a fait pour les Archives nationales mais n'importe quel service d'archives français peut l'utiliser, même des services d'archives extérieurs, même des projets internationaux. C'est un logiciel *open source*. »

00.59.10s « Développer RiC-O Converter ça demande un financement, il faut le chercher. En l'occurrence c'est le ministère qui nous l'a payé ce logiciel, ce n'est pas les Archives nationales. »

00.59.20s « Dans une deuxième phase, on envisage en priorité numéro un – mais ça on va le faire tout seul sans demander de financement je veux dire, on va le faire nous-même avec Sparna – de faire passer RiC-O Converter à la version 0.2 de RiC-O, puisque RiC-O Converter utilise la version 0.1 de RiC-O et pas la 0.2 qui vient juste d'être publiée. Et puis ensuite, on va essayer de trouver les dispositifs, les moyens financiers notamment, pour convertir des fichiers SEDA, les fichiers SEDA des Archives nationales, parce que par ailleurs, SEDA peut être utilisé de plein de manières différentes, donc nous nous avons des profils d'application SEDA qui ne sont pas ceux d'autres services d'archives. »

01.00.54s « Les mappings qui disent pour chaque élément EAD à quoi ça correspond dans RiC-O (voir https://github.com/ArchivesNationalesFR/rico-converter/tree/master/ricoconverter/ricoconverter-doc/src/main/resources) ne sont pas attachés à RiC-O Converter. Ils sont utilisables par d'autres personnes qui ne voudraient pas utiliser RiC-O Converter. Ce sont des fichiers Excel qui déterminent pour un élément EAD donné, ce que vous pouvez faire avec RiC-O, comment RiC-O vous propose de le traiter, et c'est pareil pour EAC-CPF. Donc c'est une documentation qu'on a rendu publique et qui peut servir à d'autres projets, et en particulier FranceArchives dans leur travail de préparation d'une application. FranceArchives en RiC-O va se servir probablement de ces mappings. »

Référentiels

01.03.28s « Les référentiels aux Archives nationales ce sont des notices d'autorité et des vocabulaires actuellement. Ce sont des données d'autorité, c'est pour cela que je parle d'*authority data*. Ce sont des données que l'on agrège et que l'on gère sur des agents, sur des lieux, sur des mots-matière, sur des types de documents et d'autres catégories d'entités. Une partie de ces données actuellement est stockée en EAC-CPF, c'est le référentiel des producteurs. Et le reste est actuellement stocké dans une DTD maison. Nous, on a nous-même, mon adjointe et moi-même, puisque je fais du développement, on a nous-même converti ces données en données RDF conformes avant toute chose à RiC-O. Et pour certains des vocabulaires, on utilise également, en même temps, SKOS. C'est nous-même qui l'avons fait, hors du logiciel. Le

raisonnement c'était de dire : à part les notices d'autorité qui sont encore une fois prises en charge par RiC-O Converter puisque nos notices d'autorité sont en EAC-CPF, donc le référentiel producteurs des Archives nationales est pris en charge par RiC-O Converter il reste tout le reste. Notamment nous avons 60 000 lieux qui sont décrits dans des notices, nous avons plusieurs centaines de mots-matière, etc. donc pour ces choses-là, le modèle que nous utilisons actuellement c'est une DTD maison, conforme ni à RiC-O, ni à SKOS, ni à la norme sur les thésaurus ; ce n'est pas assez riche, on a beaucoup de problèmes avec cette structure de données, à tel point que l'on a commencé à construire d'autres versions de ces référentiels à l'extérieur de notre système d'information, pour pouvoir les enrichir correctement. Donc ce que je veux dire, c'est que c'est une DTD qui va tomber dans quelque temps, on va avoir besoin de la remplacer. Donc on n'avait pas envie de se lancer dans l'écriture d'un logiciel pour faire des mappings à partir d'un modèle qui ne convient pas, qui n'est pas destiné à être pérenne, qu'on va remplacer. Par ailleurs, dans le marché que l'on a passé avec Sparna, on avait la somme nécessaire pour faire de l'EAD et de l'EAC-CPF mais pas pour faire autre chose. »

01.06.33s « Nous utilisons nos propres scripts, nos propres dispositifs pour les convertir [les référentiels] en RiC-O. Après, ce n'est pas parce qu'on les a converti en RDF que la chose est finie. Donc encore une fois, on les enrichit en externe : on travaille à les aligner à d'autres jeux de données. Par exemple pour les lieux on a fait pas mal de choses là-dessus. Donc pour revenir au RDF aux Archives nationales, toutes nos métadonnées de description sont actuellement converties en RiC-O actuellement. Notre référentiel des lieux tel que nous l'avons dans le système, on l'a converti en RiC-O. Les lieux qui étaient définis par de toutes petites notices, sont devenus des instances de la classe *place* de RiC-O. Les types de document – on a un petit vocabulaire pour les types de document – sont devenus des instances de la classe *DocumentaryFormType* mais en même temps des instances de la classe *skos:concept*. Donc on utilise à la fois RiC-O et SKOS pour certains de nos référentiels. »

01.08.35s Fin de l'entretien.

1.5. Entretien avec Monsieur Edouard Vasseur

1.5.1. Guide d'entretien

Thème	**Sous-thème**	**Temps**
Parcours	Études (diplômes, lieux, dates, stages)	
	Professeur d'histoire des institutions, de diplomatique et d'archivistique contemporaines. Emplois précédents ?	
	Comment avez-vous rejoint le programme VITAM ?	
	Pourquoi avoir rejoint le programme VITAM ?	
	Qu'avez-vous appris de cette expérience ?	
Programme VITAM – Equipe & méthodologie	Quelles ont été vos missions en tant que responsable fonctionnel du programme VITAM ?	
	Une des spécificités de VITAM est d'avoir été développé à la fois par des archivistes, des records managers et des informaticiens. Pour quelles raisons ?	
	Comment s'opère la collaboration entre les deux professions ? Plus particulièrement entre vous et le responsable technique Monsieur Frédéric Brégier ?	
	Dans un article de 2016 intitulé « Table ronde Vitam au congrès de l'ICA à Séoul » écrit par Madame Mélanie Rebours dans la revue *Archivistes !*, il est indiqué au sujet des deux professions (archivistes et informaticiens) que : *« elles ont des points communs, mais aussi des distinctions qui compliquent parfois les travaux en commun »*. Quelles sont ces distinctions et comment les traitez-vous ?	
	Toujours dans l'article de Mélanie Rebours, il est écrit au sujet de cette collaboration que : *« De façon plus inattendue, elle permet aux archivistes de réseaux différents de discuter , de confronter leurs pratiques, parfois divergentes »*. Quels bénéfices en sont ressortis ? Un exemple ?	
	La collaboration interne à chaque partenaire est censée accompagner la transition des services d'archives vers le numérique et assurer la conduite du changement.	

	Méthodes employées ? Sensibilisation ? Applicables/Transposables pour RiC ?	
	<u>Méthode</u> : Qu'est-ce que la méthode Agile ? Qu'est-ce que Scrum ?	
	Comment l'appliquez-vous ?	
	Pour quelles raisons l'appliquez-vous ?	
Programme VITAM – la solution logicielle	Comment avez-vous étudié la faisabilité du projet ?	
	<u>Norme OAIS (Système ouvert d'archivage d'information)</u> : comment avez-vous utilisé cette norme ? Comment vous l'êtes-vous approprié ?	
	Pourquoi avoir choisi d'utiliser ce modèle conceptuel ?	
	Comment utilisez-vous SEDA ?	
Records-in-Contexts	Quel est votre avis sur RiC ? Quel est l'avis du responsable fonctionnel sur RiC ?	
	De la constitution de l'EGAD ?	
	Du modèle conceptuel ?	
	De l'ontologie ?	
	Votre avis sur le prototype PIAAF ?	
	Monsieur Jean-Séverin Lair suggère qu'il faudrait *« savoir mixer les modèles et sans doute avoir des référentiels sémantisés, à côté de grands gisements de données plus traditionnels »*, qu'en pensez-vous ? (article *Archivistes !* mars 2021)	
Les standards de communication	<u>Sur XML dans la profession archivistique</u> : Son adoption *« a gêné les dernières évolutions dans le domaine de l'archivage numérique »,* que le fait que les systèmes d'information récents ne font plus de XML mais du JSON *« a compliqué les développements de Vitam »*. Pouvez-vous m'en dire plus ?	

	L'intérêt de RDF pour la recherche d'information (information retrieval).	
	L'intérêt de RDF pour la recherche (research).	
	Pensez-vous que RDF pourrait favoriser la sérendipité ?	
Référentiels	Du point de vue de l'utilisateur, en quelle mesure la consultation directe des référentiels améliorerait-il la recherche (information retrieval + research) ?	
	Avantages et inconvénients d'une sémantisation très précise ?	
Représentation sous forme de graphe	Intérêt en termes de compréhension des ressources archivistiques ?	
	Intérêt en termes d'exploitation de ces dites ressources ?	
	Du point de vue de l'utilisateur, comment traiter de grands ensembles d'information ? (outils/méthodes de partitionnement de graphes : bipartition de graphes, méthode hiérarchique, algorithmes de *clustering*, etc.)	

1.5.2. Transcription chrono-thématique

00.00.02s Début de l'entretien.

Parcours

00.00.31s « Je suis ancien élève de l'Ecole des Chartes, donc archiviste-paléographe, conservateur du patrimoine spécialité archives. J'ai commencé ma carrière aux Archives nationales pendant 4 ans au Centre des Archives Contemporaines de Fontainebleau comme responsable de la gestion des fonds où je me suis occupé notamment de l'administration, du référentiel producteurs des Archives nationales qui s'appelait le N1 et donc de la base de description des versements qui s'est créée : Priam 3 ; donc j'ai commencé comme administrateur. Ensuite j'ai été pendant 6 ans chef de la mission des archives du ministère de la Culture et de la Communication, à la fois responsable de la mission archives et assurant le contrôle scientifique et technique sur les opérateurs – avec des grands guillemets autour – puis j'ai été recruté par le ministère des Armées en 2012 comme adjoint à une cheffe de projet *modernisation et gestion du cycle de vie de l'administration*. Je m'occupais de la politique des archives électroniques, et j'assurais un suivi du projet Archipel. Et donc c'est dans ce cadre-là qu'a été monté le programme VITAM et je suis devenu responsable fonctionnel à son lancement en mars 2015, fonction que j'ai assurée pendant 4 ans et demi, avant de prendre mes nouvelles fonctions au 1er octobre 2019 comme titulaire de la chaire des centres des institutions d'archivistique et de diplomatique contemporaine à l'Ecole nationale des Chartes, fonction que j'exerce aujourd'hui. »

Vos missions en tant que responsable fonctionnel du programme VITAM ?

00.02.27s « En tant que responsable fonctionnel du programme VITAM, j'étais chargé de l'expression besoin métier ; donc à la fois du recueil auprès de nos clients des exigences fonctionnelles, de leur mise en forme au sein de l'équipe en dialogue avec mes collègues architectes sous l'autorité de Jean-Séverin Lair directeur du programme, ensuite de la déclinaison en spécifications à destination des équipes de développement donc avec l'aide de mes collègues architectes, et ensuite de la recette du produit, de sa documentation, de sa présentation au client, du recueil des retours des clients, des recettes tests-bugs une fois ça a été ça. Plus l'animation de la communauté métier, donc on organisait des multilatérales métier qui traitaient à la fois des problématiques justement de

recueil de besoins plus partages d'expériences, réponses aux questions liées à l'implémentation, etc. Et on a mené ensuite des chantiers transverses ; moi plus particulièrement je gérais le chantier préservation. »

00.04.08s « Dans ces cas-là c'était pilotage d'un chantier, donc : organisation de réunions régulières, alimentation de la communauté de pratiques avec des sujets, des définitions de protocoles de tests, réalisation d'études. »

Pourquoi avoir choisi d'allier les deux professions : archivistes et informaticiens ?

00.04.44s « C'est indispensable. C'est indispensable pour faire un projet informatique d'archivage d'avoir à la fois des archivistes et des informaticiens. Et ça a été extrêmement riche en la matière justement donc ça a été très bien parce que ça a été souvent des compromis. Déjà c'était une expérience très enrichissante parce que nos collègues informaticiens nous ont vraiment obligé à réfléchir à quels été vraiment nos besoins, à parler du *quoi* et pas du *comment*. Nous on était chargé du *quoi*, du *pourquoi* on faisait ça. Et ce n'était pas dire : moi je veux *ça* et je le veux *comme ça*, non. C'est : qu'est-ce que tu veux faire d'un point de vue métier ? Et après c'était à eux de nous proposer des solutions techniques par rapport à leur connaissance de l'état de l'art, par rapport aux possibilités techniques qui étaient offertes par la technologie d'aujourd'hui, etc. aux choix d'architecture qui avaient été faits, connaissance de la capacité technique à installer, exploiter les choix techniques. Donc parfois on a eu de lourdes négociations par rapport à certaines fonctionnalités mais j'ai toujours trouvé ça très intéressant dans le sens où ça nous a obligé vraiment à réfléchir au *pourquoi* on le fait. »

00.06.32s « L'intérêt aussi c'était à la fois d'envisager les besoins du métier archives, mais d'envisager les besoins du métier informatique en termes de prise en charge, installation, exploitation, stockage, sécurité, reconstruction. Donc vraiment prendre en compte aussi les besoins des équipes d'exploitation. »

00.07.00s « Donc c'est pour ça que ça a été construit à deux, parce que de toute façon, un projet informatique ne peut se construire qu'avec le métier et les informaticiens. »

00.09.30s « On était quatre archivistes, j'avais trois personnes avec moi pour travailler sur ces travaux de spécification. »

00.10.16s « C'était une très belle expérience de se retrouver dans un projet informatique parce qu'on apprend ce que c'est qu'un projet informatique. »

00.10.29s « C'est le retour d'expérience que je fais en deuxième année aux élèves, que je fais en M2, c'est que oui, on apprend parce qu'on a l'image de l'informaticien,

comme de l'archiviste, mais en fait il y a cinquante métiers. Et puis on n'imagine pas les problématiques de logistique, les problématiques techniques d'exploitation, etc. »

00.11.55s « [Au sujet des métiers d'informaticien et d'archiviste] C'est la complémentarité qui est intéressante. On est deux professions totalement complémentaires parce qu'on travaille sur la donnée. Simplement, on l'attaque par un angle différent. Pour les informaticiens, il faut que ce soit disponible à l'instant T, intègre à l'instant T et nous on va penser aux usages futurs en disant : non, ça il va falloir faire autre chose, il va falloir l'envisager de telle manière. On apporte un autre regard sur la donnée. »

00.13.07s « Déjà ce qui a compliqué [la collaboration entre archivistes et informaticiens] c'est la question du vocabulaire. Il faut arriver à se comprendre, ce n'est pas simple. On n'a pas nécessairement le même mode de fonctionnement. [Ce qui a compliqué aussi c'est] Peut-être l'approche littéraire versus mathématiques. »

Confrontation de pratiques divergentes entre archivistes

00.14.23s « On venait de milieux différents. Moi je venais du monde des Archives nationales mais Archives nationales côté missions au ministère, ce qui est un cas particulier. Nathalie venait plutôt du monde *Affaires étrangères* et *Armées*. Medhi était passé plutôt par les opérateurs et puis il avait fait un passage par le privé puisqu'il était passé dans un cabinet d'avocats. Mélanie venait au départ des archives départementales puis de l'administration centrale et Gwendoline était passée par le Quai d'Orsay. Et Marion venait des AD. Elle n'a pas une formation d'archiviste au départ, elle a plutôt, ce que l'on fait à l'Ecole des Chartes, un master *Humanités numériques*. Et elle n'avait jamais vraiment eu un poste opérationnel d'archiviste, elle avait travaillé dans un service d'archives en tant qu'administrateur de bases, en tant que cheffe de projets. »

Agile & Scrum

00.18.13s « L'Agile c'est une méthode générale de développement qui est centrée sur le produit, sur la valeur ajoutée pour les bénéficiaires, et qui est construit progressivement. Donc l'idée c'est de faire de la construction progressive, ça s'oppose à la méthode cycle en V ou en W où le bénéficiaire définit son besoin de manière globale, livre son besoin de manière globale à un prestataire, qui va lui livrer un produit par rapport à ce besoin global, mais ça peut prendre deux ans pour qu'il livre quelque chose, donc ce qui peut quand même faire des décalages. La méthode Agile c'est une méthode qui livre très régulièrement. On faisait des livraisons toutes les quinze semaines à l'extérieur

et nous en interne on avait des cycles de livraisons de toutes les trois semaines. Donc toutes les trois semaines on définissait un besoin que les équipes de développement disaient : ok ça je peux le faire, ça je ne peux pas le faire, etc. Et au bout de trois semaines ils développaient et ils livraient ou ils ne livraient pas. »

00.19.36s « Scrum c'est une des méthodes de mise en œuvre d'un projet Agile, il y en a d'autres. »

00.20.14s « [Agile] C'est une méthode qui est très bien car ça oblige vraiment à se concentrer sur les besoins »

00.20.48s « [Avec Agile] On est très rapidement devant les utilisateurs qui vont critiquer. »

Des preuves de concepts ?

00.23.31s « Il y a eu deux preuves de concepts qui ont été faites. Enfin surtout une sur la faisabilité technique des choix d'architecture. »

00.23.53s « Nous on a fait un travail sur les messageries à ce moment-là, c'était le poc [proof of concept] métier en fait, et c'était plus pour tester notre capacité à travailler ensemble. »

00.24.14s « Dans le cadre de la méthode Agile, on a pu faire en effet des poc sur des points particuliers pour tester et pour voir si ça marchait et si ça marchait on l'intégrait dans le code. »

Norme OAIS

00.25.04s « L'OAIS c'est dire avec des mots compliqués ce qu'est un service d'archives ; un service d'archives au sens anglo-saxon du terme. C'est la grosse différence aussi car au sens *archives*, ce sont des choses à conserver définitivement. C'est pour ça qu'on a dû la transposer, parce que nous on travaillait sur tous les types d'archivages. »

SEDA

00.26.51s « Quand SEDA 2.0 est arrivée, on s'est beaucoup investis dedans notamment Frédéric, pour faire en sorte que ce soit une norme exploitable pour les informaticiens. »

00.28.10s « Une fois que le SEDA a été stabilisée et qu'on avait fait passer le maximum de choses que l'on voulait dans le SEDA, ça nous a servi de colonne vertébrale pour le développement parce que le modèle de données VITAM s'appuie sur le SEDA, même si ce n'est pas une transposition stricte du SEDA parce qu'on n'est pas sous la même technologie : à l'entrée c'est du XML, à l'intérieur c'est du JSON. »

00.28.42s « [Au sujet du passage du XML à JSON] Ca a quelques contraintes. Ca a été pour le coup très lourd et ça a coûté très cher dans VITAM, cette transposition de XML à JSON et de JSON à XML. »

00.29.03s « [SEDA] Ca a structuré notre réflexion. Notamment tous les contrôles qu'on a fait en entrée se sont basés sur le SEDA : Qu'est-ce qu'il faut contrôler ? Qu'est-ce que je peux contrôler ? »

00.31.15s « Les grosses contraintes ça a été : à l'arrivée on reçoit dans VITAM un bordereau en SEDA, un fichier XML, et il a fallu faire le mapping entre la grammaire XML et la grammaire JSON qui ne sont pas du tout identique. Et il a fallu le refaire en sortie, parce qu'en sortie il fallait du SEDA. »

00.31.47s « On a traîné ça, cette difficulté, comme un boulet pendant tout le projet, parce que la norme métier imposait un choix technique. »

Normes de description antérieures & *Records in Contexts*

00.33.12s « J'ai vu l'apparition des normes du Conseil international des Archives, progressivement, que ce soit ISAD (G), ISAAR, ISDF, etc. donc les difficultés d'articulation de ces normes, en effet, c'est un problème. L'idée de faire un modèle de données et qui soit un modèle de données qui articule mieux ces différents éléments, ça c'est très positif. »

00.33.51s « C'est ce que je dis à mes étudiants : c'est que ISAD (G) répond à l'informatique telle qu'elle existait dans les années 1980, qui était plutôt sous forme de fichiers séparés les uns des autres. Les normes après, les couches arrivées dans les années 1990/2000 correspondent à l'informatique relationnelle. Il fallait faire évoluer les choses pour arriver à unifier ça. »

00.34.25s « Après sur la transposition technique, je dirais : quel est le besoin métier ? »

00.34.50s « Le problème du discours sur RiC, c'est qu'on en vient à dire que l'objectif c'est de faire du RiC. Je ne dis pas que c'est un objectif qu'il ne faut pas atteindre, mais : pour faire quoi ? Quel est l'objectif ? Quelle est la finalité ? »

00.35.12s « Je ne fais pas de l'EAD pour faire de l'EAD. Quel est le besoin métier derrière ? Moi mon besoin métier c'est de décrire des archives, faire en sorte qu'elles soient recherchables. Si RiC peut m'y aider, pourquoi pas. »

00.35.33s « Quand j'entends : web sémantique, etc., il ne faut pas oublier *qui* produit la description archivistique aujourd'hui. Dans mon expérience, elle a été produite par mes collaborateurs en mission pour les archives papiers ; dans le numérique, c'est nous en tant qu'utilisateur qui produisons cette description métier, parce que ce sont des métadonnées que l'on va saisir dans les applications métier, qu'on va récupérer en partie pour faire la description archivistique plus tard. Il faut penser à ça. Donc sur RiC, intellectuellement sur ce qu'est la description archivistique, les différents concepts qui sont associés,

oui. Sur l'implémentation technique, faisons attention, voyons quelle utilité ça a, pour quoi faire, et ne nous lançons pas tête baissée là-dedans parce qu'il faut déjà arriver à décrire les archives. Est-ce que c'est un changement possible pour tout le monde ? Parce que c'est quand même très philosophique : c'est lié à des concepts qui séparent des entités qui sont reliées entre elles par des liens, etc. Intellectuellement, ça nécessite d'y réfléchir et d'y penser. Et ensuite, il faut voir où on positionne l'impact technique. »

00.37.31s « [RiC-O] Ca répond à un besoin, qui est un besoin de visibilité sur le web, là, peut-être. Mais toutes les archives ne sont pas accessibles par le web, parce qu'on gère de la donnée personnelle, on gère aussi dans les ministères de la donnée qui est éliminable, qui n'a pas vocation nécessairement à être accessible *via* le web. Donc il faut sans doute réfléchir à l'endroit où positionner RiC, la mise en œuvre de RiC. Donc, RiC pour *FranceArchives* pourquoi pas, ça me semble intéressant. Mais dire que tout le monde va faire du RiC, non. Il faut que ça réponde aux besoins de portails. Est-ce qu'il faut faire porter l'effort partout ? Ça veut dire changer tous les systèmes métier, etc. »

00.38.53s « Je préfère que les services d'archives se concentrent sur faire en sorte que les archives soient classées et décrites, parce qu'il y a encore des archives de l'Ancien régime et des archives du 19ème qui ne sont pas décrites, ou des bordereaux de versement qui ne sont pas mis parce qu'ils ne sont pas jolis, etc. C'est aussi l'arbre qui cache la forêt : qu'est-ce qui est mis à disposition des chercheurs que l'on ne challenge pas ? Parce que quand vous êtes un chercheur – enfin moi je le vois dans mes nouvelles fonctions – essayez de trouver la description des archives postérieures à 1945, eh bien bon courage. Mettons l'accent là-dessus, faisons en sorte que ce soit décrit, exportable vers un endroit qui peut-être en effet, permettra d'avoir une meilleure accessibilité utilisant des technologies comme celles de RiC, si tant est que ça tienne la charge parce qu'il faut aussi ça. Il faut que ça puisse tenir la volumétrie derrière. »

Les référentiels

00.40.56s « Un des grands non-dits dans le domaine des archives est celui des référentiels, des données de référence, parce qu'on a besoin en effet de ces données de référence et c'est un sujet dont on ne parle pas assez , dont on ne traite pas assez, que ce soit des règles de gestion, que ce soit des référentiels géolocalisés pour arriver à faire en sorte qu'on arrive à positionner quand on décrit quelque chose, une description qui soit géolocalisée et qu'un utilisateur qui arrive par une carte puisse y accéder en naviguant à partir de sa carte. Le référentiel, une fois qu'il existe on peut le réutiliser partout. Il y a des points

de concentration de ces référentiels à faire. Le tout c'est d'arriver à simplifier le travail de tout le monde. »

Intérêt(s) de la représentation graphique pour l'utilisateur final ?

00.43.43s « Le fait de pouvoir rebondir d'une ressource à une autre – archivistique, bibliothéconomique, etc. – ça pourrait en effet être intéressant. Sur la représentation graphe, le fait de tester de nouveaux modes de visualisation, oui. Moi qui suis un visuel, j'aime bien les choses visuelles, donc oui ça me semble intéressant. Après, est-ce que c'est exploitable dans le temps ? Est-ce que c'est la forme à privilégier pour tout ? Pas nécessairement parce que tout le monde n'a pas la même approche cognitive je pense, mais là il faudrait aussi voir avec des spécialistes du cognitif, mais ça mériterait d'être vraiment testé, et d'être testé avec des utilisateurs. »

00.45.43s « Quand on voit déjà les difficultés avec la salle des inventaires virtuelles des Archives nationales pour arriver à trouver quelque chose, et même sur les sites d'archives départementales pour déjà arriver à trouver *où* est-ce que je vais chercher… donc il faut aussi faire simple pour les utilisateurs je crois. [Au sujet de la représentation sous forme de graphe] Ça demande une expérimentation avec un panel représentatif d'utilisateurs. Mais d'être fait, testé, dans le cadre d'un projet avec des archivistes, des informaticiens, des spécialistes du cognitif voire des sociologues pour étudier les réactions. Mais avec des informaticiens à l'intérieur, qui sont là pour aider à gérer la masse, l'exploitation de tout ça, la faisabilité technique. »

Les référentiels comme outils d'aide à la recherche

00.49.41s « Je vais vous donner un retour d'expérience concret. J'ai eu le malheur dans mon nouveau poste de faire des recherches sur l'histoire des archives ; tapez *archives* dans la salle des inventaires virtuels, vous allez comprendre votre douleur. Globalement comme la recherche ne se fait pas uniquement sur la description des archives, en termes de XML-EAD sur les <c> et sur les <unittitle> des <c>, et que vous faites aussi une recherche sur le <archdesc> et sur donc l'introduction, toutes les introductions décrivent des archives, donc si vous faites une recherche sur *archives*, vous avez tout qui sort. Donc pour moi le service producteur est un moyen d'entrer. Parfois il est plus simple pour moi – peut-être par ma pratique d'archiviste – d'aller chercher le producteur, pour après avoir ses archives. Pour l'utilisateur lambda, ce n'est pas nécessairement son mode de fonctionnement, il va chercher un intitulé de dossier. Il ne va pas chercher nécessairement par le producteur parce qu'il ne

00.52.49s sait pas nécessairement qui a produit l'archives. Le référentiel agents me sert beaucoup mais c'est une approche, pour un type de profil. »

« L'approche référentielle c'est ça après, c'est aussi permettre une approche cartographique, parce qu'on passe sur un référentiel géographique, mais quand vous avez de la description qui utilise des informations d'époque, il faut que votre référentiel soit un peu plus intelligent : quand vous avez quelqu'un qui est né à Presbourg en 1850, tout de suite ça le positionne à Bratislava en Slovaquie et pas à Presbourg dans l'empire d'Autriche. »

00.54.52s Fin de l'entretien.

1.6. Entretien avec Monsieur David Chavalarias

1.6.1. Guide d'entretien

Thème	**Sous-thème**	**Temps**
Parcours	Études (diplômes, lieux, dates, stages)	
	Vous avez une formation initiale en mathématiques et informatique, vous vous êtes ensuite orienté vers les sciences cognitives, quelles en ont été les motivations, quel est votre intérêt personnel pour cette discipline ?	
	Vous êtes actuellement directeur de recherche au Centre d'Analyses de Mathématiques Sociales (CAMS) mais également directeur de l'Institut des Systèmes Complexes de Paris. Emplois précédents ?	
Représentation sous forme de graphe	Les difficultés auxquelles on peut faire face dans l'élaboration d'un modèle conceptuel ?	
	Dans votre interview Mooc Digital Paris, vous parlez de systèmes discrets pour faire référence aux systèmes de graphes, pourriez-vous expliciter svp ?	
	Vous dites également que la théorie des graphes s'est beaucoup développée ces dernières années. Quels en sont les facteurs ? Qu'est-ce qui a favorisé ce développement/cet engouement pour les graphes ?	
	Ontologie : Quand on élabore une ontologie, il y a nécessairement un modèle conceptuel de réalisé en amont sur lequel on s'appuie ?	
	Pouvez-vous expliciter les critères propres aux ontologies qui sont : la clarté, la cohérence, l'extensibilité, un engagement ontologie minimal et l'évitement de toute déformation d'encodage minimale ?	
	Comment faire émerger des entités conceptuelles qui sont implicites dans les fichiers ? L'approche ontologique peut-elle permettre la conversion d'informations latentes en informations manifestes ?	
	Selon vous, quels sont les avantages et les inconvénients de la représentation sous forme de graphe en termes de compréhension (des ressources archivistiques) ?	

	Intérêt en termes d'exploitation (de ces dites ressources) ?	
	Dans votre interview Mooc, vous dites au sujet de d'Alembert qu'il pensait déjà au caractère multi échelles, c'est-à-dire au niveau de zoom, comme quand on regarde une carte […] et qu'il faisait l'analogie entre la déformation des mappemondes une fois projetées sur un plan et la représentation des connaissances qui serait déformée si un jour on devait la représenter de cette manière-là. Il dit qu'il y aurait autant de représentations de la connaissance que de mappemondes et que chacune de ces représentations apporterait quelque chose de nouveau. Pourriez-vous expliciter svp ?	
	Du point de vue de l'utilisateur, comment traiter de grands ensembles d'information ?	
	Quels sont les outils de partitionnement de graphes les plus pertinents selon vous ? (algorithmes de *clustering*)	
	L'intérêt du langage RDF pour le *text mining* (*information retrieval* : recherche dans le sens de manière de retrouver des informations provenant d'une ou plusieurs base(s) de données).	
	L'intérêt des graphes pour la recherche (*research*). (recherche dans le sens d'ensemble d'études et de travaux ayant pour objet de faire progresser la connaissance)	
	Comment le langage RDF pourrait favoriser la sérendipité ?	
	Selon vous, quels sont les avantages et inconvénients d'une sémantisation très précise ?	
Fonctions cognitives & traitement de l'information	On entend souvent qu'« une image vaut mille mots ». Est-ce vrai d'un point de vue cognitif ? (compréhension plus rapide, d'un coup d'oeil)	
	La perception joue un rôle majeur dans le traitement de l'information. Quels sont les mécanismes impliqués ? Que doit faire le système perceptif pour qu'une scène visuelle soit compréhensible ?	
	Expliciter les niveaux de traitement de l'information (niveau sensoriel, niveau perceptif, niveau cognitif).	
	Expliciter les types de traitement de l'information ? (traitement ascendants/traitement descendants)	

	Quels sont les mécanismes mentaux de la compréhension ?	
	Quels sont les mécanismes mentaux du raisonnement ?	
	Comment fonctionne le cerveau du chercheur ? Face à ses sources, comment le chercheur parvient à produire de la connaissance scientifique (pdv cognitif, processus cognitifs élémentaires mis en œuvre) ?	
Représentation sous forme de graphe et fonctions cognitives	Dans l'interview Mooc, vous dites que *« Rien que la visualisation des graphes est instructive »*, pourriez-vous expliciter svp ?	
	Quelles sont les processus cognitifs élémentaires sollicités face à une représentation des connaissances sous forme de graphe ?	
	Quels sont les effets de la représentation sous forme de graphe sur nos mécanismes de traitement de l'information ?	
	Mémoire de travail et représentation sous forme de graphe, que pouvez-vous m'en dire ? (fonctionnement, effets)	
	Quels mécanismes mentaux sont impliqués dans l'analyse des représentations sous forme de graphe ?	
	Comment la représentation sous forme de graphe peut-elle favoriser la récupération et l'utilisation des connaissances par l'Homme ?	
	L'attention & la représentation sous forme de graphe, que pouvez-vous m'en dire en termes de fonctionnement, mécanismes mentaux impliqués, intérêts, etc. (attention sélective, attention partagée, attention soutenue)	
	Raisonnement & représentation sous forme de graphe : que pouvez-vous m'en dire ? Est-ce que ce type de représentation pourrait favoriser le raisonnement (parvenir à faire de nouveaux parallèles, faire naître de nouvelles idées, mieux se représenter un environnement/un contexte) ?	
	En tant que chercheur, est-ce que la représentation sous forme de graphe vous apparaît comme un outil d'aide à la compréhension d'une situation/d'un problème/d'un contexte ?	

1.6.2. Transcription chrono-thématique

00.00.05 Début de l'entretien

Parcours

00.00.30s « Ma formation initiale c'est mathématiques/informatique. J'ai fait une scolarité à l'ENS et je me suis assez vite intéressé aux sciences cognitives ; j'avais un goût pour l'interdisciplinarité. J'ai donc fait un DEA en sciences cognitives et suite à ça j'ai fait une thèse à l'École polytechnique sur la question des modélisations des dynamiques sociales. »

00.02.28s « En parallèle de recherches un peu théoriques et formelles sur les dynamiques sociales, je me suis intéressé à la manière de les explorer et de les reconstruire à partir des graphes numériques. »

00.06.05s « Je suis au CAMS en tant que directeur de recherche CNRS et par ailleurs je dirige l'institut des systèmes complexes. »

Modèles conceptuels

00.08.23s « Les modèles conceptuels nous permettent de réduire un petit peu le caractère foisonnant des objets naturels pour essayer d'en extraire, on peut dire, les principes actifs, les mécanismes ou les propriétés qui vont donner lieu à des observables. Et donc, le pari avec des modèles conceptuels, c'est de réussir à identifier ce qui produit dans l'observation certains traits qui sont saillants pour notre perception. »

Ontologies

00.15.25s « Quand on travaille sur les ontologies, c'est qu'on essaie de faire des catégorisations, des manières de désigner les objets naturels. C'est une catégorisation hiérarchique, ordonnée, avec des types de relations. Il ne faut pas qu'il y ait de contradiction. »

D'Alembert & la représentation de la connaissance

00.19.07s « D'Alembert parlait de la représentation de la connaissance. Donc si on prend la connaissance, qu'on l'attrape par le biais des traces qui sont laissées par les chercheurs et les savants dans leurs écrits et que l'on se dit que là, il y a de la connaissance et qu'on veut représenter cette connaissance, il faut bien comprendre que cet objet-là vit dans un monde aux dimensions *a priori* très grandes. Il n'y a aucune raison pour que ce soit un objet en une dimension, deux dimension, trois dimension. Et notamment, il peut être dans de grandes

dimensions et avec des topologies spécifiques. Typiquement, la terre est en dimension deux mais elle a une typologie de sphère, qui n'est pas pareil que la topologie de plan. Donc quand vous avez un objet qui est dans un espace de très grandes dimensions et que vous voulez le représenter, de manière à ce que vous avec vos yeux, vous puissiez le voir, comme vos yeux ne voient que des dimensions un, deux et trois — et encore trois c'est compliqué —, ça veut dire que l'on va être obligé d'écraser, de compresser l'information, et de ne retenir de cet objet, et des informations qu'on a dans cette grande dimension que certains aspects. Et donc, quand on retient que certains aspects, on va forcément en déformer certaines relations. »

00.21.13s « Pour représenter ces propriétés intrinsèques, on est obligé de faire un choix, et ce choix lui est objectif ; c'est un choix. Donc quand vous voulez représenter la terre sur un espace plan, vous êtes obligé de faire un choix de quels sont les différents modes de projection de la sphère sur le plan. Il y a des projections qui déforment les angles, d'autres vont déformer les surfaces ; ça ne va pas être les mêmes projections et en fonction de ce que vous voulez regarder, vous allez faire différentes projections. C'est à ça que se référait d'Alembert, c'est-à-dire que la connaissance vit *a priori* dans un espace de grandes dimensions, et que si on veut la représenter dans une encyclopédie par exemple, il va forcément falloir faire des choix de représentation. »

00.22.09s « Il peut y avoir aussi des choix personnels. Vous avez un objet naturel, souvent ce sont les objets des systèmes complexes donc ils ont des étendues assez importantes, ils sont composés de pleins d'éléments, il y a des interractions, etc. La première chose quand vous voulez essayer de le représenter, c'est de savoir ce que vous allez mesurer sur cet objet ; si vous prenez la science vous pouvez mesurer les auteurs, le sens des mots, l'interraction entre les mots, les dates de publication, les supports de publication… il y a plein de choses que l'on peut mesurer. Savoir quoi mesurer, c'est déjà engager une certaine subjectivité. »

00.23.28s « Une fois que vous avez fait ce choix-là qui correspond à votre question de recherche, vous faites les mesures, aussi exactes que possibles, et même après ça, vous êtes face à un objet qui est très complexe. Ces objets très complexes vivent dans un espace de grandes dimensions, et après, si vous voulez les représenter, encore une fois il va falloir perdre de l'information. »

Comment lire et comprendre un graphe ?

00.25.27s « Quand on crée une visualisation, on sait bien que le choix de certaines couleurs va attirer l'œil plus que d'autres, on peut avoir une forme qui est plus

ciselée ou au contraire un peu plus vague dans ses contours, ça va avoir des effets différents. »

00.25.52s « Les questions de densité d'information : des endroits où il y a beaucoup d'informations, des endroits où il y en a peu. Tout cela va orienter l'interprétation et le regard. Ce qui est important de mon point de vue, dans une visualisation, c'est le côté interaction et ça c'est quelque chose qui est assez récent, qui a dix/quinze années. Pendant longtemps, mais aussi parce que les ordinateurs ne le permettaient pas, on produisait des visualisations à plat, des cartes, des graphiques, et là maintenant depuis quelques années il est très courant de pouvoir interagir avec ces représentations : cliquer dessus, tourner le graphe, filtrer des choses, filtrer des choses… tout cela est très important car c'est ce qui permet justement de tourner autour de l'objet dans l'espace de grande dimension dans lequel il vit. C'est comme lorsque vous tournez autour d'un bâtiment, vous voyez Notre-Dame, vous voyez le fronton, après vous êtes derrière, vous voyez d'autres parties, et à chaque fois ça vous donne une vision différente du bâtiment et c'est parce que vous êtes capable de tourner autour que vous vous faites une représentation plus complète de la chose. Et donc ça c'est très lié à la cognition située, c'est-à-dire que la perception des choses que l'on a est très liée à notre capacité à bouger, à se mouvoir dans le monde et donc à changer les points de vue. Et donc dans les représentations ce qui est important c'est de reconstituer ce rapport aux objets, c'est-à-dire de pouvoir bouger et tourner autour. Et donc à ce moment-là on a une représentation située qui est la manière naturelle de percevoir. »

Quels outils pour interagir avec les graphes ?

00.28.39s « Alors, il y en a plein. Dans un graphe disons qu'il faut toujours, enfin surtout quand il est complexe, avoir une option de zoom, c'est-à-dire de pouvoir le voir de loin, de près, et d'afficher des informations différentes quand c'est de loin ou de près, de pouvoir filtrer les liens qui lient les éléments du graphe entre eux. »

00.29.11s « Les algorithmes qui permettent justement d'étaler un graphe sur un plan dépendent évidemment du nombre de connexions qu'il y a entre les éléments du graphe. Et donc quand vous faites varier ce nombre de connexions, vous faites apparaître des sous-structures qui sont très intéressantes et donc en respatialisant on peut les identifier de visu ou alors en relançant l'algorithme de clustering. »

00.29.47s « En fait, il faut bien voir qu'un graphe c'est un objet qui a un caractère multi-échelle c'est-à-dire que les liens dans un graphe sont souvent pondérés

c'est-à-dire qu'il y a plusieurs valeurs des liens — des liens forts, des liens faibles — et en fait la structure de ces liens, à différents niveaux — si on prend que les liens forts, que les liens faibles, un mélange de tout —, cette structure-là est très importante pour comprendre les propriétés du graphe. Et donc pour les comprendre, le mieux c'est de pouvoir filtrer et de pouvoir réarranger le graphe en fonction de ce qui reste. »

Text mining

00.30.42s « Text mining en français c'est *fouille de données textuelles.* C'est l'analyse informatique et mathématique des données textuelles par des aspects quantitatifs, c'est-à-dire qu'on va analyser du texte et essayer d'extraire des relations grammaticales, des relations de sens, on va essayer d'extraire des réseaux de graphes, comprendre quels sont les sujets automatiquement dans un article, au cœur du résumé d'article ; tout cela c'est ce qu'on appelle le text mining. »

00.31.22s « Vous avez un matériau qui est le texte, qui est numérisé — donc l'ordinateur peut le lire — qu'est-ce que vous pouvez faire avec ? C'est ça le text mining. »

00.31.35s « Pas uniquement [le text mining ne repose pas uniquement sur le langage RDF], le langage RDF c'est un des résultats d'un text mining. C'est-à-dire que le langage RDF va vous dire : *Jean*, *pomme*, la relation c'est *a mangé* par exemple. Et ça c'est déjà le résultat de l'analyse d'un texte où il y avait d'écrit : Jean a mangé la pomme et est parti se balader. Mais on aurait pu analyser ce texte de plein d'autres manières différentes. »

Est-ce qu'une image vaut mille mots d'un point de vue cognitif ?

00.33.34s « D'un point de vue cognitif, la question c'est celle de l'attribution du sens. Donc comment est-ce qu'on attribue du sens aux choses que l'on voit, que l'on perçoit ? Et en fait, la vision est associée à une interprétation presque proprioceptive des choses et donc il y a une compréhension qui passe par la visualisation qui parfois ne peut pas du tout être exprimée en mots ou assez difficilement donc dans ces cas-là effectivement une image vaut mille mots, ça ne sert à rien d'essayer d'écrire en long et en large. »

00.34.34s « Toutes les capacités cognitives de l'œil, du cerveau à détecter des nuances de couleurs, des nuances d'orientation sont automatiques. Par exemple la détection des contours c'est quelque chose qui est automatique, qu'on arrive très bien à faire. »

00.35.08s « Donc à chaque fois que l'on va faire appel à des fonctions cognitives assez bas niveau qui sont de l'ordre de la vision, de la détection de formes, de contours, ou même après de proprioception c'est-à-dire où avec la

compréhension d'une forme visuelle, on va donner une image. Et c'est vrai qu'il y a beaucoup de situations qui sont dans ce cas de figure-là. »

Plusieurs niveaux de traitement de l'information ?

00.36.37s « Le niveau sensoriel, vous avez déjà tout un ensemble de fonctions cognitives de très bas niveaux : détection des contours, des couleurs, etc., ça vous ne pouvez presque pas vous empêcher de faire ces traitements. »

00.37.18s « Ensuite, il y a l'interprétation que l'on fait de tout cela avec notamment tout ce que l'on connaît déjà et que l'on a rencontré dans notre vie, et donc ça c'est le fait que notre cerveau apprenne en permanence. Il faut savoir que la rétine fait un traitement local, puis projette tout cet influx nerveux sur les aires cérébrales et ces premières aires cérébrales, traitent la manière dont nous allons détecter ces premiers signaux, détecter les contours, etc. Quand on regarde la manière dont elles traitent les signaux, 90% de l'information vient du cerveau, 10% qui vient de l'extérieur. C'est-à-dire que notre cerveau réutilise énormément de choses, tout ce qu'il a déjà vu ou qu'il connaît pour interpréter les scènes visuelles. »

00.39.59s « Et puis ensuite, il y a le raisonnement qui va utiliser tout ça pour construire du sens encore plus élaboré. »

Visualisation d'un graphe & compréhension

00.42.22s « Quand vous regardez un graphe, votre cerveau catégorise directement les formes qu'il voit dans ce graphe, et ça, ça vous permet directement de créer du sens dans le graphe. »

00.43.18s « [Exemple avec les graphes politiques] Si je vous décrivais les relations de ces deux cent mille nœuds comme une longue liste vous ne comprendriez rien du tout, par contre si je vous montre le graphe, là tout de suite vous comprenez la structuration de l'espace politique. Donc la lecture d'un graphe, en général, directement, vous donne de l'information. ».

00.44.30s « Nos sens sont précablés pour extraire des corrélations spatiales dans les choses que l'on voit ; c'est quelque chose qui se fait sans aucun effort, c'est quelque chose qui est fait de manière parallèle au niveau de nos neurones. Et donc quand on regarde un graphe, on bénéficie de tout ce traitement, et donc apparaît au niveau conscient, ces corrélations sous forme de sens un peu structuré ; on voit qu'il y a des ensembles et c'est en cela que voir un graphe nous enseigne quelque chose, pour autant qu'il soit bien spatialisé. Si je vous montre un graphe mis n'importe comment ça ne va rien vous apporter, mais si le graphe a été travaillé et la disposition des nœuds du graphe suit une certaine logique, vous êtes capable d'attraper cette logique même si vous avez

deux-cent milles nœuds. »

RDF & sérendipité

00.47.38s « Le langage RDF est une manière fine de structurer des relations entre objets donc ça veut dire que dans le langage RDF, on encode une certaine connaissance. »

00.48.04s « Avoir la possibilité d'interroger une base avec un langage RDF va forcément offrir plus de possibilités et effectivement permettre de découvrir des choses d'une manière peut-être plus puissante et aussi rattraper des éléments d'information auxquels on n'aurait pas pensé. Donc oui on peut dire que d'une certaine manière ça favorise la sérendipité. Tous les dispositifs d'interraction, qu'ils soient visuels ou textuels, qui se basent sur une préstructuration des données vont vous aider à découvrir plus facilement les choses. »

00.48.55s Fin de l'entretien.

2. Questionnaires

2.1. Réponses de Monsieur Daniel Pitti

About your career

1. **Studies (diploma, places, dates, internships).**

University of California, Berkeley. MLIS. 1987.

University of California, Los Angeles. C.Phil. in the History of Religions. 1984.

University of California, Los Angeles. MA in the History of Religions. 1981.

University of California, Davis. AB in Religious Studies. 1977.

2. **You are currently chair of EGAD. How did you get to this position? (date, motivation of your application, your expectations).**

In August 2012 at the ICA Congress in Brisbane, Australia, I was asked by Henri Zuber, then president of the ICA Progamme Commission; Claire Sibille-de Grimoüard (Service interministériel des Archives de France); and Vitor Manoel Marques da Fonseca (Arquivos Nacional do Brasil) to chair the Experts Group on Archival Description (EGAD), and that the focus of the work would be to develop a standard to replace the existing ICA description standards: ISAD(G), ISAAR-CPF, ISDF and ISDIAH.

Given the amount of work such an endeavor would require, I was reluctant to accept the responsibility. In the end, I was persuaded to do so.

I would characterize my expectation as cautiously optimistic about the success of the activity. In retrospect, I did severely underestimate the time it would take to develop the new standard. This was due to a variety of factors, including the following:

- Most of the members have fulltime professional positions, and given their expertise and competence, they tend to have large demands on their time, challenging them to find time to do the voluntary work of EGAD. This was compounded by the international economic crisis that began in the fall of 2008, as it resulted in many repositories to a decrease in staffing.
- The intellectual and technical challenges of developing the Records in Contexts (RiC) standard were complex and detailed, and the decision to develop the standard as two parallel facets, a high-level Conceptual Model and a formal Ontology based on it increased the challenge.
- The EGAD functions by consensus, which is to say, the EGAD does not vote on decisions, but discusses each issue until there is general agreement and acceptance. While this is less efficient than limited discussion followed by a vote, employing the consensus method ensures over all harmony in the group, and leads to more open and in-depth analysis and discussion. Further, all decisions are provisional rather than final, which is to say, if new insights and ideas emerge, past decisions can and are frequently revisited. It is a slow, deliberative process that combines intellectual, technical and social factors. The objective

is to create something together that none could create alone, with each member contributing based on different perspectives, expertise and professional experience.

3. **What exactly does your job entail? What kind of decision do you take?**

Chairing involves administrative and process management duties. These include: communications with ICA staff; writing reports; applying for funds to support face-to-face meetings; organizing and leading face-to-face meetings; and organizing and hosting frequent teleconference meetings. In addition to the administration and management duties, I have also been involved in drafting substantial portions of the text of the Conceptual Model.

4. **Previous positions?**

My professional activities have been in two related areas: archives and digital humanities. With respect to archives, I have devoted the bulk of my time to developing and promoting archival description standards; and designing and developing systems based on the standards. I conceived of Encoded Archival Description (EAD) in the early 1990s, and then served as its principal technical architect until 2010. In the early 2000s, I began collaborating on the development of an encoding standard for the description of corporate bodies, persons and families, and served as the principal technical architect of Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF) until 2002. Based on EAD, I collaborated in the founding of the Online Archive of California and subsequently advised in establishing other EAD-based archival consortia in the United States, and Archives Portal Europe.

With respect to digital humanities, in collaboration with humanities scholars, graduate students, and programmer/developers, I designed and supervised the implementation of a large number of projects including: A Worldview in Words: Lexical Categories of the Mopan Maya; Digital Yoknapatawpha; Collective Biographies of Woman; Dante's Divine Comedy; The Lives of Saints; The Arapesh Grammar and Digital Language Archive; Leonardo Da Vinci's Treatise on Painting; Walt Whitman Archive; The Chaco Digital Initiative; Exemplary Women Project; Salem Witch Trials; Uncle Tom's Cabin Archive; The Samantabhadra Collection of Nyingma Literature; William Blake Archive; Pompeii Project; Piers Plowman Archive; Rossetti Archive; and Valley of the Shadow.

About the InterPARES comments

5. **InterPARES comment : *"[RiC-CM] started with no analysis of both the actual level of application of the ICA standards in different countries [...]. This is a crucial factor in a decision whether to integrate a particular standard in the new model, and how it should be integrated."* What were the preliminary questions considered by EGAD ?**

The InterPARES comment is correct, in that there was no systematic survey of the use of the ICA description standards county-by-country. Whether this is a crucial factor is arguable.

Based on available, largely anecdotal, evidence, of the four standards, ISAD(G), ISAAR-CPF, ISDF, and ISDIAH, *only* ISAD(G) has been and continues to be broadly influential globally. The second edition was translated into 17 languages, though no Asian languages. In turn, ISAD(G) was the foundation for EAD, and EAD has been translated into a dozen languages, including Chinese.

Of the other three standards, ISAAR-CPF has had some success. It serves as the foundation for EAC-CPF, which is or has been integrated into existing archival management platforms, and also serves as the foundational standard for the Social Networks and Archival Context Cooperative (SNAC).

Standardization of archival description varies considerably internationally, and whether or not there has been substantial focused investment in standardization is influenced by a number of factors including in particular: investment in established local practices that are prohibitively resource-intense to change; and available human resources and expertise. While many archives have embraced standardization, many under resourced archives have not, and it is an ongoing ethical, intellectual and political challenge to devise methods and strategies for assisting archives, large and small, that have insufficient resources to adequately process, preserve and provide access to the records for which they are responsible.

At the beginning of the EGAD deliberations, we published *Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives* (available in French, English and Arabic here: https://www.ica.org/en/egad-resources). It provides a good overview of our preliminary considerations.

6. **I've had a meeting with Claire Sibille de Grimoüard three weeks ago and she mentioned the fact that there are countries currently that still try to completely adopt ISAD/G. Thus, there is an idea of a two-speed standardization. How can these countries adopt RiC?**

As described above, there is a big gap between the relatively well-resourced archives and those under-resourced. This is essentially a microcosm of the gap between the developed versus the developing countries, and the gap within countries between the rich and the poor.

RiC is aspirational, that is, it does not canonize existing practice, but lays the foundation for next generation archival description. If successful in being adopted, such adoption will be gradual. It will likely begin with well-resourced archives, by adoption in readily available tools (such as ArchivesSpace and AtoM) that ameliorate the intellectual and technical challenges, and by integration into existing description content standards. Based on my experience with EAD, it will take many years to become established and significantly dispersed. Assuming success, there will be many countries and archives within countries that will be unable to participate in the benefits without specific methods, strategies and tools to assist them. This, I think, is a long-term ethical and political challenge for ICA and the international archival community. Ultimately, we have a professional duty both to develop next-generation archival standards that position archives to be full, equal partners in providing access to cultural heritage, and at the same time,

to ensure that over time the underrepresented countries and demographic groups are fully represented in the global archival social-document network. Our collective memory needs to be inclusive of all human history. It the social justice dimension of archival professional principles and practice.

7. **RiC wants to gather the archival community by providing best practices but, regarding the current situation, can he widen the gap between the different countries?**

I think it will initially widen the gap. This, though, is inevitable when extending the boundaries of existing practice. I think the long-term strategy, though, will require focused effort at closing the gap.

8. **As chair of EGAD, did you choose the members? What were the criteria to select them?**

The members were chosen in consultation with the president of the ICA Programme Commission. EGAD is a partial successor to the ICA Committee on Best Practices and Standards (CBPS), and at the time of forming EGAD, the existing members of CBPS were given the option of transitioning to EGAD. It was thought important to have continuity between the old and the new. As chair, I recruited additional members based on expertise and experience that were anticipated to be important, and also attempted to be as diverse as possible in national representation. I knew many of the new members, though some were recommended by others whose judgement I trusted. A key factor is that each member had to have a reputation for successful collaboration, of being able to politely work with others. An important social consideration is that you want the members of a group to feel comfortable speaking openly, and to argue, as argument is essential, with respectfully. When the group comes to consensus, it is often consensus on an idea that has been formed and shaped by multiple members.

About EGAD

9. **How many members? Which countries are represented?**

21 members from 14 countries.

Australia, Austria, Brazil, Canada, Finland, France, Germany, Hong Kong, Italy, Romania, Spain, Switzerland, United Kingdom, and the United States. One member from the Ivory Coast resigned last year.

10. **What did EGAD answer to the InterPARES comment : *"RIC-CM has been developed without significant input from Africa and Asia (there is no Asian representative, and a single member from Ivory Coast representing the whole Africa)"?* Would not it be simpler to have one representative per country?**

The EGAD did not respond directly to the InterPARES comment, as the draft RiC-CM 0.2 draft already contained the following observation:

"EGAD has twenty-one members from thirteen countries (please see Appendix II for a list of members). While the members are broadly representative of the global archival community, there are many areas with long and distinguished histories of administration and governance, and concomitant traditions of record creation, use, and management that are not represented, such as Asia, southwest Asia, and eastern Europe. And while Africa, central Europe, and South America are represented, the representation could be broader. Given this, the members of EGAD welcome broader international participation in the ongoing development of an international archival description standard, so as to strengthen the intellectual foundation of the standard, and ensure that the standard represents a broad consensus and addresses the needs of the global archival community."

The InterPARES comment thus, added nothing to our understanding of our own shortcomings.

There are 195 countries in the world. I do not think having one representative from each would be realistic or manageable.

11. What did EGAD answer to their question : *"Do the countries represented by two members have two votes when decisions are made?"*

EGAD, as stated above, operatives by consensus. Thus the answer is no, given there have been no votes.

12. Are there any palpable conflicts of interest within EGAD (maybe some members who would want to impose their archival traditions)?

Each member has a perspective formed by the context within which he or she lives and works. This is, in fact, a good thing. Developing a standard is ultimately political, in the good sense of the word: a community building, consensus building exercise. A standard should be representative and thus involve as wide a range of perspectives, understandings and experiences as possible. The overall attitude cultivated in the group is that none of knows everything, but each of us has knowledge to contribute to the deliberations. In addition to creating something together none of us could create alone, I think if you were to ask each member of EGAD if they have learned a great deal from one another, the answer would be yes. From my perspective, I am deeply indebted to others, as my knowledge and understanding has been expanded and deepened by the dialogues that I have had with others. I could enumerate a list of people that have had a profound influence on me, and many on this list would be past and current members of EGAD.

There are a lot of "know a lot" members of EGAD, but no "know it all." The latter is anathema to the success of a group.

13. **InterPARES comment : *"In RiC-CM it is stated that "once the model is stable, it will be translated into French and other languages". [...] However, RiC-CM aims at becoming a professional standard that should be adopted worldwide, so the focus on French appears neither necessary nor adequate"*. What did EGAD answer to this comment?**

As with all of the other InterPARES comments, we did not respond. The authors of the InterPARES comments know quite well that the two official languages of ICA are English and French, and so the French is mentioned explicitly as coming first after English. And then "other languages" to follow. There will immediately following other Western languages. And then there will be a concerted effort to recruit additional translators. Such will take some time, as one would expect.

14. **InterPARES comment : *"RiC-CM does not consider "archivist" as a core descriptive entity [...] it seems impossible to define a model of descriptive elements if a model of the archivist's role is not defined. Instead, EGAD considers this as an action that can be postponed"*. What did EGAD answer to this comment?**

RiC does not consider "archivist" as a core descriptive entity because archivists and records managers, for that matter, are, at a high level, Agents, and then in most if not all cases, archivists and records managers hold Positions, a kind of Agent in RiC that is within a Corporate Bodies, that perform the activities (Activity) associated with managing, processing and describing record resources, based on rules, both those devised by the controlling Corporate Body but also by professional standards and practices. Archivists and records managers are well covered in RiC, they are just not elevated to having some transcendent role that makes them different from other people performing activities informed by occupational education/training and subject to professional rules. Further, you engage in managing (an Activity) your own records, I presume. Based on the InterPARES criticism would this mean that everyone that engages in activities associated with managing, processing and describing records should be identified as a kind of Agent?

Further, RiC-CM 0.1 had the following text:

"While almost all archivists accept the theoretical validity of Provenance, many have become increasingly self-conscious and self-critical about the role that archivists and the application of archival principles play in what is remembered and how it is remembered. This criticism necessarily encompasses more than description, as it also includes appraisal, the evaluations that lead to what is chosen to be remembered and what is not. Viewed broadly, appraisal determines which records are kept and which are not. Memory is selective: what archivists choose to keep directly determines whether the memory is representative of society. These determinations have profound epistemological and ethical implications. Appraisals by both records managers and archivists, as well as accidental loss or dispersal of records, also directly

determine the relative integrity or lack thereof of the fonds. Both the records kept and those lost or intentionally removed are part of the story of the fonds, and thus an important facet of the context necessary for evaluating and understanding records by users."

We did not respond to the observation, as, again, what was observed had already been explicitly addressed in the draft standards. Further, RiC-CM 0.1 did not deferred until a subsequent draft on describing description only because it did not have time to address this before meeting a deadline for releasing the draft. It is addressed in RiC-CM 0.2. And it was clearly stated in 0.1 that the draft was incomplete.

15. **InterPARES comment : *"[…] description is not always the work of archivists. […] But there appears to be nothing in the standard that would support representation of different and possibly conflicting viewpoints; indeed, EGAD does not reflect the movement toward user participation".* What did EGAD answer to this comment?**

Again, this was addressed RiC-CM 0.1 in two paragraphs following the two above:
"Increasingly, archivists observe that the archival perspective is one among many possible perspectives that may be employed in the understanding of records, that they themselves are performing their jobs in a particular historical (cultural, social, material) context, that their judgements and acts are shaped and informed by the contexts within which they live and work. At the same time, archivists increasingly are observing that the contexts in which records emerge and in which they exist over time are irreducibly dynamic and complex.
When taken together, the two observations lead to the conclusion that archival description is not and never will be perfect, but they also foreground the challenge of making it better, of improving it, of making it possible to more fully express the complexity of context and contexts, of making explicit the different roles of the records managers and archivists in selecting and forming memory, and of being accommodating to the perspectives and values of others."
RiC-CM 0.1 did reflect the movement towards user participation.

16. **Why InterPARES stated that *"[…] users are not present at all in RiC-CM. Description serves two groups: the archivists […] and patrons"*?**

Finally, an InterPARES observation with which EGAD largely agrees. Users come in a variety of kinds and may differ considerably over time. Thus there are users of records in the context in which records are created and used in the activities of an Agent, and later users of the records in subsequent contexts.

Though a great deal of attention is not focused on users, as such, RiC-CM 0.2 has the following two paragraphs:

Under Audiences:

"Finally, RiC-CM is intended to be of interest to the research users of archives. Description based on RiC will enhance users' experience and understandings of records by enabling a fuller representation of the contexts within which the records were created and used over time. RiC

will also benefit researchers interested in using RiC to describe archival records from their own disciplinary perspective. Though RiC primarily focuses on description that is based on archival principles and responsibilities, it may be used to support scholarly descriptions of individual records or sets of records that are based on other perspectives and requirements."

And then later addressing various purposes of description:

"The description created to facilitate management and preservation also serves those interested in records as witnesses to life and work activities, to people, to human-made and natural events, to things made, things studied, things done, and more. Anything and everything may be the subject of records. For people who want to use records as evidence for scholarly, business, personal, or other purposes, description facilitates discovering, locating, identifying, retrieving, evaluating, and understanding them. Such ongoing use and reuse of the records becomes part of the history of the records; it re-contextualizes them. The use and reuse generate other records, thereby extending the social-document network."

It is worth noting that users of records are agents engaged in an activity, using records, and RiC-CM could be applied to describing these agents and their activities in relation to the records, as context.

About the standardization process

17. **As chair of EGAD, you can be seen as the conductor of the group. So how did you get to a consensus despite the different cultures represented and the language barrier ? What was the strategy adopted?**

The members of the group each have distinct cultural backgrounds and most members do not speak English as their first language. Though the members represent different archival traditions, the traditions represented all broadly embrace the Principle of Provenance, and so this is foundational in the think. Of course, there are significant differences in the understanding of Provenance and thus our starting point was to agree on a broad understanding, and from this we focused on the idea of "context". Along the way, we began to think in terms of the plural "contexts," as records do arise in a particular context, but throughout their existence, the context continues to dynamically change.

The language in which we conduct our work, English, is an ongoing challenge. It is particularly a challenge in teleconferences. The overall strategy is to remind native speakers of English to speak slowly, clearly. And to be patient in giving non-native English speakers an opportunity to find the words to express their ideas.

Language also presents a challenge with respect to the terminology we choose for various concepts. For example, there was early advocacy of the word "business" to describe the function or activities in with an agent engages. But this is a native English word and is not easily translated into Romance and other languages. We spent a lot of time carefully analyzing the etymology of terms, and we gravitated towards adopting Latin-based words, as, given the far reach and influences of the Roman Empire, the Latin-based terms tended to have acceptable equivalence terms in other languages.

18. Is the consensus adopted an acceptance of a predominant archival tradition or a subtle mix of specificities from different archival traditions?

I think the predominant archival tradition is the tradition that arose in the 19th century in Europe at the time of the emergence of modern nation states and modern historiographic methods. Ultimately, I think this is problematic, as noted in the text of RiC-CM:

"… there are many areas [of the world] with long and distinguished histories of administration and governance, and concomitant traditions of record creation, use, and management that are not represented [in the membership of EGAD]."

It is problematic because there is a clear Western bias that does not take into account other administrative traditions. There are complex historical factors, very much worthy of research, that may explain the dominance of Western administrative practices. Unquestionably, there is the dissemination of Western practices through colonialism, through imposition of such practices following wars, and economic and cultural domination. RiC-CM reflects this moment in time, and as we go forward, the dominance of the West may well recede.

As for RiC-CM, there is a mix of influences, though all derive from Western practices, including the Australian shift that emphasizes Activity (or function) for managing records, the "series approach."

19. By creating a conceptual model (which is a simplified way of explaining concepts) and by getting to a consensus (which has to reflect a shared understanding of the members), can we lose archival specificities (cultural specificities)?

This has been a longstanding critique of standardization. When I first began working with archives, the prevailing view was that each archive, the records it held, and its users were unique, and that standardization would prevent the archivist from addressing the uniqueness.

I gave a presentation in Edinburgh in the 1990's at an international EAD conference organized by someone opposed to the adoption of EAD. Her reason for objecting to EAD was that the English were using it! (Standards are indeed political, in many different ways). In my presentation I said that it was the records in our archives, the artifacts in our museums, the books in our libraries that documented the historical and cultural uniqueness of each of us as a people; and not the standards we used to curating these resources.

We feel that RiC-CM is sufficient general and flexible that the specificity of the provenance of the records described can be accommodated without obscuring the specificity. This said, we also recognize that the archival perspective is one among many, and thus like all perspectives employed in the description, understanding and interpretation of records it is reductive, which is to say, it does not fully reflect the irreducible complexity of the records as such. There are clearly many benefits to standardization, but there are also inherently the risks of distortion, of misrepresentation, of misunderstanding, of bias. Description and how we described needs to be subject to ongoing revision, correction, adjustments as our shared understanding develops.

About RiC-CM & RiC-O

20. In what way the Australian, the Spanish and the Finnish initiatives in terms of conceptual model have inspired the EGAD work?

Each had substantial influence on the model but given extensive deliberations and the synthesis following the deliberations, it would be difficult, in most cases, to pinpoint specific influences. Broadly, Australian theory had considerable influence on Activity (or function) and also Mandate. The Spanish model was influential with respect to how it was organized and presented, and in some areas also on specific definitions, for example, Rule. The Finnish model was influential in much the same way as the Spanish model. But there were also many, other influences from other cultural heritage standards, including InterPARES. The objective was to develop a comprehensive, coherent model that drew on what we thought were excellent ideas.

21. Have you been inspired by other national initiatives? Or from other professional communities?

The initiatives cited above would be the major sources of influence, though given the wide range of experience of the members, there are undoubtedly others. Various W3C recommendations, for example, OWL and PROV among them. Also the cross-cultural heritage preservation metadata standard PREMIS, and a variety of Linked Open Data initiatives.

22. And what about the ontology? Have you been inspired by other ones ? And if it is the case, what does EGAD take from one, what does it take from another?

Here again, there are a number of different influences. CIDOC CRM was of significant influence, as there are many very intelligent and clever approaches to particular modeling challenges. The Object Reuse and Exchange data model contributed a key component used in modeling the relation between a Record and Record Sets, the "proxy."

Florence Clavaud, as the primary author of RiC-O, will be able to provide you with a more extensive list of ontologies that informed the development. In most cases, the accurate characterization would be "adapted" rather than "take," as use of an idea almost always involves subtle differences in understanding.

23. Did EGAD use tools such as *user stories* (to reach an agreement about RiC features from an end-user perspective) ? Did EGAD use another kind of tools?

"User" is always an ambiguous term, as, broadly considered, there would be the "users" of RiC-CM in the course of professional work, and then there would be "users" of the description produced by the professionals. The latter usually being characterized as "end-users."

This said, we did not formally create "user profiles" or "use case narratives" for either type of user, though we did very much hold in mind a range of users of both types.

24. How did you choose, name, and define the key high level concepts : entities (Record, Agent, Activty, Instantiation), and their attributes (identifier, name, description, etc.)?

The names were often based on current terminology, but when we differed in defining the terms, or though the historical term imprecise or ambiguous, we chose another name. As mentioned above, we often took into consideration the challenge of translation, and in this regard leaned on Latin-based names.

The identification of the core entities was primarily based on an analysis of existing ICA standards augmented by reviewing a variety of professional literature. With respect to attributes, we again paid attention to current practices, but also our experience with a variety of projects, and attempted to identify new attributes when existing attributes did not address something we considered important, or we may have differentiated an existing attribute into more than one attribute when the existing was imprecise or ambiguous. We also spent a lot of time differentiating a Record from the physical Instantiation of the record based on the practical observation that in many cases, for preservation and other reasons, an original instance of a Record may be "re-instantiated" and the new instantiation was considered functionally equivalent, or not. We also spent a lot of time contemplating the description of instantiations in light of digital instantiations, and in relation to Premis, that does into more detail in describing the physical in support of preservation. We wanted to address intellectual high-level description that might be extended by Premis or other standards. The rationale being that the more detailed description of digital things is not archival context, but technical management data that is also an issue for libraries, museums and others that must manage and preserve digital instantiations. It may some time and experience to further develop where to draw the line between contextual description and technical management metadata.

25. What were the difficulties faced by EGAD members during this process?

The primary practical challenge is that everyone (or almost everyone) has a full-time job, and demanding jobs, and this made it a challenge to "steal" time to work on RiC. And then the rhythms of members lives differ, and so two or three people might find a good time to work, but a fourth person needed might not be available.

And there was also simply the intellectual challenges, the analysis, the discussions, language barriers. As challenging as these were, I think if you were to ask, the members will say it has been both exceptionally intellectually and social satisfying. In many respects, I think it the highlight of my professional life, and I have been fortunate enough to work with many intelligent, thoughtful, generous people over the years.

26. How did you hierarchically organize the key concepts? What was the approach taken to organize these concepts, what was the method followed by EGAD?

This is a complex question. In many respects it is simply an intellectual exercise broader/narrower hierarchical analysis. Such is a primitive in human conceptualizing of the world we experience.

27. How EGAD designed the relations? How did you imagine all the relations possible between the entities? Did you use special methods or tools?

This was perhaps this single most difficult challenge. Drafting this section was the brillian work of Florence Clavaud. Others and I provided considerable feedback, and Florence and I together refined the relation categories and hierarchy, as well as the names. Both Florence and I agree they are not perfect, but that we have advanced the boundary of our understanding in a systematic and coherent fashion. But it is inevitable that over time, there will be need for additional refinement and revision. The one observation about relations is that people have a tendency to identify possibilities endlessly, differentiating endlessly. And so we want to discipline the process by being intentionally minimal, and yet it is still a fairly large number. One thing we clearly avoided was identifying relations between persons that were qualitative and transitory: no "best friend forever," and such. Some relations are transitory, but none should be both transitory and qualitative.

28. How/in what way the development of the ontology can improve the conceptual model?

The ontology served to test the logic and completeness of the higher-level conceptual model. The formal requirements of ontology development is rigorous, and practicing such in the less formal conceptual modeling is exceptionally difficult to master and sustain.

29. During the call for comments, what advice did you receive to improve both the conceptual model and the ontology?

Well, 260 pages if they were to be printed. From 62 individuals or groups from around the world. The comments varied from helpful and actionable to not.

30. Why EGAD did not create a proof of concept for RiC (Deming PDCA approach) ? Is it due to a matter of time?

Over the years, I have designed a lot of standards and projects. Successfully. Though I never thought that I was applying Deming PDCA, the process was always iterative, one always identifies the objective, the functional requirements, and then begin thinking and designing how to fulfill the requirements, and then prototyping to test the design, and then revisiting the requirements. Over and over, sweeping back and forth. Developing the ontology served as a prototyping and testing, and it in turn was tested against existing description to ensure that it would accommodate existing description. We also attempted along the way to apply the model to an existing thing needing describing, another form of prototyping.

31. Did the US National Archives also create a pilot as the National Archives of France (NAF) with the PIAAF?

Not that I know of and I interact almost daily with two members of the NARA staff. And I served on the Advisory Committee on Electronic Record Archives for many years, until it was dissolved. NARA is active with the SNAC Cooperative, and in fact provides the training program for SNAC.

32. Is RiC compatible with the records management standards?

ISO 23081 is currently under revision and the ISO committee has requested meeting with EGAD to discuss collaborating. Unfortunately, the pandemic prohibited planning for a meeting. One member of EGAD, Adrian Cunningham is also on the ISO committee. From his perspective he feels RiC has "leap frogged" ISO 23081, and thus a revision of ISO 23081 necessarily needs to take RiC into consideration. Aligning the two would be great for both the archive and the records management communities, as both are records managers, and in the world of electronic records, they need to work cooperatively. I think.

33. How did you transform the conceptual model into an ontology? Which tools and methods did you use? What were the first steps?

Florence would be best to answer this question, as she did the bulk of the work. But in essence, the high-level ontology and a layer or two below come from RiC-CM, and then there is extensive detailed extensions to make the ontology both comprehensive, but also fully usable. There are a lot details and alternatives in RiC-O not found in RiC-CM. It is also very difficult in an ERM to be as expressive as is possible and necessary on an ontology.

As for tools, Florence used Protégé. It is a great tool, and several of us also use it. It enables easily reviewing and providing feedback. Reading raw OWL is not for the faint of heart.

34. Why did you decide to consider places as geohistorical entities?

This seems rather straightforward, as the places are geographic and they have a history, place as jurisdictional entities, but also structures as well as natural geographic features. They exist, are identifiable, have histories, and can be geo-located (over time if necessary). A place has an identity.

35. Why did you choose to consider some relations as entities in their own right (with a date and a description)? What is the point behind this choice?

The relations are relations. Date might be considered either an attribute or an entity. We chose to treat it as an entity simply because representing date, particularly in a digital environment, has necessary complexities, and to express these complexities, you need attributes. It is a practical decision more than an intellectual decision. The method of modeling dictates some decisions. Further, when you move from the CM to the O, some things modeled

one way shift to being modeled another. And so, because of the nature of OWL and ontological primitives, CM relations become O classes. Mostly.

36. Are all the relations considered as entities in their own right?

In RiC-O, yes, but not in RiC-CM.

37. How did you align archival concepts with the ones of libraries and museums (which methods, technical solutions)?

In RiC-CM, the relation is in a comment; in RiC-O, the relation is formally expressed using OWL or a related standard. Mapping and alignment is an ongoing activity. With respect to methods, this is a matter of analyzing other conceptual models and ontologies and making a judgement with respect to the degree to which a concept in RiC is similar or the same as a concept in another model. In general, we are disinclined to say a concept in RiC is the same as a concept in another model, as there are often subtle though significant differences, as such alignment is always arguable.

38. Why EGAD states that "*RiC is intended to be of interest to the research users of archives, in particular to scholars interested in reusing archival records*"?

Though RiC is specifically targeted at records managers, either in the context of origination of records, or subsequently in archives, the model is sufficiently flexible to be applied by other perspectives. In particular, based on our experience and familiarity with digitial humanities scholarship, the concepts of Record Set may be employed to identify sets of records selected and group together for any scholarly purpose. Such sets, for example, may have Record members from selected from two or more fonds because they document different activities or subjects. Such use by other than records managers may require extension RiC concepts in order to accommodate particular research interests and activities. Having had extensive experience with digital scholarship projects both in design and implementation, the distinction between a Record and an Instantiation will be useful, as many projects use digital surrogates of records of original records held in repositories. A good example is the Walt Whitman Archive, a documentary editing project, that has digital surrogates of Whitman records (manuscripts) held by nearly 80 different repositories.

About RiC-O Converter, the ontology converter

39. The National Archives of France has developed "RiC-O Converter" to convert EAD finding aids to RDF finding aids. Why don't you create such a tool within EGAD?

The development of such tools is outside of the scope of the EGAD mandate, and properly belongs in the archival and records management communities or vendors that provide services to them.

40. Will each institution have to develop its own software/its own tool to convert its finding aids? Why?

No, as the French developed converter can be adapted by other users of EAD or EAC-CPF. While a fundamental purpose of EAD and EAC-CPF is for the community to create standard, consistent archival description, each is sufficiently flexible that there are often differences from one implementation to another that make it a challenge to write one conversion script that will work for all. But … one conversion script can work mostly for all, provided the base script is adjusted or adapted, at the margins, to the encoding practices employed in a specific implementation. I know this because I have written conversion scripts that attempt to uniformly process EAD EAC-CPF instances from multiple sources.

About the RDF format

41. According to you, what are the advantages of the RDF format for the information retrieval?

The key value of RDF lies in being able to interrelate descriptions in cultural heritage by way of shared (widely and commonly used) authority files and controlled vocabularies. For example, using Wikidata and other identifiers to interrelate archival records, museum objects, and library objects when the creator of the records and objects is the same person. A good example of effectively doing is can be found in the BnF project data.bnf.fr wherein the different resource descriptive systems of different curatorial domains are interrelated. Of course, this is more easily done within a single institution that controls all of description system components that it is accomplished in the wide-open world on the Internet, where the various nodes are independently controlled. But making it work in the wide-open world is a key aspiration of Linked Open Data. But it is more easily imagined and said, than done. Nevertheless, the quest is leading to cross-domain dialogue, cooperation and collaboration that is good in itself, and may, over time, lead to vastly improved information discovery for those seeking information.

42. And what about its advantages for research?

RDF and ontologies (both formal domain ontologies such as RiC-O and CIDOC CRM) and ad hoc syncretic or eclectic ontologies (such as data.bnf.fr) that are cobbled together from diverse existing ontologies, are being employed in researcher/scholar driven projects, for example, in Canada, the Orlando Project, and in the U.S. the Enslaved project. Each of them is developing an project specific ad hoc ontology.

And with respect to the research discovery activity, see Q1 above. Ultimately, what is implicit in the Linked Open Data + ontologies movement is a quest for integrated discovery of artifacts of all kinds that serve as the basis, narrowly, for research and scholarship, and more broadly for anyone seeking information out of curiosity or need.

43. Do you think that the RDF format could favour serendipity?

Sure. Whatever we humans put together to facilitate discovery will inevitably lead to people being able to identify "connections," as a product or byproduct of the discovery system.

I think in many respects we may be making what was once serendipity (an accidental byproduct, say, of going the library shelves to find a specific book, but because library classification and shelving conventions bring similar books together, discovering a book that we judge to be better than the one we sought) less accidental, making explicit the relations between things that were implicit. For example, in SNAC, by intentionally mining and extracting data from diverse sources and attempting to surface the implicit social-document network, we surface relations that a researcher would have had to piece together going from one discovery system to another and following clues. Two scholars that SNAC employed as advisors, Ed Ayers, an expert on the American Civil War, and Kenneth Price, a Walt Whitman scholar, immediately discovered in SNAC resources that years of research had not revealed. But this said, serendipity will always emerge, I think, at the margins. Systems designed with a specific objective in mind frequently lead to accidental discoveries. I think we make want was once serendipitous discovery increasingly systematically discoverable, but there is always, I think, a serendipitous or accidental discovery dimension remains, as humans will identify relations in interpreting the information that were not explicitly identified. The phenomena being descriptively represented are irreducible in their complexity.

About the graph representation

44. The objective is to create an independent platform for the graph visualisation or to create a functionality which could be integrated into an archival information system?

I think our understanding of the utility of graph visualizations if nascent. For the most part, I think the visualization are often aesthetically pleasing, but not particularly useful. In a very abstract way, they do convey the intricate ways in which people are related to one another, and to the artifacts that document their lives.

The more important value of graph representation of information is that it is more expressive of the reality that we are trying to describe. Still reductive, ultimately, but less so than markup and relational database technologies. At the same time, to employ effectively, graph technologies require a much greater degree of precision, of specificity than other methods of representation. They are, thus, more labor intensive, time consuming to implement effectively.

45. If a platform is created : what stage has been reached in the creation of this platform? What can you say about its progress/advance ? What are the difficulties faced by the EGAD members?

Florence Clavaud is more familiar than I with nascent implementations of RiC. From my own perspective, we are implementing some RiC compatible features in SNAC, in particular a

subset of the RiC relations, though systematically extended to serve semantics desired by the SNAC community.

46. What are the advantages of graph representation for the understanding of the archival resources (from a user point of view)?

I addressed this above in Q1, I think. Though I would add, for the user, a fuller understanding of context, which is essential in identifying, understanding and interpreting records. Also the broader interrelating of records beyond the current self-contained, isolated descriptions of fonds, series, and so on, that are currently not interrelated other than implicitly. Records exist in the world related to other Records, and the relations transcend the traditional fonds, series and so on. Any reduction we make is, well, a reduction, and thus it leaves relations implicit rather explicit. A graph approach to description enables us to represent the fact that records are part of a vast, global social-document network. I think users have known this implicitly, as many researchers spend years trying to piece identify both social and document entities and piece them together by identifying relations of many kinds.

47. From the user point of view, how to deal with such big graphs? How to deal with such large information sets?

One step at a time.

2.2. Réponses de Madame Florence Clavaud

RiC-O

1. **Comment le prototype du British Museum « Research Space » vous a inspiré pour RiC-O ? Quels points en particulier ?**

Nous ne nous sommes pas inspirés de Research Space pour RiC-O. Je pense que vous voulez parler de PIAAF et non de RiC-O ici, mais même si vous parlez de PIAAF qui a été conçu et réalisé en 2017, Research Space, qui est une plateforme ne nous a pas servi de source d'inspiration.

RiC-O est un modèle de métadonnées (comme EAD ou EAC-CPF ou encore FRBRoo), pas un outil.

Comme il s'agit d'une ontologie de domaine, pour le construire, tout comme pour construire RiC, on s'attache à identifier quelles catégories d'entités existent dans le domaine concerné, ici l'univers des archives, quelles sont les caractéristiques de ces entités (comment on peut les définit formellement) et quelles relations elles sont susceptibles d'entretenir entre elles. Toutes ces informations sont ensuite à représenter formellement en systèmes poly-hiérarchiques de classes et de propriétés, à organiser convenablement pour construire un modèle formellement pertinent. Cela demande de connaître le langage formel de représentation à utiliser (en l'occurrence OWL, qui intègre lui-même les composantes du langage RDFS, qui lui-même utilise les briques de base définies dans RDF), donc bien évidemment de beaucoup travailler pour s'approprier ce langage, comprendre les impacts des choix que l'on fait en terme d'inférences logiques, etc. Cela demande aussi de tester ce que l'on fait, de le confronter avec le monde réel.

2. **Vous êtes-vous inspirés d'autres ontologies ?**

Comme pour RiC-CM (pour lequel nous avons commencé par étudier les normes de métadonnées existants pour les archives, outre les normes du CIA, telles que par ex. la norme espagnole CNEDA, puis regardé les modèles élaborés par d'autres communautés), bien sûr, nous avons étudié attentivement divers modèles ontologiques, comme CIDOC-CRM, FRBR, PROV-O, PREMIS, etc. L'objectif était cependant, encore une fois, de définir une ontologie couvrant l'ensemble des besoins des professionnels des archives. Aucun de ces modèles préexistants ne convenait, ces modèles préexistants n'avaient pas les archives comme point central d'intérêt, ne définissaient par les archives avec la précision nécessaire et selon la perspective métier qui était la nôtre. Nous avons donc élaboré un modèle conceptuel et une ontologie de domaine. Cela ne veut pas dire que les composants de RiC ne seront pas mappés avec ceux d'autres modèles. La feuille de route de RiC-O inclut, pour l'élaboration de la version 1.0 normative qui sera publiée fin 2021, l'établissement de mappings (correspondances, et notamment équivalences) avec les composants d'autres ontologies. Voir à ce sujet les pages suivantes : https://ica-egad.github.io/RiC-O/why-use-RiC-O.html et https://ica-egad.github.io/RiC-O/next-steps.html

Représentation sous forme de graphe

3. **Selon vous, quel est l'intérêt de la représentation sous forme de graphe pour la compréhension des ressources archivistiques ?**

Voici quelques éléments de réponse :

Pour l'utilisateur final, un graphe d'entités liées a pour premier avantage de dénommer et définir clairement (c'est d'ailleurs le sens du mot « sémantique » dans l'expression « web sémantique ») ce dont on parle. Tous les composants de RiC-CM et de RiC-O ont un nom et une définition. Toutes les entités que l'on trouve dans un graphe archivistique, par ex. une personne ou un lieu, devraient également recevoir une description riche (par les attributs ou relations qu'on leur assigne).

RiC rend explicites des relations qui sont implicites ou semi-implicites dans les instruments de recherche actuels. Par exemple (et ce n'en est qu'un parmi tant d'autres), la relation d'inclusion (d'une série dans un fonds, d'un dossier dans une série, d'une pièce dans un dossier) est représentée dans un instrument de recherche ISAD(G) classique par une pure convention typographique (l'indentation) et/ou par une table des matières, présente soit après le corps de l'IR, soit à côté de ce corps dans une vue HTML. Par ailleurs, dans un fichier EAD, les archivistes ont choisi de la représenter par l'arborescence XML, mais ce que représente cette arborescence n'est pas explicité dans un fichier EAD. Pour la relation séquentielle (une série vient après une autre) c'est pareil : seules des conventions de présentation et/ou un choix technique tacite les expriment. Dans un graphe d'entités liées conforme à RiC, ces relations sont clairement explicitée.

RiC a aussi l'intérêt de se concentrer sur le monde décrit, pas sur les règles de description. Il y est question de Record Resource, Record Set, Record et Instantiation, pas d'unité ou de niveau de description. Dans les instruments de recherche classiques, les métadonnées que l'on présente avec soin ont finalement tendance à 'masquer', d'une certaine manière, la réalité documentaire. RiC invite à remettre le focus sur les documents. Par exemple, avec RiC, on peut représenter un contenu informationnel (Record) comme étant lié à de zéro à n Instantiations (zéro, lorsque le contenu informationnel est connu mais que la manifestation physique a disparu, comme dans le cas d'une charte originale ou d'un ensemble de documents d'archives privées qu'on ne connaîtrait que par un inventaire du 19e ; une instantiation, si l'on a affaire à des originaux sur papier ; une deuxième instantiation si on en a une copie à l'identique telle qu'une image numérique). RiC considère par ailleurs les instruments de recherche ou notices descriptives comme… des documents (Record). Ainsi un Record peut être associé par la relation 'is or was described by' à autant d'instruments de recherche-documents qu'il en existe pour le décrire.

RiC permet enfin une plus grande précision. Avec 22 entités, 41 attributs et 78 relations (plus les relations inverses) on dispose d'un arsenal permettant d'atteindre une granularité bien plus importante qu'avant.

La compréhension des ressources archivistiques est aussi affaire d'ergonomie et de sciences cognitives, donc d'interfaces. Ainsi, à partir d'un graphe d'entités liées, il est possible (mais c'est affaire de conception et de développement applicatif) de montrer, au sens propre du terme, par des interfaces agréables et dynamiques, au sein d'un diagramme, toutes les entités de contexte d'un document, en incluant dans l'interface une ligne de temps afin de représenter aussi la dimension temporelle, essentielle, dans laquelle le document et ces entités ont existé à un moment donné.

4. **Selon vous, quel est l'intérêt de la représentation sous forme de graphe pour l'exploitation des ressources archivistiques ?**

Voici quelques éléments de réponse.

Dans un graphe d'entités liées, les entités de contexte des documents ou ensembles documentaires (les agents, les lieux, les fonctions et activités…) deviennent autant de points d'entrée différents et intuitifs vers la description des documents.

Un moteur de recherche adapté (utilisant SPARQL, si le graphe d'entités liées est stocké en RDF) peut parcourir ce graphe en interrogeant aussi bien les nœuds (entités) que les relations, ce qui permet de poser des questions qu'il est extrêmement difficile voire impossible de poser dans un SIA classique, par ex. : donne-moi la liste de toutes les (entités) collectivités *qui ont dans le temps, succédé* à tel bureau *dans l'exercice de* telle (entité) fonction ; ou bien donne-moi les (entités) documents *produits par* des (entités) organisations créées avant telle (entité) date et *après* telle (entité) date et *ayant (si on le connaît, donc optionnellement)* telle (entité) statut juridique, et affiche aussi leurs fonctions, leur organisme de tutelle et leur historique si elles en ont un.

Les jeux de données liées ouverts (Linked Open Data) peuvent aussi (devraient, logiquement) être rendus accessibles à distance et exploitables par des applications tierces, au minimum par SPARQL (qui est aussi un protocole de recherche), idéalement par des APIs dédiées. Les possibilités d'exploitation sont donc démultipliables.

5. **Du point de vue de l'utilisateur, comment traiter de grands ensembles d'information ?**

Nous avons déjà de grands ensembles de métadonnées (ainsi aux AN nous avons près de 30000 instruments de recherche, plus de 15000 notices d'autorité et une vingtaine de vocabulaires), nos métadonnées augmentent chaque jour, et nos systèmes sont capables de gérer et de publier ces grands ensembles. Cependant il peut s'avérer très gagnant de faire évoluer les modèles de métadonnées, les infrastructures de stockage et de gestion (trop cloisonnées) et les applicatifs en front-end, et d'ouvrir les silos de métadonnées vers l'extérieur.

L'idée générale qui conduit à envisager ces évolutions est d'améliorer la qualité des métadonnées, ce qui implique aussi de travailler à leur précision (par ex.. indexer plus

systématiquement la description) et à la granularité de leur structuration, pour en faciliter l'utilisation. RiC peut largement aider, comme cadre général de référence, à avancer dans cette direction. Passer d'un modèle et d'un format de métadonnées à un autre plus précis et plus riche (par ex. d'EAD à RiC ou RiC-O) peut se faire automatiquement par script (cf. par ex. le projet RiC-O Converter, qui est performant, fiable et rapide). Mais ces processus purement automatiques, même très intelligemment conçus, ont leur limite, on ne peut pas tout attendre d'eux ; on peut gagner en expressivité (expliciter des relations, faire émerger des entités) mais on ne peut que difficilement enrichir par ces processus le contenu des métadonnées préexistantes. Le travail d'amélioration de la qualité des métadonnées reste donc avant tout un travail humain qui demande des investissements en temps. Il peut aussi se faire en partie par des technologies de machine learning (reconnaissance d'entités nommées par ex.) , là aussi les processus ne sont jamais entièrement automatisés actuellement.

La notion de qualité conduit immanquablement, pour définir et mener à bien des programmes réalistes, à réfléchir au niveau de qualité suffisant (*good enough*) pour l'utilisateur. Cette question est complexe et sort du cadre de cet interview. Il est clair qu'aujourd'hui l'utilisateur des services d'archives a besoin de solutions plus performantes de recherche (plus de clarté, moins de bruit, moins de silence). Plus les métadonnées seront riches, indexées avec des référentiels d'entités précisément décrites, et finement structurées, plus les interfaces de recherche (construction dynamique du questionnaire de recherche, critères optionnels de recherche, extension de la recherche sur un concept à la recherche sur un concept plus générique…) et de navigation (présentation en onglets par type d'entités, navigation au sein des référentiels ou au sein du graphe obtenu pour un résultat de recherche donné…) gagneront en clarté, en fluidité et en flexibilité, plus l'utilisateur y gagnera. Des dispositifs simples d'extraction de jeux de métadonnées dans différents formats depuis un panier de recherche devraient aussi faire partie des dispositifs.

6. **Savez-vous d'ores et déjà si des outils de partitionnement de graphes seront proposés ? Si oui, lesquels ? (bipartition de graphes, méthode hiérarchique, algorithmes de *clustering*, etc.)**

Cette question ne concerne pas RiC ou RiC-O à proprement parler, et dépasse de loin le cadre de cet interview. Elle est du ressort des applications qui vont manipuler et afficher les graphes, pour lesquelles beaucoup de choses restent à explorer (R&D). Dans le cadre du projet PIAAF nous avions commencé à y réfléchir.

RiC aux AN, PIAAF

7. **Au sujet de l'expérience PIAAF, vous dites avoir beaucoup appris de la pratique et du point de vue des institutions partenaires et *« pu mesurer combien ceux-ci déterminent les métadonnées créées »*[581]. Pouvez-vous expliciter s'il vous plaît ?**

Les métadonnées, quelles que soient la rigueur scientifique qui préside à leur production, reflètent toujours le point de vue et/ou les besoins de leurs auteurs.

Par exemple, telle administration décrite dans un référentiel de bibliothèque sera considérée en tant qu'auteur sur la longue durée et décrite très rapidement, tandis qu'un service d'archives sera beaucoup plus sensible à l'évolution de ses fonctions et de son environnement administratif, déterminants pour comprendre sa production documentaire ; de ce fait pour une seule entité dans le référentiel de la bibliothèque on en trouvera peut-être trois ou plus, liées par des relations chronologiques, dans le référentiel du service d'archives. Un lieu pourra être considéré simplement comme un concept dans un méta-portail, et sera alors représenté comme tel dans un thésaurus géographique conforme à la norme ISO 25964 et utilisant le format RDF SKOS. Dans le contexte d'un projet de recherche en histoire ou d'une institution patrimoniale conservant des documents anciens, le même lieu sera vu et modélisé, non pas comme un concept mais comme une entité géo-historique à part entière, ayant des attributs spatio-temporels.

Voir aussi la section 1.3. Premier bilan de ces opérations, dans la documentation PIAAF (https://piaaf.demo.logilab.fr/editorial/contexte-technique#volumetrie-corpus).

8. **Comment les Archives nationales utilisent-elles les premières versions de RiC actuellement ?**

Les AN utilisent la dernière version en date de RiC et de RiC-O. J'ai déjà eu l'occasion de donner des éléments de réponse à cette question : PIAAF ; conversion et enrichissement des référentiels ; conversion des fichiers EAD et EAC-CPF ; travaux en cours pour préfigurer un data.archives-nationales.fr ; fourniture de jeux de données RDF à des projets de recherche externes ; travail sur les métadonnées de description des archives numériques, etc. Voir la bibliographie et notamment l'article publié dans la Lettre de l'AAF, n° 136, p. 24-25.

9. **Qu'est-ce que RiC a changé aux Archives nationales (SIA, habitudes de travail, du point de vue des utilisateurs) ?**

---

[581] Florence Clavaud, Pauline Charbonnier, *Records in Contexts aux Archives nationales : enjeux et premières réalisations*, *op. cit.*, p.22.

Mis à part PIAAF, les utilisateurs n'ont pas encore eu affaire à des graphes d'entités liées conformes à RiC aux AN. La nouvelle interface de la base de données Léonore (https://www.leonore.archives-nationales.culture.gouv.fr/ui/) utilise cependant des données sur les lieux (notamment leurs coordonnées géographiques) issues du référentiel des lieux des AN en RDF. L'interface web de recherche dans les métadonnées RDF du projet ALEGORIA n'est pas encore disponible (elle le sera cette année). Plusieurs autres projets de recherche sont en cours sur des jeux de métadonnées des AN qui seront rendus accessibles au travers d'applications RDF dédiées, dans les prochaines années. Voir aussi l'article déjà cité, p. 25, et le démonstrateur mentionné dans l'interview oral.

Au sein des AN, les travaux relatifs à RiC sont essentiellement du ressort de la Mission référentiels. Les impacts sont encore relativement faibles sur la vie quotidienne des autres services. Cependant, dans le cadre des projets de recherche cités ci-dessus, plusieurs services des AN ont commencé à travailler avec la Mission référentiels dans le but de produire des métadonnées EAD plus riches et plus finement structurées qu'habituellement, qui sont ou seront ensuite converties en RDF/RiC-O. Cf. aussi les travaux qui sont prévus cette année sur les archives nativement numériques, et le projet de logiciel de reconnaissance d'entités nommées. Cf. aussi tous les programmes relatifs à la construction de référentiels (fonctions notamment).

Globalement, depuis deux ans, au fil de divers projets comme PIAAF ou RiC-O Converter, ou encore d'événements comme les datathons et hackathon, les collègues producteurs de métadonnées aux AN (qui sont environ une centaine d'archivistes), ont pris conscience de la nécessité d'améliorer la qualité des métadonnées produites, notamment par l'indexation. C'est bien sûr plus largement aussi un enjeu pour tout le réseau des services publics d'archives (cf. le *Guide d'indexation* en cours d'élaboration par un groupe de travail monté par le SIAF).

Le plan stratégique 2021-2025 des AN inclut l'élaboration et la mise en oeuvre d'un plan qualité pour les métadonnées, la restructuration, l'enrichissement et la publication des référentiels, l'inscription des métadonnées dans le web de données, parmi ses objectifs. Cela passera forcément, et au minimum pour les référentiels, par des changements importants au sein du SIA dans les infrastructures de stockage, les modèles de données et les fonctionnalités.

RiC, le portail FranceArchives & le portail européen des archives

10. Est-il déjà utilisé ? Si oui, de quelle manière ?

La liste des utilisations actuelles de RiC dans le monde est là : https://ica-egad.github.io/RiC-O/projects-and-tools.html.

Tout n'y est pas forcément précisément détaillé, d'autre part divers projets ne peuvent pas encore y être mentionnés car au stade de la réflexion préalable, de la recherche de financement ou de la passation de marché. Enfin pour rappel RiC est encore un draft ; souvent les institutions ou projets qui s'intéressent à RiC attendent la version 1.0 normative pour vraiment lancer des opérations concrètes.

Comme déjà dit et écrit ici et ailleurs, il faudra plusieurs années à FranceArchives pour « passer à RiC-O ». En ce qui concerne Archives Portal Europe, je n'ai pas d'information ; la responsable est informée et a d'ailleurs appris avec beaucoup d'intérêt la décision de FranceArchives.

J'ajoute que le comité ad hoc de la Society of American Archivists (TS-EAS) qui maintient les modèles EAD et EAC-CPF et dont je fais partie, va élaborer pour 2025 une nouvelle version majeure d'EAD, qui prendra en compte, entre autres, RiC.

RiC-O Converter

11. **Quelle est la politique d'URI pour les Archives nationales (stratégie de construction et de gestion d'URI) ?**

Voir à ce sujet le PDF de la présentation faite le 28 janvier 2020 (https://f-origin.hypotheses.org/wp-content/blogs.dir/2167/files/2020/02/20200128_4_RiCOConverter.pdf). Nous avons défini une stratégie que nous appliquons afin de générer des URIs RDF pour les données RDF que nous produisons. Nous nous appuyons sur les identifiants locaux uniques que le SIA, par chance, génère systématiquement et gère de façon fiable, pour les unités de description des fichiers EAD et les entrées de l'ensemble des référentiels. Lorsque le SIA ne fournit pas d'identifiant (cas par exemples des noms des agents ou des relations entre entités EAC-CPF), le logiciel RiC-O Converter génère un identifiant à partir d'autres infos présentes dans les métadonnées.

Ceci ne concerne pas l'ensemble des ressources décrites par les AN.

Globalement les AN doivent encore consolider et étendre leur stratégie d'assignation d'URI ; par ailleurs plus concrètement cela impliquera sans aucun doute de doter l'institution d'un dispositif robuste et flexible, fondé probablement sur ARK, et capable de déréférencer convenablement les URI selon leur contexte d'utilisation comme l'a fait la BnF.

12. **Les Archives nationales doivent échanger des données sur les agents avec la base ISNI. Qu'est-ce qu'être membre du consortium ISNI apporte aux Archives nationales du point de vue de sa politique d'URI ?**

Rien en fait !

La base ISNI fournit un identifiant pour un nombre conséquent de personnes et collectivités, que nous stockons dans nos notices d'autorité (élément <entityID> des notices EAC-CPF). Mais nos notices ont leurs propres identifiants, qui nous servent de base pour générer des URIs pour les agents que nous décrivons. Dans les jeux de données RDF que nous produisons, les identifiants ISNI sont bien sûr conservés, mais pas pour construire les URIs. Voir par exemple :

https://github.com/ICA-EGAD/RiC-O/blob/master/examples/examples_v0-2/NationalArchivesOfFrance/src/eac-cpf/FRAN_NP_050789.xml
et
https://github.com/ICA-EGAD/RiC-O/blob/master/examples/examples_v0-2/NationalArchivesOfFrance/rdf-xml/agents/FRAN_Agent_050789.rdf

Les identifiants ISNI sont par contre des clés de liage importantes pour par exemple aligner nos notices avec celles de la BnF (si une de nos notices stocke un id ISNI et qu'une notice de la BnF stocke le même ISNI, alors les deux notices peuvent a priori être considérées comme décrivant la même entité). Voir aussi le démonstrateur PIAAF (https://piaaf.demo.logilab.fr/alignments)

13. Dans la présentation de RiC-O le 28 janvier 2020, il est dit au sujet des URI que *« des choix provisoires ont dû être faits »*. Pouvez-vous expliciter s'il vous plaît ?

Les choix sont dits provisoires dans l'attente d'un dispositif technique dédié, global, robuste et flexible, cf. plus haut. La stratégie que nous avons adoptée y sera très probablement totalement conservée.

Référentiels des AN

14. Quelle est la situation actuelle des référentiels des Archives nationales ?

Je ne peux pas passer des heures ici à réécrire ce que j'ai écrit ailleurs, désolée. Voyez notamment le support et l'enregistrement audio de mon intervention de février 2021 lors d'une journée d'étude à Lille, https://geriico.univ-lille.fr/detail-event/le-web-semantique-et-le-patrimoine-culturel-de-la-convergence-des-donnees-au-croisement-des-connai/ ainsi que l'article que j'ai publié dans le n° 254 de la Gazette.

15. Est-ce que tous les référentiels - ou du moins les principaux - ont une structure hiérarchique (thésaurus) ?

Les référentiels dits d'indexation ont actuellement, soit la forme d'une liste plate d'autorité, soit la forme d'un thésaurus.

16. Comment parvenir à une structure hiérarchique : procédé(s)/outil(s)/méthode(s) ?

Pour les référentiels pour lesquels passer à une structure de thésaurus est appropriée, hiérarchiser est un travail scientifique qui revient à se demander si tel concept peut être

considéré comme spécifique d'un autre. Puis on représente formellement cette hiérarchie, en utilisant SKOS. En l'occurrence, aux AN, nous utilisons à la fois SKOS et RiC-O pour structurer nos référentiels d'indexation. Un type de document, par ex. y est défini comme étant à la fois une instance de la classe Concept de SKOS et une instance de la classe Documentary Form Type de RiC-O.

17. Quel est l'intérêt des thésaurus pour la consultation et l'exploitation des ressources archivistiques ?

Cette question dépasse le cadre d'un interview sur RiC. En sciences de l'information, disposer de thésaurus normalisés, c'est-à-dire conformes à la norme ISO 25964, est considéré comme important pour faciliter l'indexation et améliorer sa qualité (puisque les notions du domaine y sont représentées clairement, précisément définies, associées à des termes choisis pour les véhiculer, y compris des termes alternatifs, dotées de notes d'application, reliées entre elles et mappées avec des notions d'autres thésaurus) et la recherche d'information. (puisqu'on peut alors, éventuellement, permettre à l'utilisateur final d'accéder directement au thésaurus, de naviguer dans sa structure, de consulter les définitions etc. avant de choisir un concept pour lancer une recherche , ou encore, proposer à l'utilisateur d'étendre sa recherche aux notions liées à une notion donnée).

18. A quel stade se situe l'avancée de la conception d'un référentiel des agents unique ? Quelles sont les étapes à suivre pour y parvenir ?

Pour l'instant, nous sommes en attente des évolutions techniques, internes au SIA, qui permettront d'implémenter ce référentiel. En attendant, nous accumulons hors du SIA des notices descriptives d'agents que nous importerons dès que ce sera possible. Par ex. ce printemps nous allons travailler à produire des notices d'autorité à partir des 10000 premiers enregistrements vérifiés de la base Léonore. Nous travaillons également, hors du SIA, à enrichir les quelques 3000 notices d'agents qui sont actuellement stockées dans les 2 référentiels personnes physiques et morales, hors du référentiel des producteurs. Un groupe de travail vient par ailleurs de se monter aux AN pour réfléchir au sort de plusieurs bases de données nominatives anciennes ; notre proposition est d'en extraire des notices sur les personnes que ces bases mentionnent. Nous avons enfin toujours en chantier la construction d'un référentiel national des notaires.

19. Au sujet du référentiel agents, vous avez dit lors de la journée d'étude du 28 janvier 2020 qu'il y avait une *« nécessité de faire évoluer les fonctionnalités, l'interface et le modèle de données »*. Pouvez-vous expliciter s'il vous plaît ?

Il faut passer de trois référentiels à un référentiel unique, utilisable depuis n'importe quel endroit dans n'importe quel IR pour indexer un agent, quelle que soit la relation qu'il a avec les documents décrits, alors que ce n'est pas possible aujourd'hui (changement dans les fonctionnalités). Il faut une interface d'édition des notices qui permette pour certains agents de produire des notices plus légères que celle des producteurs (évolution du modèle de données et du workflow de validation). Il faut un référentiel des relations pour lier les agents à la description des documents, ou à celle d'autres agents. Ensuite, au sein de la SIV, il faut rendre directement interrogeable ce référentiel des agents.

20. L'idée de développer un référentiel national couvrant les fonctions des collectivités territoriales a été évoquée lors de cette même journée d'étude. Avez-vous pu commencer à le développer ? Si oui, quelles sont les méthodes suivies pour sa mise en œuvre ?

Pour l'instant nous travaillons à la construction d'un référentiel des fonctions des collectivités de l'Etat, pour les Archives nationales et pour tout l'écosystème des services versant des archives aux AN (missions des archives dans les ministères, opérateurs nationaux). Ce travail collectif et collaboratif est long et complexe et va nous prendre encore quelques années.

21. Lors de la rencontre annuelle des correspondants archives en 2016, vous aviez dit que *« l'interopérabilité se construira aux Archives nationales quasiment exclusivement sur les référentiels »*. Pouvez-vous expliciter s'il vous plaît ?

Voir mon article dans le n° 254 de la Gazette, notamment la section « Se servir des référentiels pour créer de l'interopérabilité interne et externe ». Cf. la phrase suivante : « : « ces référentiels serviront à créer de l'interopérabilité ; ils sont globalement les seuls à pouvoir jouer ce rôle, les documents conservés étant le plus souvent, contrairement aux livres conservés en bibliothèque, des objets uniques. »

22. Sur quel cadre normatif vous appuyez-vous pour la production des référentiels ?

RiC-CM et RiC-O ; ISO 25964 et SKOS.

23. Quel est le processus d'application de la norme ISO 25964 ?

Eh bien comme déjà dit nous tâchons de rendre nos vocabulaires, nouveaux ou existants, conformes à cette norme. Comme dit plus haut cela implique au minimum de déterminer de quels concepts nous avons besoin pour décrire les documents conservés, de donner une définition à chaque concept, voire une note d'application et des exemples, de lui associer les

termes qui vont l''exprimer, de le positionner dans une hiérarchie de concepts, de le lier éventuellement à d'autres concepts du même référentiel ou d'autres référentiels. Puis de tester ce qui a été créé en voyant ce que cela donne quand on index des instruments de recherche, d'ajuster si nécessaire, etc. Cela se passe par itérations et cela mobilise pas mal de monde.

24. Quels sont les résultats de l'étude menée courant 2020 pour évaluer les impacts d'une évolution du SIA des Archives nationales vers le support du format SKOS ou d'un format compatible avec ISO 25964 ?

Nous sommes toujours dans l'attente de cette étude technique détaillée.

25. Comment utilisez-vous la plateforme GINCO du Ministère de la Culture dans l'élaboration des vocabulaires contrôlés ?

Lorsque nous n'avons pas le choix (lorsque nous ne pouvons pas travailler dans le SIA parce que le modèle de données ne convient pas ou que nous souhaitons travailler avec des personnes extérieures aux AN) il nous arrive d'utiliser la plateforme Ginco du ministère, pour produire un thésaurus. Nous pouvons ensuite exporter au format SKOS ce thésaurus, et l'enrichir d'autres données. Il reste alors à réimporter dans le SIA les données en question.

26. Comment appliquez-vous la feuille de route sémantique du Ministère de la Culture ?

Nous l'appliquons par toutes nos actions, qui s'inscrivent totalement dans cette feuille de route dans toutes leurs composantes ; sans oublier bien évidemment l'élaboration du modèle RiC-O. Voir toutes mes réponses ci-dessus… Nous participons aussi à l'élaboration de cette feuille de route.

27. Selon vous, en quelle mesure la consultation directe des référentiels améliorerait-il la recherche au sein des fonds d'archives pour l'utilisateur final ?

La consultation directe, déjà possible, du référentiel des producteurs est déjà une porte d'entrée utile dans la SIV des AN, notamment pour faire des recherches au sein des archives des notaires parisiens, qui est entièrement décrit dans ce référentiel. Lorsque ce référentiel des producteurs aura été enrichi des notices décrivant toutes les entités d'administration centrale du Moyen Age à nos jours, au plus haut niveau (ministères, gouvernements, cabinets, directions techniques et opérationnelles, opérateurs nationaux, autorités indépendantes, juridictions…) il rendra encore plus de services. Cependant pour penser à l'utiliser il faudra que l'utilisateur ait

CC BY NC ND

en tête le fait que les documents sont groupés par provenance, ce qui n'est pas le cas de la majorité de nos usagers.

Il est évident, de mon point de vue, qu'un accès par lieux, par types de documents ou par fonctions, pour ne citer que ces référentiels, apporterait beaucoup. Voyez déjà l'organisation par onglets (par types d'entités) dans PIAAF et imaginez des onglets « lieux », « agents » ; « fonctions », etc. Voyez aussi mon article dans le n° 254 de la Gazette. Voyez enfin, à titre d'exemple encore peu ergonomique, la consultation du référentiel des fonctions sur le site web des AN d'Australie (cliquer sur Advanced search) : https://recordsearch.naa.gov.au/SearchNRetrieve/Interface/SearchScreens/AdvSearchMain.aspx.

Projet ALEGORIA

28. L'enjeu de ce projet est de parvenir à lier des données RDF avec d'autres. Comment assurez-vous l'articulation entre les métadonnées des Archives nationales et celle de l'IGN : méthode(s)/outil(s)/procédé(s) ?

Premièrement, nous utilisons RiC-O comme unique modèle pour la production des jeux RDF de métadonnées à partir des descriptions de l'ensemble des photos aériennes ou documents graphiques sélectionnés pour le projet ALEGORIA; quel que soit leur lieu de conservation (IGN, AN, Musée Nicéphore Niepce, etc.). Ce qui confère à ces jeux de données une grande homogénéité et permet, par ex. de construire facilement des recherches fédérées sur les documents (Record) décrits dans l'ensemble du corpus. D'autre part l'indexation des lieux ou objets représentés par les photos se fait en utilisant des référentiels de lieux (ceux des AN et ceux de l'IGN) que nous avons alignés. Idem pour les types de lieux, pour les techniques de production ou pour les types de documents.

29. Comment faites-vous pour aligner les concepts archivistiques avec les concepts des bibliothèques/des musées ou ceux de l'IGN par exemple ?

En l'occurrence pour ALEGORIA, hors référentiel des lieux, comme les référentiels étaient assez petits, l'alignement s'est fait manuellement. Pour les lieux on a utilisé diverses techniques pour interroger via SPARQL les jeux de données IGN et INSEE puis intégrer les résultats dans nos jeux de données. Plus généralement, nous utilisons beaucoup OpenRefine pour faire des alignements divers. Cela demande toujours une série d'interventions humaines, rien n'est totalement automatisable.

Table des matières

TABLE DES SIGLES ET ABREVIATIONS UTILISES 7

SOMMAIRE .....ERREUR ! SIGNET NON DEFINI.

INTRODUCTION GÉNÉRALE 1

LA NORMALISATION DE LA DESCRIPTION ARCHIVISTIQUE DE 1961 A 2008 5

1. Des origines bibliothéconomiques 7
1.1. Les règles de catalogage anglo-américaines de 1967 7
1.2. L'arrivée de l'automatisation, quels impacts sur la normalisation de la description bibliographique ? 9
1.3. *International Standard Bibliographic Description* (ISBD), un pas vers la réconciliation des pratiques américaines et britanniques ? 11
1.4. Les règles de catalogage anglo-américaines de 1978 13
2. Le rayonnement du monde anglo-saxon pour la normalisation de la description archivistique..... 16
2.1. Lutte contre l'idiosyncrasie : les efforts pionniers des Etats-Unis.....16
2.2. Rupture avec les origines bibliographiques : la vision britannique 20
2.3. L'initiative canadienne, un compromis entre l'approche américaine et britannique ? ..23
3. ICA, chronique de la description archivistique internationale 28
3.1. De la matière première au produit fini : normalisation de la description archivistique 28
1.1.1. De l'importance des principes.....28
1.1.2. ISAD/G, une norme couteau suisse.....30
1.1.3. Un format de description archivistique fait *par* des archivistes *pour* des archivistes ...31
3.2. La description contextuelle sur la voie de l'émancipation : vers une logique de réseau de relations 34
3.2.1. Vers l'indépendance de la description du producteur 34
3.2.2. Le 13ème Congrès international des archives et la création du Comité sur les normes de description : quels apports pour la discipline ? 35
3.2.3. Un format pour la description du producteur 37
3.2.4. Influence du 15ème Congrès de l'ICA sur la normalisation de la description archivistique 37

CONCLUSION DE L'ETAT DES CONNAISSANCES 39

BIBLIOGRAPHIE 42

1. LA NORMALISATION 42
2. LA DESCRIPTION BIBLIOGRAPHIQUE 43
3. L'EMPIRE BRITANNIQUE..... 45
4. LA DESCRIPTION ARCHIVISTIQUE 45
5. L'ENVIRONNEMENT NUMERIQUE DES ARCHIVES 49
6. L'UTILISATEUR FINAL DES ARCHIVES 51
7. LA TERMINOLOGIE 52
8. LE NUMERIQUE 53
9. LA LINGUISTIQUE..... 53
10. L'APPROCHE ENCYCLOPEDIQUE..... 53
11. LES NEUROSCIENCES ET LES SCIENCES COGNITIVES 54
12. LA CARTE HEURISTIQUE 56
13. LA SEMIOLOGIE GRAPHIQUE 56
14. LA *DATA VISUALISATION* 56

ETAT DES SOURCES 58

1. Sources écrites 58
1.1. Échelle internationales 58
1.2. Australie.....59
1.3. Canada.....59
1.4. États-Unis.....59
1.5. France.....60

1.6. Roumanie 61
1.7. Royaume-Uni 61
2. Sources par voie d'entretiens 62
3. Sources par voie de questionnaires 63

RECORDS IN CONTEXTS : CAP SUR LE WEB 3.0 64

1. RiC : origines, conception, implémentation 65
1.1. L'EGAD, une mosaïque de traditions archivistiques ? 66
1.1.1. Constitution et méthodes de travail de l'EGAD 66
1.1.2. Elaborer un modèle conceptuel : par où commencer ? 70
1.2. Redéfinir les fondamentaux de la description archivistique 72
1.3. « Qui contrôle les métadonnées, contrôle le Web » 77
1.3.1. Du modèle conceptuel à l'ontologie 77
1.3.2. Les principes de RiC-O 80
1.3.3. RiC-O : tout pour plaire ? 81
2. De la théorie à la pratique, première expérimentation de RiC à l'échelle nationale 84
2.1. Le PIAAF : expérimenter l'interconnexion inter-institutionnelle des données culturelles 84
2.1.1. Plusieurs acteurs, de multiples enjeux 84
2.1.2. Représenter en RDF : identifier ses besoins et s'approprier RiC-O 86
2.1.3. Les enjeux de la *data visualisation* 87
2.1.4. PIAAF : organisation et méthodologie 87
2.2. Un convertisseur et un composant d'exploration open-source pour les métadonnées archivistiques 89
2.2.1. RiC-O Converter : un logiciel pour la conversion des métadonnées archivistiques 89
2.2.2. Sparnatural : un composant d'exploration basé sur SPARQL 91
2.3. Des référentiels en mutation 93
2.3.1. Passage d'un référentiel des producteurs à un référentiel des agents 94
2.3.2. Les référentiels d'indexation liés au référentiel des agents : des référentiels de concepts 95
a) Le référentiel des fonctions 95
b) D'un référentiel des lieux à un référentiel d'entités géo-historiques 96
c) De nouveaux référentiels à venir ? 96
2.3.3. Favoriser l'intéropérabilité des données culturelles 97
3. Représentation sous forme de graphe : un moyen d'optimiser la compréhension contextuelle ? 100
3.1. Graphes et réseaux de neurones, un air de déjà-vu ? 100
3.2. *Data visualisation* : de l'importance d'une ergonomie attractive pour une information qui « saute aux yeux » 104
3.2.1. Systèmes mnésiques et processus attentionnels 104
3.2.2. Principes de Gestalt, sémiologie graphique : pour un guidage exogène de l'attention 106
3.3. Explorer le graphe : proposer des outils d'interaction 108

CONCLUSION DE L'ETUDE DE CAS 110

CONCLUSION GÉNÉRALE 113

ANNEXES 116

1. Entretiens 116
1.1. Entretien avec Madame Claire Sibille de Grimoüard 116
1.1.1. Guide d'entretien 116
1.1.2. Transcription chrono-thématique 120
1.2. Entretien avec Monsieur Thomas Francart 130
1.2.1. Guide d'entretien 130
1.2.2. Transcription chrono-thématique 135
1.3. Entretien avec Monsieur Jean-Séverin Lair 150
1.3.1. Guide d'entretien 150
1.3.2. Transcription chrono-thématique 154

1.4. Entretien avec Madame Florence Clavaud 163
1.4.1. Guide d'entretien 163
1.4.2. Transcription chrono-thématique 167
1.5. Entretien avec Monsieur Edouard Vasseur 181
1.5.1. Guide d'entretien 181
1.5.2. Transcription chrono-thématique 184
1.6. Entretien avec Monsieur David Chavalarias 192
1.6.1. Guide d'entretien 192
1.6.2. Transcription chrono-thématique 195
2. Questionnaires 201
2.1. Réponses de Monsieur Daniel Pitti 201
2.2. Réponses de Madame Florence Clavaud 218

TABLE DES MATIERES 230

RÉSUMÉ

Inspirée des exemples états-uniens, anglais et canadiens, la description archivistique internationale telle que nous la connaissons à travers ISAD/G, ISAAR/CPF, ISDF et ISDIAH a connu de nombreuses évolutions. Aujourd'hui, elle est sur le point d'être bouleversée avec l'arrivée de la nouvelle norme de l'ICA : *Records in Contexts* (RiC).

Par le biais d'entretiens auprès des personnes clés dans la création de RiC, nous avons analysé le contexte historique et technologique dans lequel il s'inscrit. Les enjeux du web sémantique et de la visualisation des données sous forme de graphe ont également fait l'objet d'entretiens auprès d'experts en représentation graphique et en sciences cognitives ainsi qu'auprès d'informaticiens.

L'étude de la perception visuelle, des systèmes mnésiques et des processus attentionnels nous a permis de questionner les effets de la représentation sous forme de graphe sur la compréhension contextuelle des fonds. Il s'agit alors de penser le travail de *data visualisation* de la future interface de visualisation, ainsi que les fonctionnalités que cette dernière pourrait proposer afin d'optimiser l'expérience de navigation de l'utilisateur final.

mots-clés : normalisation ; Conseil international des Archives ; web sémantique ; graphe ; *data visualisation* ; cognition ; perception visuelle

ABSTRACT

Inspired by American, English and Canadian examples, the international archival description as we know it through ISAD/G, ISAAR/CPF, ISDF and ISDIAH has experienced many developments. Today, it is about to be upset with the arrival of the new standard of the ICA : *Records in Contexts* (RiC).

Through interviews with key people in the creation of RiC, we analyzed the historical and technological context to which this new standard belongs. The semantic web and graph representation issues were also the subjects of discussions with graphic and cognitive science experts and computer scientists.

The study of the visual perception, the memory systems and attentional processes allowed us to question the effects of graph representation on the contextual understanding of the archival resources. Then, it is about thinking the data visualization work of the future visualization interface and the features it could offer to optimize the end-user's browsing experience.

keywords : standardization ; International Council on Archives ; semantic web ; graph ; data visualisation ; cognition ; visual perception

Présidence de l'université
40 rue de rennes – BP 73532
49035 Angers cedex
Tél. 02 41 96 23 23 I Fax 02 41 96 23 00

université angers

ENGAGEMENT DE NON PLAGIAT

Je, soussignée Estelle Cepparo
déclare être pleinement consciente que le plagiat de documents ou d'une partie d'un document publiée sur toutes formes de support, y compris l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.
En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce rapport ou mémoire.

signé par l'étudiant(e) le **28** / **05** / **2021**

Cet engagement de non plagiat doit être signé et joint à tous les rapports, dossiers, mémoires.

Présidence de l'université
40 rue de rennes – BP 73532
49035 Angers cedex

université angers