HAL
open science

La norme ISO 21127 au regard de l'offre logicielle d'A&A Partners : analyse prospective des besoins des musées et propositions d'évolution des solutions WebMuseo

Julien Possompès

To cite this version:

Julien Possompès. La norme ISO 21127 au regard de l'offre logicielle d'A&A Partners : analyse prospective des besoins des musées et propositions d'évolution des solutions WebMuseo. domain_shs.info.docu. 2017. mem_01723673

HAL Id: mem_01723673

https://memsic.ccsd.cnrs.fr/mem_01723673

Submitted on 5 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

Equipe pédagogique Stratégies

INTD

MEMOIRE pour obtenir le Titre enregistré au RNCP

"Chef de projet en ingénierie documentaire"

Niveau I

Présenté et soutenu par

Julien Possompès

le 07/12/2017

La norme ISO 21127 au regard de l'offre logicielle d'A&A Partners

Analyse prospective des besoins des musées et propositions d'évolution des solutions WebMuseo

Jury :
Monsieur Gérald Kembellec, Maitre de conférences, INTD-CNAM
Monsieur Emmanuel Pierru, Directeur technique, A&A Partners

Promotion 47 (2016-2017)

Paternité Pas d'Utilisation Commerciale - Pas de Modification

Remerciements

Merci à Emmanuel Pierru pour l'acuité de ses analyses et son aptitude à les transmettre au cours de nos nombreux échanges, à l'occasion de ce mémoire et plus généralement depuis mon arrivée dans l'entreprise.
Ma sincère reconnaissance va à mon employeur ainsi qu'à mes collègues qui m'ont permis de mener à bien ces deux années de formation en alternance.

Je remercie mon directeur de mémoire Gérald Kembellec pour ses remarques et conseils de lecture inspirants.

Merci à Nadia Raïs, Xavier Aimé et Philippe Bourdenet dont l'enseignement m'a permis de découvrir le web de données et les ontologies.

Je remercie Safia Ziour dont les références bibliographiques m'ont mis le pied à l'étrier au début de l'été et Marianne Bordreau pour sa relecture finale et son soutien indéfectible.

Merci à ma famille et à mes proches pour leur patience et leurs encouragements.

Merci aux équipes de l'INTD pour ces deux années de formation dont je garderai un excellent souvenir.

Merci aux camarades des deux promotions pour leur bienveillance et leur humour ; tout particulièrement à mes co-équipières Catherine Ventenat et Mickaëlle Fils Marie-Luce, aux philosophes de la machine à café Agnès Cunegatti, Camille Bault, Firas Riche, Vincent Rappeneau, Karine Ueda, Violaine Serra-Lacombe, François Haas, et à mon AVS Elisabeth Plus.

Résumé / Abstract

Ce mémoire propose une étude du modèle de référence conceptuel élaboré par le CIDOC (Comité international pour la documentation) et une analyse de son intérêt pour société A&A Partners, développeur de solutions web de gestion et de valorisation du patrimoine culturel. L'étude explique le fonctionnement théorique de l'ontologie CIDOC CRM, présente les retours d'expérience de son utilisation par le Smithsonian American Art Museum et le British Museum puis se concentre sur la question de la modélisation des liens entre objet patrimonial et connaissance. L'analyse met en évidence la convergence entre les enjeux métiers de la société A&A Partners et certaines problématiques du web sémantique culturel (gestion des référentiels, mises à jour des données, intégrité de l'information). Le positionnement d'A&A Partners vis à vis du web sémantique culturel est celui d'un fournisseur de données structurées exploitables au sein d'entrepôts de triplet RDF générés et exploités via des ontologies telles que le CIDOC CRM.

Descripteurs : LODLAM (Linked Open Data in Librairies, Achives & Museums), ontologie, modèle conceptuel, interopérabilité, web de données, musée, gestion de collections patrimoniales, ingénierie des connaissances.

This master thesis proposes a study of the conceptual reference model developed by the CIDOC (International Committee for Documentation) and an analysis of its interest for A&A Partners, developer of web solutions for management and valorization of cultural heritage. The study explains the theoretical functioning of the CIDOC CRM ontology, presents the feedback from its use by the Smithsonian American Art Museum and the British Museum and then focuses on the question of modeling the links between heritage objects and knowledge.The analysis highlights the convergence between the business challenges of A&A Partners and certain issues of the semantic web (management of repositories, data updates, integrity of information). A&A Partners' positioning for the cultural semantic web is that of a provider of data structured and workable within RDF triplet warehouses generated and exploited via ontologies such as CIDOC CRM.

Keywords : LODLAM (Linked Open Data in Librairies, Achives & Museums), ontology, conceptual models, semantic Web, interoperability, museum, heritage collection management, knowledge management.

Sommaire

Remerciements **2**
Résumé / Abstract **3**
Sommaire **5**
Table des illustrations **7**
Introduction **9**
Première partie : Etude du modèle conceptuel ISO21127:2014 ..... **11**
1 Qu'est-ce que le CIDOC CRM ? **11**
1.1 Un consensus international et transdisciplinaire 11
1.1.1 Expertise métier et documentaire 11
1.1.2 Genèse du modèle conceptuel 11
1.2 Un modèle pour l'interopérabilité des inventaires patrimoniaux 13
1.2.1 Relier les descriptions des biens culturels 13
1.2.2 Un vocabulaire socle du LODLAM 16
1.2.3 Un modèle extensible (FRBRoo, PRESSoo, etc.) 16
1.2.4 Un schéma compatible 17
1.3 Les concepts fondamentaux du modèle 19
1.3.1 Entités et propriétés 20
1.3.2 Endurants et Perdurants 20
1.3.3 Appellation et Types 21
1.3.4 Acteurs et Choses 21
1.3.5 Choses matérielles et Objets conceptuels 22
1.3.6 Evénements 22
2 Principales applications **23**
2.1 Le Smithsonian American Art Museum 23
2.2 Le British Museum 25
3 Réunir Patrimoine et Récit historique **26**
3.1 La source, fondement de l'information historique 26
3.1.1 Modéliser et exploiter les ressources historiques 26
3.1.2 Alignement avec CIDOC CRM : conserver la notion de source 27
3.2 L'objet "conteneur de connaissance" 29
3.3 L'archive support d'un discours 31
Deuxième partie : Positionnement et évolutions possibles pour A&A Partners **35**

4 L'entreprise et ses solutions 35
4.1 Le contexte institutionnel 35
4.1.1 Les préconisations du SMF (Service des musées de France) du ministère de la Culture 35
4.1.2 Les portails culturels 36
4.1.3 Le programme HADOC (HArmonisation pour la production des DOnnées Culturelles) 37
4.2 La plateforme WebMuseo 39
4.2.1 WebMuseo Gestion 40
4.2.2 WebMuseo Publication 42
4.3 L'accompagnement des utilisateurs 43
4.3.1 Formation à l'utilisation 43
4.3.2 Migration de données 43
4.3.3 Développements spécifiques 46
5 Les échanges de données un enjeu métier 47
5.1 Exports unilatéraux : publication, archives 47
5.1.1 Mise en ligne des collections muséales 47
5.1.2 Export enrichi pour archivage (Ressources EFEO au CINES) 48
5.2 Import de données et mises à jour : l'exemple d'e-ReColNat 48
6 Compatibilité et propositions d'évolutions 55
6.1 Un existant favorable à l'exposition sémantique 55
6.1.1 Eléments favorisant la qualité des données 55
6.1.2 Compatibilité native du modèle WMG avec le CIDOC CRM 56
6.1.3 Notices d'œuvres et représentation des connaissances 58
6.2 Principales notions à modéliser 59
6.2.1 Précisions sur la modélisation 59
6.2.2 Caractéristiques générales des objets 59
6.2.3 Etapes de la vie d'un objet : de sa conception à sa conservation patrimoniale 63
6.2.4 Documentation 70
6.3 Limites de la modélisation en CIDOC CRM 72
6.4 Enseignements vis-à-vis de la gestion du cycle de vie 72
6.5 Evolution vers des fonctionnalités d'éditorialisation 72
Conclusion 73
Bibliographie 76

Table des illustrations

Figure 1 : Le « layer cake » du Web Sémantique 15
Figure 2 : Graphe des liens sortants du LAWD vers d'autres vocabulaires du LOD (Capture d'écran du site http://lov.okfn.org et annotations JP) 18
Figure 3 : Schéma général du CIDOC-CRM [15, OLDMAN, CRM Labs 2014].....19
Figure 4 : Structure générale des entités du CIDOC-CRM [22, TCHIENEHOM 2017) 20
Figure 5 : L'ontologie du SAAM [21, Szekely et al. 2013] 24
Figure 6 : Modélisation CRM d'une unité de connaissance "meurtre de Brutus" issue d'une source documentaire - d'après Atelier-rencontre avec Martin Doerr, 15-17 novembre 2016 28
Figure 7 : Mise en évidence des entités pivots pour la conversion en CIDOC CRM (entourées en rouge) d'après Schéma simplifié de l'ontologie SyMoGIH. (LARHRA 2016) 29
Figure 8 Processus de traitement des médias par le dispositif Campus-AAR 32
Figure 9 Structure sémiotique d'un objet ou corpus média [27, STOCKINGER, Peter, LALANDE, Steffen et BELOUED, Abdelkrim, 2015] 33
Figure 10 : Schéma sémantique simplifié du modèle HADOC (Katell Briatte) 38
Figure 11 : Processus de migration des données (EP, A&A Partners) 44
Figure 12 : Connecteur e-ReColNat label ARGENT : export des notices (JP) 49
Figure 13 : Connecteur e-ReColNat label OR : export et ré-import des notices (JP) 49
Figure 14 : Champs d'applications WebMuseo Gestion "Museum", e-ReColNat et Darwin Core (EP, A&A Partners) 51
Figure 15 : Arbre taxonomique de l'espèce Canis lupus, Linnaeus, 1758 dans WebMuseo (JP).....52
Figure 16 : Arbre taxonomique en temps réel (EP, A&A Partners) 53
Figure 17 : Arbre taxonomique historisé (EP, A&A Partners) 54
Figure 18 : Verrouillage de la détermination (EP, A&A Partners).....55
Figure 19 : Modélisation des caractéristiques générales des objets (JP) 62
Figure 20 : Modélisation de la création de l'œuvre (JP).....64
Figure 21 : Modélisation possible d'un laps de temps (JP).....66
Figure 22 : Modélisation des acquisitions (JP) 67
Figure 23 : Modélisation des transferts de garde (JP) 68
Figure 24 : Modélisation de la localisation (JP) 69

Figure 25 : Modélisation des constats d'états (JP).....70
Figure 26 : Modélisation de la documentation associée (JP).....71

Introduction

Société d'édition de logiciel de gestion, publication et syndication web à destination des musées et des gestionnaires de collections muséales, A&A Partners souhaite identifier les principaux enjeux de la norme ISO 21127:2014 présentant la version officielle du modèle de référence conceptuel du CIDOC (Comité International pour la DOCumentation).

A quoi sert le CIDOC-CRM ? Quel est l'écart entre ce modèle et les solutions webmuseo d'A&A Partners ? Ces dernières sont-elles facilement compatibles et selon quels principes ? La norme ISO 21127 contiendrait-elle la formalisation de concepts qui permettrait des améliorations fonctionnelles ? Dans quelle mesure le marché (c'est-à-dire les gestionnaires de collections) est-il en attente de solutions compatibles CIDOC-CRM ?

Afin de répondre à ces questions, nous commencerons par dresser une fiche d'identification de la norme : genèse au sein du CIDOC, caractère novateur et ouvert dans le contexte du web de données liées, présentation des concepts fondamentaux du modèle.
Nous observerons ensuite dans quelle mesure ce modèle a été utilisé pour l'exposition de données muséales par le SAAM (Smithsonian American Art Museum) et par le British Museum.
Conscient que l'intérêt du web sémantique réside dans l'enrichissement mutuel des gisements d'informations, nous nous intéresserons à trois expérimentations de modélisation des liens entre objet patrimonial et connaissance : les assertions historiques pour le SyMoGIH (Système modulaire de gestion de l'information historique), les récits en lien(s) avec le patrimoine (Nantes 1900), le discours porté par une archive dans le cadre de MEDIHAL. Face à ces bases de connaissance très spécialisées, le CIDOC CRM peut-il apporter une valeur ajoutée ?

Nous présenterons ensuite l'entreprise A&A Partners, son contexte et ses savoir-faire pour évaluer dans quelle mesure le CIDOC CRM lui permettrait de participer au web de données liées et d'améliorer ses modèles de données. Une attention

particulière sera portée à la problématique de l'échange de données, cruciale pour ancrer la réflexion dans la réalité opérationnelle.

Première partie : Etude du modèle conceptuel ISO21127:2014

1 Qu'est-ce que le CIDOC CRM ?

1.1 Un consensus international et transdisciplinaire

1.1.1 Expertise métier et documentaire

L'ICOM (International Council of Museums) est une organisation non-gouvernementale émanant de l'UNESCO (United Nations Educational, Scientific and Cultural Organization). Créé en 1946, il s'agit de la seule organisation de musées et de professionnels de musées à l'échelon mondial. Elle a pour mission de "*promouvoir et protéger le patrimoine culturel et naturel, présent et futur, tangible et intangible*[1]". Avec 26 000 membres répartis dans 151 pays, l'ICOM est un réseau unique de professionnels de musées. Dans l'ensemble des préconisations de l'ICOM, la documentation des collections (identité, gestion physique et administrative) apparaît comme une exigence essentielle de la bonne tenue d'un musée.

Le CIDOC (Comité International pour la DOCumentation) fondé en 1950 (majoritairement par des francophones), est aujourd'hui l'un des 30 comités internationaux de l'ICOM. Il rassemble 550 membres de 70 pays[2] et leur permet d'échanger au cours d'événements réguliers ou ponctuels, de formaliser des consensus et de diffuser définitions et recommandations en matière de documentation et de gestion de collections muséales. La force du CIDOC est d'allier expertise documentaire et expertise métier dans une démarche d'échange et de construction du consensus.

1.1.2 Genèse du modèle conceptuel

Le CIDOC-CRM (Conceptual Model Reference) est l'aboutissement d'un travail de plus de 10 ans du DSWG (Documentation Standards Working Group). Son objectif initial était de créer un modèle général de données pour les musées, mettant l'accent sur l'échange de données. Ainsi, jusqu'en 1994 ce groupe de travail élabore

[1] http://www.icom-musees.fr/index.php/page/index/Quest-ce-que-lICOM (consulté le 31/07/2017)
[2] http://open-world.ch/CIDOC/SummerSchool/Remote/101%20Introduction%20to%20documentation/

un modèle de données relationnel. A partir 1996, le DSWG s'engage dans une approche orientée objets afin de gagner en expressivité et en extensibilité.

En effet, le modèle de 430 entités-relations était devenu trop lourd à utiliser et à maintenir[3]. La nouvelle approche devait "*mettre sur pied un modèle aussi simple et compact que possible, mais tout de même suffisamment riche pour rendre compte de toutes spécificités des différents types de collections muséales*[4]". Le CRM a d'emblée été conçu selon une méthode pragmatique et empirique tirant ses principes de l'analyse de l'existant. Ce modèle n'est pas applicable en l'état aux données des musées, qui doivent passer par une opération préalable de mise en concordance manuelle des rubriques source avec les champs du système cible, dont une part de part de plus en plus importante devrait être automatisée au fur et à mesure des projets[5].

La première édition complète du modèle de référence conceptuel CIDOC (appelé CIDOC-CRM) est publiée en 1999. Afin de l'imposer en tant que standard international et validé, le CIDOC le propose à l'ISO (International Organization for Standardization) dès 1999.

Depuis 2000, la rédaction et la maintenance de cette norme sont portées par le CIDOC-CRM-SIG (Specific Interest Group).

> *"Tout organisme intéressé à prendre part à cette réflexion peut demander à y nommer un représentant. On y rencontre plutôt des informaticiens et des spécialistes du traitement de l'information que des conservateurs de musée ou de bibliothèque - mais les profils des participants sont assez variés"* (16 LE BOEUF, 2013)

Notons pour l'implication française l'AFNOR (Association Française de NORmalisation), le C2RMF (Centre de recherche et de restauration des musées de France) et l'IRCAM (institut de recherche et coordination acoustique/musique), la BnF (Bibliothèque nationale de France).

La formalisation de la sémantique inhérente aux données muséales s'est trouvée confortée par les principes du web sémantique formulés notamment par Tim Berners-Lee à partir de 2001[6].

Le CIDOC-CRM est accepté comme projet de travail par l'ISO en 2000 (ISO/TC46/SC4) et publié officiellement comme norme en 2006 sous la référence

---

[3] https://www.slideshare.net/patrickleboeuf/prsentation-du-cidoc-crm-20151109/3 (consulté le 31/10/2017)

[4] LE BOEUF 2013

[5] [13] OLDMAN, CRM Labs 2014, p.5

[6] Berners-Lee et al., "The Semantic web" in Scientific American, 2001

ISO 21127:2006. Il est en constante évolution et sa version ISO a déjà été mise à jour une première fois en décembre 2014 ISO 21127:2014. Cette dernière est basée sur la version 5.0.4[7] du CIDOC CRM. La dernière version de travail mise à disposition gratuitement en anglais depuis le site du CIDOC-CRM est la 6.2[8] et date de mai 2015.

En France c'est en 2011 qu'un groupe de travail de l'AFNOR a réuni des représentants du ministère de la Culture et de la Communication, du Centre national d'art et de culture Georges Pompidou et de la BnF pour participer à la révision de la norme (effective en 2014).

La norme ISO est traduite en français sous la direction de Patrick Le Boeuf et mise en vente par l'AFNOR. Le modèle est régulièrement mis à jour en fonction de ses extensions (FRFRoo, PRESSoo, CRMgeo, etc.) et des projets le mettent concrètement en oeuvre notamment les musées de l'AAC (American Art Collaborative) ou le British Museum.

> *"The CRM Special Interest Group (SIG) now meet on a regular basis to maintain the standard, resolve issues and incorporate new practice into the model. It is an international and democratic committee open to new proposals from the user community."*
> [17, OLDMAN, CRM Labs 2014 p.2]

1.2 Un modèle pour l'interopérabilité des inventaires patrimoniaux

1.2.1 Relier les descriptions des biens culturels

Le CIDOC-CRM est une ontologie, ou modèle conceptuel, qui formalise et explicite la description et la gestion des biens patrimoniaux et culturels.

> *"The CIDOC CRM ontology provides a real world, empirically based representation aimed at harmonizing heterogeneous data. However, the CIDOC CRM method of harmonisation retains the individual nature of the data proving a semantic framework or context that supports the full variability and richness of the information and brings to life the concealed and implicit relationships that exist between things".*
> [17, OLDMAN, CRM Labs 2014 p.1]

Chaque concept est précisément décrit dans une note d'application (scope note). Toutefois, contrairement à d'autres ontologies de domaine, le CIDOC-CRM reste générique dans ses entités et très souple dans ses capacités de liaisons (peu

[7] http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf
[8] http://www.cidoc-crm.org/Version/version-6.2

d'entités obligatoires[9]). Dans sa version 6.2 de mai 2015 le modèle définit seulement 94 classes et 168 propriétés. L'ensemble des combinatoires possibles permet d'interpréter et de formaliser dans un langage unique des sources hétérogènes à plus d'un titre : format des données, type de collections, métiers pratiqués, échelle de catalogage.

> *« En technologie de l'information, une ontologie est l'ensemble des définitions logiques des concepts d'un domaine et des relations qu'ils entretiennent. »* [Menon 2016]

L'objectif est d'offrir un modèle d'interopérabilité avec l'ensemble des ressources patrimoniales et culturelles (musées, archives, bibliothèques, recherche). En effet plutôt que de construire des correspondances entre différentes bases hétérogènes, le CIDOC CRM propose un modèle commun restituant les points de vue et la granularité de chaque institution culturelle. Ce n'est pas un simple format de métadonnées prescrivant une liste de champs (comme le Dublin Core) mais un cadre sémantique qui décrit des entités plus générales (y compris des événements) et les relations entre eux. Cependant, s'il permet un accès homogène, il n'homogénéise pas pour autant les données par rapport au type de contenu représenté.

> *"Rather than describing a limited number of common fields, as in many digital representations, the CRM describes objects more fully (reflecting full institutional documentation) including the semantic meaning of the data".*
> [17, OLDMAN, CRM Labs 2014, p.13]

L'ontologie CIDOC CRM permet de faire correspondre des instances de personnes, de lieux, de choses, d'événements et de périodes en utilisant l'information et le contexte autour de ces entités et d'aller plus loin qu'un d'un simple rapprochement de chaînes de caractères.

Le CRM répond ainsi à plusieurs préoccupations :

- Réduire le silence documentaire en explicitant et formalisant la connaissance implicite contenue dans les sources.
- Encapsuler avec la donnée les informations permettant de la comprendre, et ce indépendamment d'un logiciel ou d'une machine. Les données structurées selon le CIDOC CRM restent toujours lisibles par l'homme.
- Permettre à chaque institution culturelle (musée, bibliothèque, service d'archives) de documenter ses collections selon ses priorités internes

[9] Le CIDOC CRM est donné comme exemple d'ontologie formelle par le bureau italien du W3C : http://www.w3c.it/talks/2005/openCulture/slide18-0.html

(scientifiques, métier, légales ou stratégiques) tout en participant de façon efficace au Linked Open Data culturel.

Ce qui apparaissait comme une faiblesse inhérente à l'hétérogénéité des données culturelles est vue comme une richesse calculable et exploitable à condition d'être formalisée.

Il est alors concevable de :

- générer des liens "intelligents" entre différents gisements d'informations complémentaires,
- détecter de nouveaux faits exploitables par les chercheurs,
- rendre accessible des connaissances très spécifiques.

Toutefois, le CIDOC CRM reste un objet théorique qui doit être implémenté en OWL (Ontology Web Language) ou en RDFS (Schéma RDF) pour être une ontologie exploitable dans le cadre du web de données.

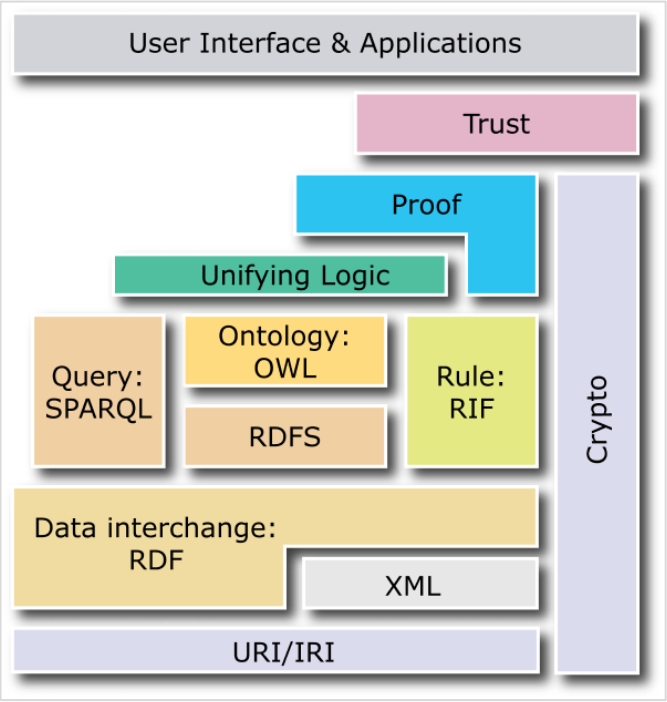

Figure 1 : Le « layer cake » du Web Sémantique[10]

Le web sémantique s'appuie sur les fondamentaux du langage : les mots (URIs et les valeurs littérales encodées) sont assemblés selon une syntaxe (triplet RDF : sujet, prédicat, objet), pour produire un texte (jeu de données RDF), dans un

[10] https://fr.wikipedia.org/wiki/Semantic_Web_Stack (consulté le 6/12/2017)

système d'écriture (XML). Le langage SPARQL sert à interroger et à compiler les jeux de données RDF[11].

L'ontologie joue le rôle de dictionnaire pour encoder/décoder finement le sens des jeux de données RDF. Elle apporte une valeur ajoutée dans le traitement sémantique mais n'est pas le seul canal de production et d'exploitation des *triplestore*.

Ainsi, R2RML (RDB to RDF Mapping Language) propose un modèle pour convertir des bases de données relationnelles en RDF. La récente recommandation SHACL (SHapes Constraint Language) du W3C offre un cadre pour spécifier des règles de conformité des graphes RDF[12].

Les langages SWRL (Semantic Web Rule Language) et SPIN (SPARQL Inferencing Notation) ont été conçus pour générer de l'inférence, c'est-à-dire de nouveaux liens entre les triplets exposés.

1.2.2 Un vocabulaire socle du LODLAM

L'un des objectifs revendiqués par les concepteurs du CIDOC CRM est de lier les biens culturels sans modifier leurs champs métiers respectifs. Les collections muséales sont décrites en insistant sur l'unicité de chaque objet ; les archives sont appréhendées par fonds et en attachant une grande importance à leur contexte de production ; les ouvrages de bibliothèques sont catalogués principalement en fonction de leur contenu informationnel. Chacune de ces trois disciplines a ses objectifs métiers propres et reste légitime dans ses méthodes de description. Néanmoins ces identités métiers fortes génèrent des silos qui freinent mécaniquement la capacité à rechercher et exploiter l'information culturelle. L'idée principale du LODLAM (Linked Open Data for Libraries Archives and Museums) est d'utiliser les standards du web sémantique pour lier en aval les différentes descriptions. Pensé dès le départ dans cette optique, le CIDOC CRM propose une modélisation sémantique qui peut être augmentée soit par extension et/ou par adjonction d'autres standards de description.

1.2.3 Un modèle extensible (FRBRoo, PRESSoo, etc.)

L'expressivité du CIDOC CRM a été assez vite reconnue par le monde des bibliothèques. L'IFLA (**I**nternational **F**ederation of **L**ibrary **A**ssociations and

---

[11] Voir [7, Menon, 2016]

[12] Voir la présentation de ce standard par Thomas Francart : http://blog.sparna.fr/2017/01/02/shacl-rdf-shapes-constraint-language-enfin-la-possibilite-de-valider-des-donnees-rdf/ (consulté le 6/12/2017)

Institutions) a impulsé dès 2003 une réflexion visant à transposer le modèle descriptif FRBR (Functional Requirement for Bibliographic Records) dans la grammaire orientée objet du CIDOC CRM. FRBRoo (FRBR orienté objet) est une extension du CIDOC CRM. Chacun des deux modèles garde "*son autonomie dans sa communauté d'origine*" tout en étant fortement interopérable [16, LE BOEUF 2013].

Au cours des échanges entre professionnels des bibliothèques et ceux des musées trois cas d'alignements se sont présentés :

- équivalence complète et introduction de la classe CIDOC CRM telle quelle
- entité FRBR plus complète et déclaration d'une sous-classe spécifique à FRBRoo
- entité FRBR plus générique ou hors champ et modification du CIDOC CRM [19, LE BOEUF, 2009]

Plusieurs autres modèles constituant des enrichissements à destination d'autres disciplines sont officiellement référencés sur le site du CIDOC-CRM[13] :

- PRESSoo, une extension de FRBRoo pour la description des périodiques
- CRMarchaeo pour documenter les fouilles archéologiques
- BRMba pour décrire les vestiges bâtis
- CRMsci pour décrire l'information scientifique

Dans cette même page sont référencés des compléments plus opérationnels et transdisciplinaires :

- CRMinf pour générer des raisonnements des inférences
- CRMgeo pour exploiter la ressource de formalisation topographique fournie par Open Geospatial Consortium (OGC).
- CRMdig pour exploiter les métadonnées des ressources numériques illustrant les objets décrits selon le CIDOC CRM

1.2.4 Un schéma compatible

L'encodage du CIDOC CRM est disponible selon deux formats sémantiques : RDFS par le SIG (Special Interest Group) et OWL par l'université allemande d'Erlangen[14].

RDF Schema (ou RDFS) permet de « *déclarer des propriétés, de les organiser en hiérarchie et de spécifier leur signature en termes de domaine et de portée* » [8, Menon, 2016].

---

[13] http://www.cidoc-crm.org/collaborations (consulté le 09/10/2017)
[14] http://lov.okfn.org/dataset/lov/vocabs?q=crm (consulté le 10/10/2017)

OWL (Web ontology Language) est le format le plus complet pour formaliser une ontologie. *« Il prend le relai là où l'on a besoin de rendre compte plus en détail de la structuration conceptuelle d'un domaine d'application.* » [8, Menon, 2016].

De plus, le jeu de métadonnées CRM Core, fonctionnel tout en palliant les lacunes et imprécisions du Dublin Core simple est proposé depuis 2005. Il permet de rendre les ressources exploitables selon la sémantique du CIDOC CRM.

L'objectif du modèle LAWD (Linking Ancient World Data Ontology) est d'articuler les données culturelles utilisées et publiées non structurées ou selon d'anciennes pratiques avec les standards internationaux comme Dublin Core, la collaboration Open Annotation et CIDOC-CRM.

Disponible depuis la plateforme LOV (Linked Open Vocabularies), LAWD y est représenté comme un pivot entre les différents standards de description des ressources disponibles sur le LOD (Linked Open Data). C'est la transposition ecrm mise à disposition en OWL par l'Université d'Erlangen qui est présentée comme référence.

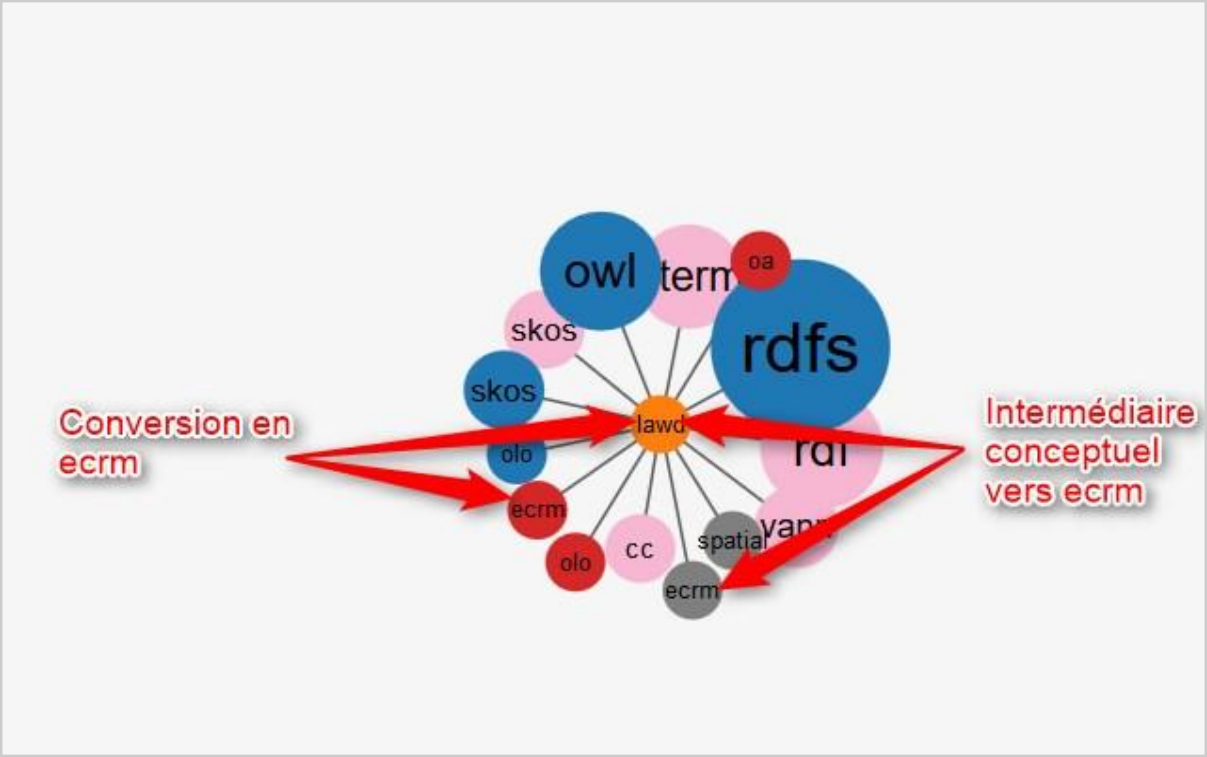

Figure 2 : Graphe des liens sortants du LAWD vers d'autres vocabulaires du LOD[15] (Capture d'écran du site http://lov.okfn.org et annotations JP)

[15] http://lov.okfn.org/dataset/lov/vocabs/lawd (consulté le 10/10/2017)

Deux types liens sortants vers ecrm (version OWL de l'Université d'Erlangen) sont présentés : en rouge "Imports", en gris "Has Equivalences with". Nous interprétons qu'il est ainsi possible d'exporter des instances formalisées en LAWD vers un entrepôt structuré en eCRM et d'aligner les concepts définis par LAWD sur ceux formalisés par eCRM.

1.3 Les concepts fondamentaux du modèle

Le CIDOC-CRM permet de décrire :

- des objets (identifiant, type, titre, matériau, dimension, note),
- leur historique (origine, découverte, localisation, événement, propriétaires successifs, transfert de garde, état de conservation, mesures),
- ainsi que les relations qui existent entre objets ou partie d'objets (bibliographie, composition, similarité, autre représentation, inscription)[16].

La principale caractéristique du CIDOC-CRM est son approche centrée sur l'événement : "*Tout ce qui est décrit dans un inventaire muséographique est le produit d'une suite d'événements*". [16, LE BOEUF 2013].

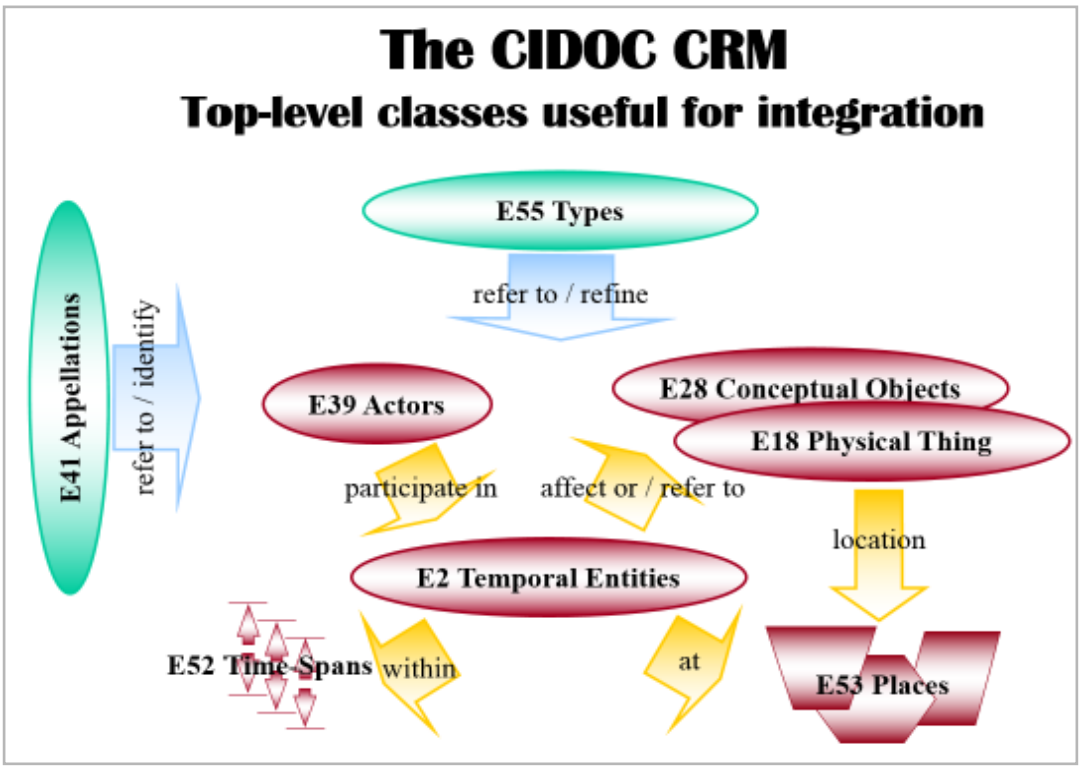

Figure 3 : Schéma général du CIDOC-CRM [17, OLDMAN, CRM Labs 2014]

[16] (

Le schéma général publié par le CIDOC montre bien que chaque entité (Acteur, Objets conceptuel, Chose matérielle, Evénement, Lieu) peut être nommée (E41_Appellations) et catégorisée (E55_Types).
Les acteurs participent aux événements qui ont lieu chacun dans un laps de temps, à un endroit. Ces événements se manifestent comme ayant un effet sur les objets conceptuels et/ou les choses matérielles.

1.3.1 Entités et propriétés

Une entité "*est une chose, un objet, ou une réalité, voire une substance au sens philosophique, toujours de nature et de propriétés indéterminées, et apparemment dotées d'une forme d'individualité, d'identité ou d'unité*[17]. [22, TCHIENEHOM 2017]
Une propriété est une caractéristique d'un objet. Un objet rouge est dit posséder la propriété de « rougeur ». "*La propriété peut être considérée comme une forme de plein droit de l'objet, lequel est en mesure de posséder d'autres propriétés*[18]".
Dans une ontologie telle que le CIDOC-CRM, les entités sont liables entre elles selon un nombre défini de propriétés. Chaque entité admet un nombre limité de propriétés. Chaque propriété est applicable à une liste précise d'entités.

1.3.2 Endurants et Perdurants

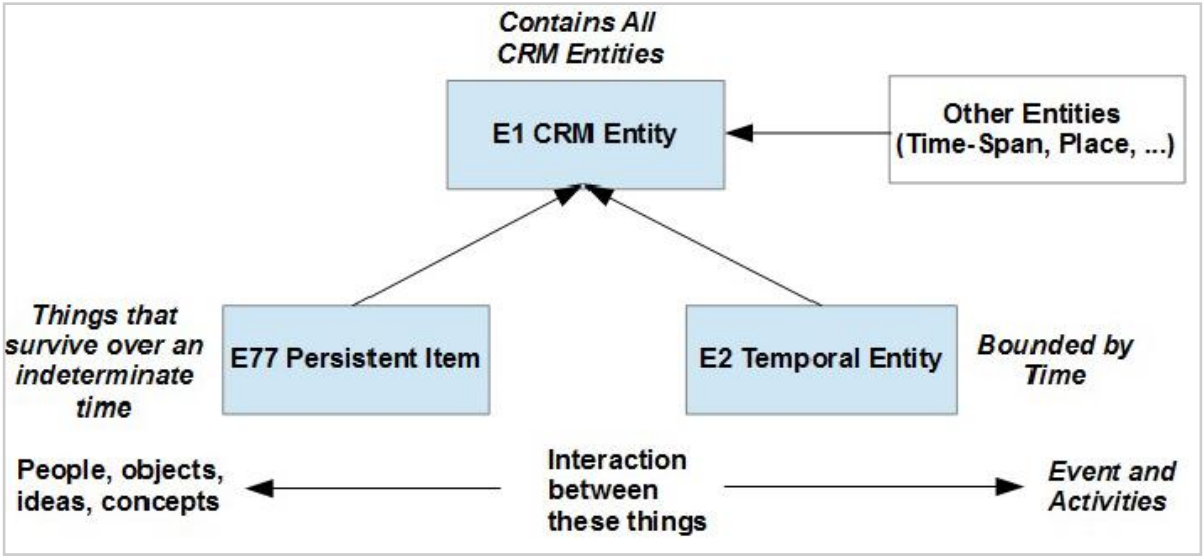

Figure 4 : Structure générale des entités du CIDOC-CRM [22, TCHIENEHOM 2017)

[17] fr.wikipedia.org/wiki/Entité (consulté le 2/08/2017)
[18] fr.wikipedia.org/wiki/Propriété_(philosophie) (consulté le 2/08/2017)

Comme d'autres ontologies de domaine comme par exemple DOLCE[19], le CIDOC-CRM établit une distinction fondamentale entre ce qui a une durée indéterminée (les endurants héritant de la super-classe E77 Persistent Item) et ce qui s'inscrit dans une durée limitée (les perdurants héritant de la super-classe E2 Temporal Entity).

Les personnes (E39_Actors), objets (E18_Physical_Thing), concepts (E28_Conceptual_Object) sont des endurants. Les événements (E5_Event), périodes (E4_Period) ou activités (E7_Activity) sont des perdurants. Le CRM est un modèle sémantique orienté événements dans la mesure où il décrit les lois d'interactions entre ces deux branches d'entités. Les relations sont également possibles au sein de chaque branche.

> *"For example, a person's identity endures regardless of his/her death. Death is a temporal concept just as events that occurred during a person's physical life are also examples of things bounded by some period of time. Leonardo da Vinci no longer exists physically but his identity survives."*
> [17, OLDMAN, CRM Labs 2014, p.13]

1.3.3 Appellation et Types

Un acteur comme n'importe quelle entité, ne s'identifie pas à son nom. Il existe en soi et se voit rattacher une appellation (E82_Actor_Appellation). De la même façon, l'existence d'une chose n'est pas confondue avec son nom. Elle peut avoir une ou plusieurs appellations (E41_Appellation), qui en fonction de sa nature et de ses caractéristiques seront un titre (E35_Title) ou une appellation d'objet conceptuel (E75_Conceptual_Object_Appellation).

Ainsi le fait que l'appellation soit elle-même une entité permet de faire co-exister et de lier plusieurs variantes. Par exemple, les appellations d'acteurs "Raphaël", "Raffaello Sanzio", "Raffaello Santi", "Raffaello da Urbino", "Raffaello Sanzio da Urbino" font bien référence au peintre de la *Fornarina*.

> *E21_Person **P131 est identifié** par E82 Appellation d'acteur {instance = "Raffaello Sanzio"}*

De même qu'elle peut être nommée à l'infini, chaque entité peut être indexée selon un type (E55_Types).

1.3.4 Acteurs et Choses

Le modèle distingue deux catégories d'endurants : les acteurs (E39_Actor) et les choses (E70_Thing).

---

[19] http://www.loa.istc.cnr.it/old/DOLCE.html (consulté le 31/10/2017)

L'acteur peut être une personne physique (E21_Person) ou un Groupe (E74_Group), une personne morale (E40_Legal-Body) par exemple.

1.3.5 Choses matérielles et Objets conceptuels

Le parti pris fondamental du CIDOC-CRM est d'opérer une distinction entre Objet conceptuel (E28_Conceptual_Object) et Chose matérielle (E18_Physical_Thing). Une oeuvre sera donc toujours à la fois un objet conceptuel créé et une chose matérielle produite.

1.3.6 Evénements

La description de l'état matériel d'un objet (E3_Condition_State) se conçoit nécessairement sur une période donnée. Les différents changements d'états peuvent ainsi être retracés au cours du temps.

La Période (E4_Period) est définie comme "*un ensemble cohérent de phénomènes ou de manifestations culturelles limitées dans l'espace et dans le temps"* [15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel, 2014]

L'événement est subsumé de la période (et hérite donc de toutes ses propriétés). Sa caractéristique selon le CIDOC-CRM est d'entraîner *« un changement d'état : avant que le peintre ne peigne son tableau, ce tableau était dans l'état "non existant" ; après cet événement, le tableau est dans l'état "existant" ».* [16, LE BOEUF, Patrick, 2013, p.8].

Les événements sont classés en trois catégories non exclusives : Début d'existence (E63_Beginning_of_Existence) ; Fin d'existence (E64_End_of-Existence) ; Activité (E7_Activity).

La distinction entre Objet conceptuel et Chose matérielle est prolongée par celle de la Création et de la Production. En effet, la Création (E65_Creation) est l'activité par laquelle un agent humain donne naissance à un nouvel objet conceptuel (E28_Conceptual_Object), tandis que la Production (E12_Production) permet le début d'existence d'une Chose matérielle fabriquée (E24_Physical Man-Made Thing). *« Création et Production sont donc deux cas particuliers, ayant un caractère d'intentionnalité, de Début d'existence* ». [14, LE BOEUF, Patrick, 2013, p.8].

2 Principales applications

2.1 Le Smithsonian American Art Museum

Comme c'est le cas pour la plupart des musées et *a fortiori* pour les institutions telles que le Smithonian American Art Museum (SAAM), les conservateurs gèrent (identifient, décrivent, notent les déplacements et événements) leurs collections via un logiciel dédié. Dans le cas du SAAM, c'est l'application propriétaire TMS (The Museum System) éditée par la société américaine Gallery Systems. Il s'agit d'un système de gestion de base de données relationnelle d'une centaine de tables. L'information y est structurée, riche et en partie confidentielle (avec des détails sur les aspects financiers par exemple). Afin de s'assurer de la publiabilité des données, la conversion en triplet RDF s'est limitée aux données déjà diffusées sur le site web du SAAM soit 41 267 oeuvres pour 8 261 artistes.

> *"In order to avoid issues about private data, we only use the tables that museum uses to populate their Web site. All the information in these eight tables already appears on the museum Web site, so the museum is amenable to publishing it as linked data. The structure and format of these data tailored to the needs of the Web site and some fields need to be decoded to produce appropriate RDF."*
> [21, SZEKELY, et al., 2013, p.595]

Pour modéliser sémantiquement ces données, le SAAM s'est en premier lieu tourné vers le "méta-modèle" EDM constitué par le projet Européana[20] en ré-utilisant des standards couramment utilisés :

- SKOS[21] (Simple Knowledge Organization System) pour la classification des oeuvres d'arts, artistes et lieux
- *Dublin Cor*e pour le socle d'informations
- FOAF[22] (Friend Of A Friend) et RDA Group 2 Elements[23] pour représenter les informations biographiques
- ORE[24] pour agréger les données sur les objets

Pour créer l'ontologie SAAM, le modèle EDM a été étendu en ajoutant des sous-classes spécifiques (identifiants d'objets par exemple) ou en utilisant des classes et des propriétés de *schema.org* pour représenter les données géographiques.

---

[20] https://pro.europeana.eu/resources/standardization-tools/edm-documentation (consulté le 10/10/2017)
[21] https://www.w3.org/TR/2009/REC-skos-reference-20090818/ (consulté le 10/10/2017)
[22] http://xmlns.com/foaf/spec/ (consulté le 10/10/2017)
[23] http://metadataregistry.org/schema/show/id/15.html (consulté le 10/10/2017)
[24] https://www.openarchives.org/ore/ (consulté le 10/10/2017)

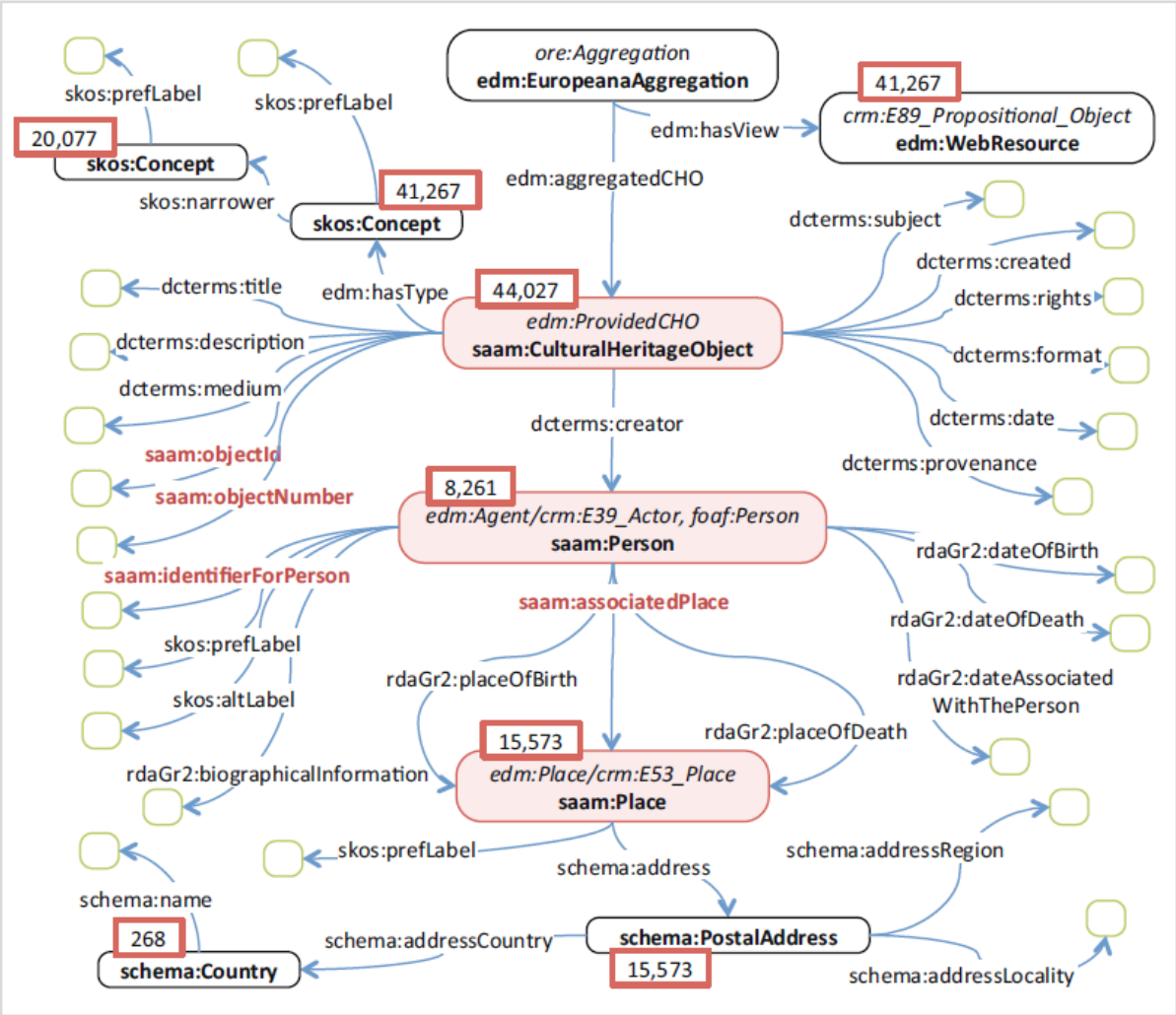

Figure 5 : L'ontologie du SAAM [21, Szekely et al. 2013]

Le CIDOC CRM comme le modèle EDM ont été jugés vastes et complexes tout en ne couvrant pas complètement les besoins du SAAM. Nativement composite, c'est EDM qui a été pris comme point de départ à la mise au point de l'ontologie SAAM. Le CIDOC CRM pourtant extensible n'a pas été retenu.

La migration des données du SAAM en RDF a été réalisée via un outil *ad hoc* de mappage KARMA[25], selon les étapes suivantes :

1. Reformatage des données (concaténation, séparation)
2. Concordance des colonnes de la base du SAAM avec les classes de l'ontologie SAAM
3. Liaison automatique des classes par inférence
4. Extraction et exploitation de recettes d'interprétation des données lors du mapping

[25] https://github.com/usc-isi-i2/Web-Karma (consulté le 10/10/2017)

Une fois les triplets exposés, la stratégie du SAAM a consisté à s'insérer dans l'écosystème du LOD en faisant porter *« son effort sur une politique d'alignement systématique des URIs avec deux acteurs du domaine culturel, le Getty Research Institute (2 110 URI alignés) et le Rijksmuseum (551 URI alignés). Deux acteurs extérieurs au domaine patrimonial ont été ajoutés, le New York Times (70 URI alignés) et DBpedia (2 194 URI alignés). Cette stratégie a permis la mise en place d'un espace numérique communautaire, interopérable sur le plan technique et en partie interopérable sur le plan culturel ».* [11, B. JUANAIS et J.-L. MINEL, 2016]

Les URIs (Uniform Ressource Identifier) servent à identifier des ressources sur le Web. Elles peuvent exister sous la forme d'URL (Uniform Ressource Locator) ou d'URN (Uniformm Ressource Name). Ce dernier type permet d'identifier les ressources indépendamment de leurs chemins d'accès tout en respectant un certain format[26].

En septembre 2014, le SAAM rejoint l'American Art Collaborative[27], consortium financé par la fondation Andrew W. Mellon. L'AAC rassemble 14 musées américains qui souhaitent ainsi former une masse critique de données ouvertes. Dans ce contexte, c'est le modèle conceptuel du CIDOC qui joue son rôle de liant sémantique[28].

2.2 Le British Museum

Le British Museum s'est engagé à partir de 2007 dans un vaste projet de numérisation et de diffusion en ligne de ses collections. Les collections en ligne comptent à ce jour « *deux millions d'objets enregistrés, dont un tiers est documenté par au moins une image en ligne ».*[29]"

Sous la direction de Dominic Oldman, le projet Research Space aboutit à l'exposition massive des collections sous la forme de triplets RDF.

Si la version OWL de l'Université d'Erlangen (eCRM) inspire très fondamentalement les équipes du ResearchSpace project, le CIDOC-CRM est intégré au sein d'une ontologie spécifique au British Museum, elle-même destinée à servir de référence.

> *"We now use local terminology to provide reification (a way of clarifying a CRM property without extending the CRM itself) and typing of CRM classes*

---

[26] Voir [8, MENON, 2016]

[27] http://americanartcollaborative.org/ (consulté le 10/10/2017)

[28] http://americanartcollaborative.org/educational-briefings/ontologies-and-the-cidoc-crm/ (consulté le 10/10/2017)

[29] http://www.britishmuseum.org/about_us/news_and_press/press_releases/2011/semantic_web_endpoint.aspx (consulté le 10/10/2017)

and properties. For example, instead of creating property extensions like 'bequeathed_by', 'donated_by', 'purchased_from', and so on, for different types of acquisition, events are simply typed and CRM properties reified using British Museum terminology. This is implemented in such a way as to allow other organisations to simply replace BM terminology with their own local vocabularies."
[20, OLDMAN, Dominic, MAHMUD, Joshan et ALEXIEV, Vladimir, 2013]

Au niveau des triplets RDF, seules les URIs de Prédicats renvoient vers le eCRM (par exemple : http://erlangen-crm.org/current/P14_carried_out_by) tandis que les Sujets et les Objets renvoient systématiquement à l'ontologie du British Museum. Ainsi les renvois bibliographiques (par exemple : http://collection.britishmuseum.org/id/bibliography/7238) sont internes alors qu'ils pourraient être alignés sur des bases bibliographiques en open data[30].

S'appuyer sur le modèle conceptuel de référence du CIDOC a permis au British Museum de s'extraire des contingences techniques et ainsi de favoriser la pérennité de ses données.

Au-delà de l'exposition en RDF, la modélisation CRM a eu pour effet d'augmenter la qualité des données produites par le musée en identifiant certaines incohérences ou lacunes.

3 Réunir Patrimoine et Récit historique

3.1 La source, fondement de l'information historique

3.1.1 Modéliser et exploiter les ressources historiques

Les conservateurs étudient leurs collections selon leurs problématiques métiers spécifiques. Il s'agit en premier lieu pour un musée d'identifier l'objet, de décrire ses caractéristiques, de retracer son parcours et de suivre les événements qui peuvent l'altérer. Une notice reste toujours perfectible et « complétable » mais elle représente de façon *a priori* fiable un objet matériel de collection.

De leur côté, les historiens envisagent l'archive comme une source partielle (et parfois partiale) d'information sur le passé. Le premier travail de l'historien consiste à "critiquer" la source c'est-à-dire à la contextualiser et évaluer son degré de fiabilité. Cette critique est déjà une analyse, elle-même potentiellement discutable, de l'archive.

Lors de la mise en place du projet SyMoGIH (Système Modulaire de Gestion de l'Information Historique) par le LARHRA (LAboratoire de Recherche Historique

[30] Voir le mémoire de master de Karine Steinberg [7, Steinberg 2016]

Rhône-Alpes) ses concepteurs avaient à trouver une réponse à deux objectifs contradictoires : la ré-utilisabilité des données d'un côté et de l'autre le sourçage de l'information.

L'un des fondements du modèle conceptuel de données SyMoGIH est que l'information historique ne provient pas directement des sources mais de leur analyse. Ainsi les contenus extraits d'une archive ne sont pas dissociables de cette dernière. La mutualisation, le raisonnement n'est possible qu'au niveau de l'information qui est une unité de connaissance opérant la synthèse entre plusieurs sources. Ce modèle se base sur des **assertions historiques sourcées** et non sur la représentation partagée mais fixée d'un domaine de connaissance. Initié en 2007 selon la méthode Merise le modèle conceptuel SyMoGIH a été formalisé en ontologie depuis 2012. Celle-ci est structurée autour de trois classes : Objet, Rôle, Unité de connaissance.

3.1.2 Alignement avec CIDOC CRM : conserver la notion de source

Bien qu'étant une ontologie de haut niveau (c'est-à-dire générique et potentiellement très expressive), le CIDOC CRM n'est pas conçu pour modéliser ce travail d'analyse de l'objet en tant que source.

Son alignement avec le CIDOC CRM est en cours d'examen[31]. Il s'agit notamment de savoir si SyMoGIH doit être alignée sur FRBRoo qui est elle-même une extension du CIDOC CRM ou si elle doit devenir une extension à part entière. Les multiples héritages de propriétés rendent l'exercice complexe et amènent les historiens à "*s'interroger sur la nature des entités*[32]" de SyMoGIH.

3.1.2.1 Respecter l'unicité de la source

En effet, les seuls « Endurants » dont l'historien soit sûr ce sont ses sources (Objet concert, Bibliographie, Objet digital, Unité documentaire). Les acteurs dont il est question dans une archive, par exemple un extrait de baptême, n'ont pas d'existence en eux-mêmes dans la sémantique de SyMoGIH. Or pour le CIDOC CRM l'existence d'une personne n'est pas remise en question. E21_Person hérite de E39_Actor qui hérite de E77_Persistent_Item et l'ancre dans les Endurants.

---

31 https://groupes.renater.fr/wiki/histoire_numerique/ontologie/aligning_symogih_org_cidoc_crm

32 https://groupes.renater.fr/wiki/histoire_numerique/ontologie/aligning_symogih_org_cidoc_crm

Du SyMoGIH au CIDOC CRM les acteurs ne sont donc plus considérés comme des contenus entièrement dépendants de la ressource, mais comme des Endurants autonomes. En revanche, *« l'unité de connaissance "naissance" est un perdurant identifié [...]. Cet objet, qui est une assertion de l'historien d'après une source, est une représentation de la réalité*. »

A l'issue de l'Atelier-Rencontre entre les membres du LARHRA-PHN et l'ISL-CCI à Herklion du 15 au 17 novembre 2016[33], il a pu être précisé que l'unité de connaissance était une information datée et sourcée.

Ainsi l'assertion "Brutus a tué César le 15 mars de l'an 44 av. J.-C. dans la Curia Pompeia sur le Champ de Mars à Rome selon Plutarque dans sa Vie de César en 66 ap. J.-C" est conforme à la contrainte de sourçage de l'ontologie SyMoGIH et transposable en CIDOC en décomposant l'entité temporelle (perdurant) "meurtre de César" en autant de sources.

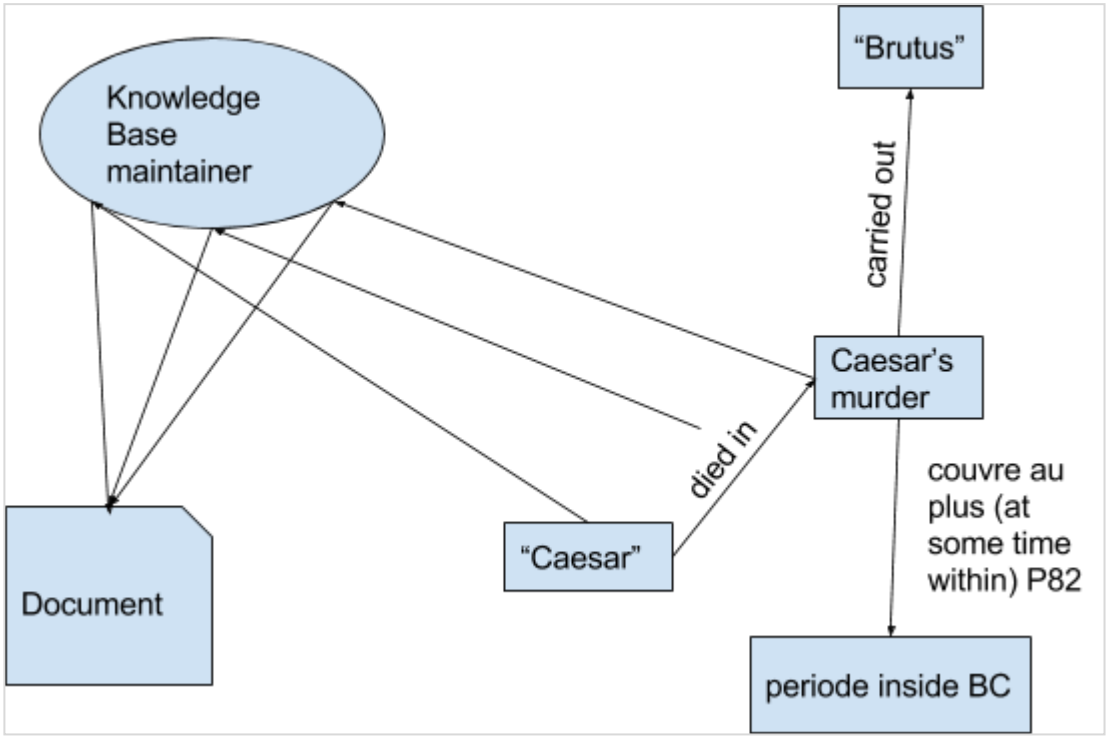

Figure 6 : Modélisation CRM d'une unité de connaissance "meurtre de Brutus" issue d'une source documentaire - d'après Atelier-rencontre avec Martin Doerr, 15-17 novembre 2016

[33] Le compte rendu de cet atelier-rencontre est disponible sur le wiki de la liste « Histoire_Numérique » : https://groupes.renater.fr/wiki/histoire_numerique/ontologie/workshop_201611 (consulté le 23/10/2017)

3.1.2.2 Affirmer l'Information comme Entité temporelle

Pour résoudre le problème, il serait possible de considérer l'entité Information du SyMoGIH comme une sous-classe d'Entité temporelle, dans le cadre d'une extension du CIDOC CRM. Dans ce cas la combinaison information, source et date formerait une unité de connaissance sans qu'il y ait besoin d'ajouter une classe spécifique "Unité de connaissance" au CIDOC CRM.

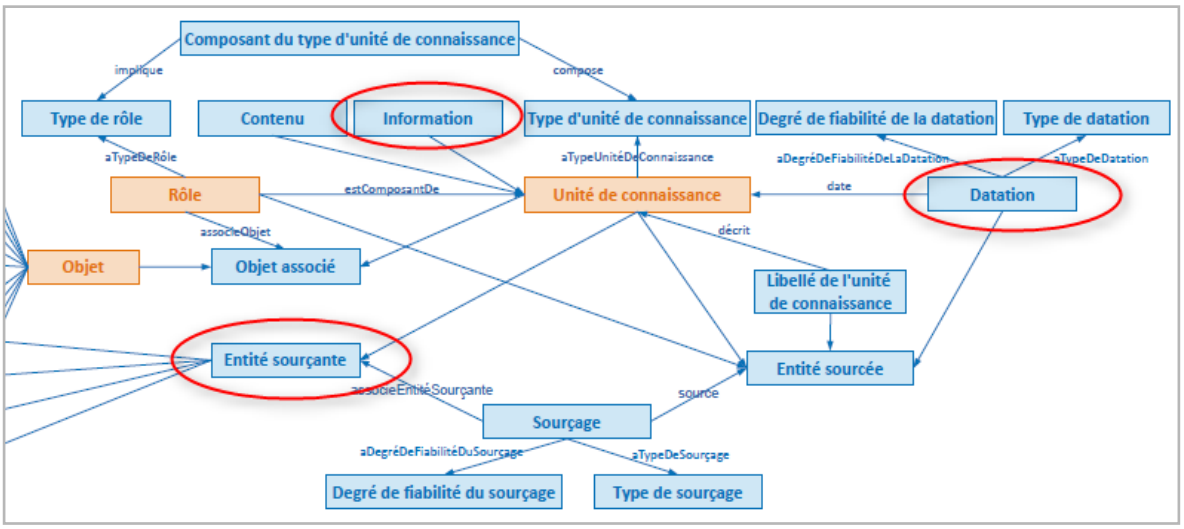

Figure 7 : Mise en évidence des entités pivots pour la conversion en CIDOC CRM (entourées en rouge) d'après Schéma simplifié de l'ontologie SyMoGIH. (LARHRA 2016[34])

3.1.2.3 Généraliser les rôles pour les lier aux propriétés

En organisant les rôles SyMoGIH par grandes catégories il serait possible de les relier avec les propriétés du CIDOC CRM.

En résumé, la transposition vers le CIDOC CRM via une extension spécifique ou via FRBRoo s'oriente vers le découpage suivant :

- Objet > E77_Persistent_Item (Endurant)
- Type d'information > E2_Temporal_Entity (Perdurant)
- Rôle > Propriétés

3.2 L'objet "conteneur de connaissance"

Nous avons vu que le CIDOC CRM était envisagé par les historiens du LARHRA (LAboratoire de Recherche Historique Rhône-Alpes) comme un outil pour valoriser les résultats de la recherche historique via l'écosystème du Web de données liées.

[34] Extrait du poster présenté à Lyon en mai 2016 au Forum Humanités Numériques Lyon / Saint-Etienne, par l'équipe du Pôle Histoire numérique du LARHRA UMR5190

Ce lien entre recherches historiques et inventaires patrimoniaux est également au cœur du projet Nantes1900[35]. Ce projet initié par le musée d'histoire de Nantes a pour but « *d'améliorer la présentation de la maquette du port, à l'aide d'outils interactifs, pédagogiques et technologiques*[36] ». L'originalité de ce dispositif de médiation est de reposer sur une base de données évolutive, capitalisant les connaissances historiques en rapport avec le territoire représenté par la maquette. Les sources valorisant les collections sont de tout type : images, synthèses historiques, apports des visiteurs qui sont invités à participer en proposant leurs connaissances et documents. Autre différence avec des dispositifs médiation plus classiques, Nantes1900 propose des scénarios guidant la découverte de la maquette mais permet également au visiteur d'approfondir un aspect à tout moment en requêtant la base de connaissance.

Le projet supposait de réaliser un triple décloisonnement :

1. entre le récit de l'historien et l'objet de collection de musée
2. entre le visiteur et l'objet exposé
3. entre la valorisation et les sources de la connaissance

Privilégiant la dimension opérationnelle, les concepteurs du projet ont écarté a priori l'usage de thésaurus externes ou d'ontologies. Pour alimenter la base de connaissance, une étape importante de découpage et de structuration des textes historiques était indispensable. Trois types de descripteurs ont été employés : temporels (norme ISO 8601), sémantiques, et géographiques. Pour être valide une information historique doit posséder au moins une de ces trois composantes qui permet de calculer la proximité entre deux items. L'outil propose des rapprochements automatisés qui facilitent le travail d'interprétation des historiens et des médiateurs. Il est toujours possible et indispensable de gérer manuellement la proximité entre deux items. Plus que la production de raisonnements logiques, la valeur ajoutée du dispositif réside dans sa capacité à capitaliser la connaissance. Il permet aussi la confrontation des différents points de vue. L'objet est à la fois conteneur de connaissance et jalon dans la trame de plusieurs récits possibles.

---

[35] [26] QUANTIN, Matthieu, LAROCHE, Florent et KEROUANTON, Jean-Louis, 2016.
[36] http://www.chateaunantes.fr/fr/nantes-1900 (consulté le 23/10/2017)

Nantes1900 et le SyMoGIH partagent des préoccupations communes : traçabilité des sources, modélisation de la donnée historique, cohabitation d'interprétations différentes ou de récits contradictoires.
Toutefois, le projet nantais a abouti à une base de connaissance *ad hoc*, qui n'a pas vocation première à participer au web de données liées mais à l'enrichissement mutuel des récits produits par les historiens et des objets patrimoniaux de la production des données à leur diffusion.
Les enrichissements potentiels avec d'autres collections et d'autres ressources sémantisées par ailleurs serait une nouvelle étape qui supposerait de raccorder ou d'aligner la base de connaissance Nantes1900 sur une ontologie formalisant une vision partagée de la discipline historique et de la description des biens patrimoniaux.

3.3 L'archive support d'un discours

La description et l'analyse de contenu est indispensable à la gestion des archives audiovisuelles. En effet, une représentation imagée (photographique par exemple) d'un objet de collection ou la table des matières d'un livre donne assez facilement une idée de ce qu'ils sont ou de leurs contenus informationnels. En revanche, pour être préhensibles et ré-utilisables les archives audiovisuelles doivent être analysées et modélisées. Cette contrainte liée à la nature du bien culturel "archive audiovisuelle" permet d'objectiver les niveaux de discours et de générer une base de connaissance non seulement sur les sujets abordés mais également sur la façon dont ils ont été traités. Décrire un item (objet, livre, archive), c'est le doter d'un ensemble de données textuelles qui "*documentent, conservent et transmettent le discours d'un acteur social*" [27, STOCKINGER, LALANDE, BELOUED, 2015]. Les méthodes développées dans le cadre du Campus AAR (Archives Audiovisuelles de la Recherche) sous l'impulsion de Peter Stockinger viennent enrichir la réflexion sur les modèles d'analyse et de représentation des biens culturels en général.

Quatre étapes de traitement sont nécessaires pour modéliser une représentation exploitable d'un document média dans le cadre du Campus AAR. En premier lieu l'auteur ou le propriétaire d'un fonds de vidéos le dépose dans MEDIHAL qui constitue la base de données de Campus AAR. A cette occasion il associe des métadonnées aux items déposés. La deuxième étape est celle du moissonnage de la base MEDIHAL (ressources et métadonnées) par le Studio Campus et de l'attribution d'un modèle de description compatible avec des vocabulaires externes

en fonction de la nature de l'archive et des besoins de ré-utilisation. Il peut y avoir plusieurs modèles pertinents pour une même archive. Ce n'est que dans un troisième temps que le travail d'analyse du contenu est engagé. L'analyste localise les corpus de vidéo en fonction des sujets, les analyse et enregistre ses descriptions. Dans un quatrième temps, le gestionnaire éditorial peut ouvrir les ressources ainsi enrichies à des portails OAI et publier les analyses sur Campus-AAR.

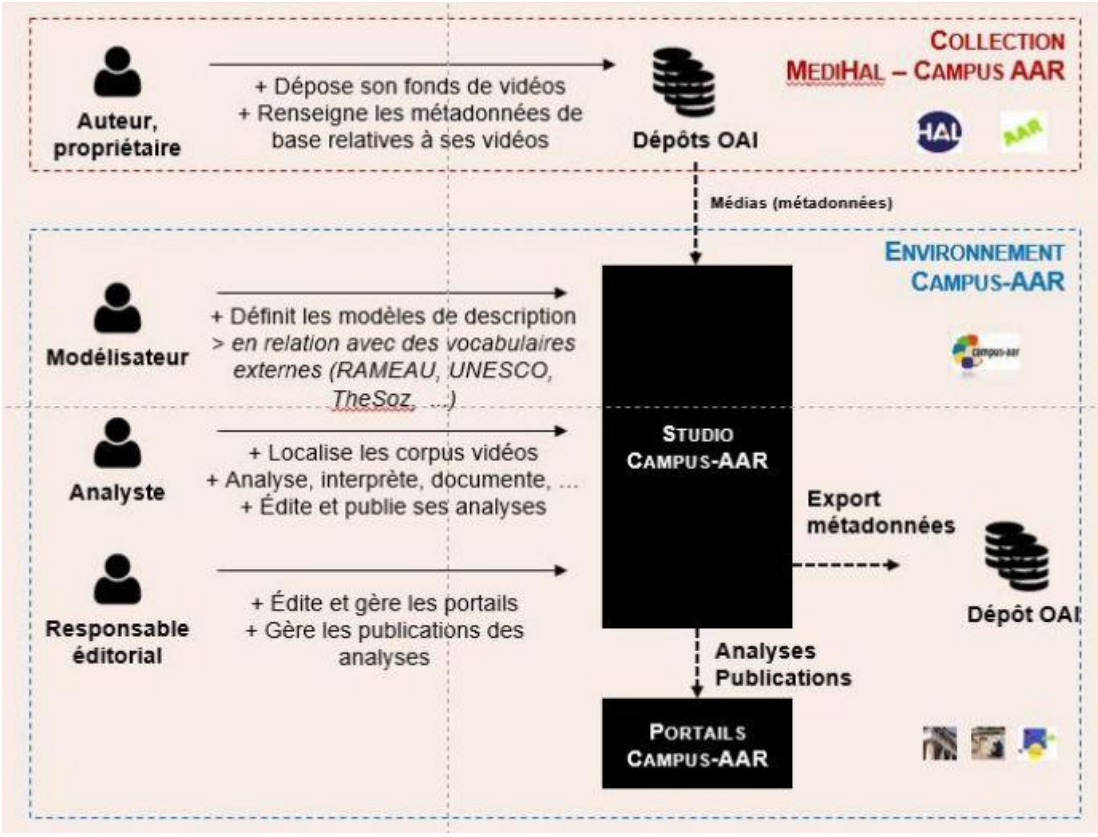

Figure 8 Processus de traitement des médias par le dispositif Campus-AAR[37]

Pour indexer un objet ou un corpus d'objets média l'analyste s'appuie sur des modèles d'analyses, en l'occurrence le métalangage ASA (Atelier de Sémiotique Audiovisuelle). Dans cette ontologie, « *l'objet "texte" n'est pas du tout restreint à son avatar culturel du texte imprimé (du livre, du journal...) mais recouvre plutôt toute "entité signifiante*" » [27, STOCKINGER, LALANDE, BELOUED, 2015]. Dans le

[37] https://f.hypotheses.org/wp-content/blogs.dir/1730/files/2016/09/Campus-AAR-poster-Interagir-comprendre-creer-Espace3.pdf (consulté le 26/10/2017)

schéma ci-dessous l'objet ou le corpus média est compris selon trois dimensions structurantes :

1. objet compositionnel (partie d'une séquence, collection ou fonds)
2. objet stratifié (le discours et son expression)
3. objet d'une pratique signifiante (usage, incluant potentiellement celle de son ou ses analystes et de son ou ses ré-utilisateurs)

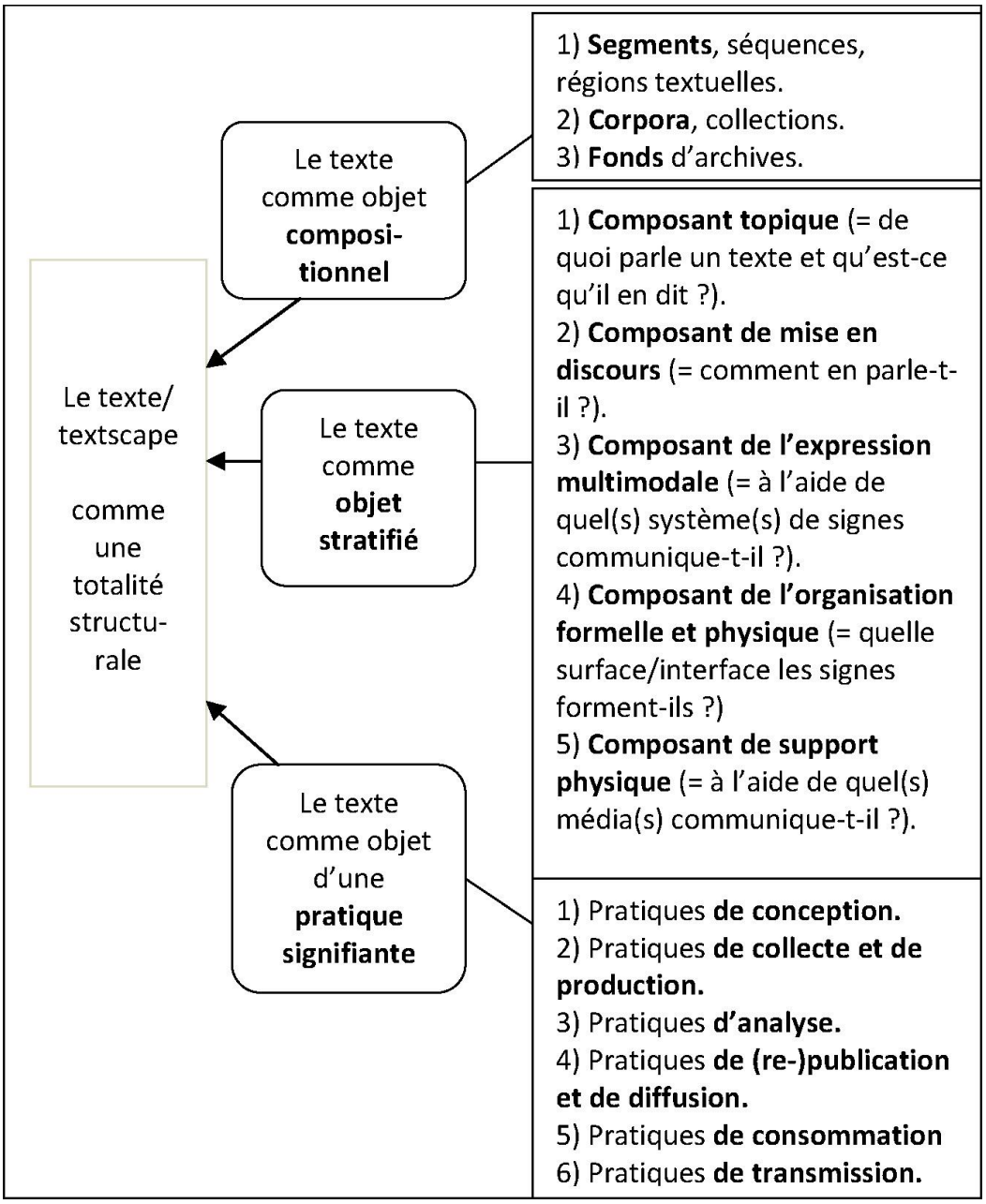

Figure 9 Structure sémiotique d'un objet ou corpus média [27, STOCKINGER, Peter, LALANDE, Steffen et BELOUED, Abdelkrim, 2015]

Cette conception de l'objet audiovisuel permet de dégager de l'archive un « discours-objet » (sujet, rhétorique, mise en scène audiovisuelle, métadonnées) et

de conserver la trace de la procédure d'analyse (analyse libre, contrôlée, spécialisée, liée à un standard externe).

L'ontologie ASA fonctionne comme une bibliothèque de modèles d'analyse (ou scénarii) hiérarchisés selon trois « topiques du discours d'archives identifiées » représentant trois domaines identifiés du discours :

- AGORA pour le patrimoine audiovisuel des sciences humaines et sociales
- ARC (A la Rencontre des Cultures) pour la diversité culturelle - INALCO
- AHM (Archives Histoires des Mathématiques) pour la préservation et la valorisation du patrimoine scientifique (CNRS)

Deuxième partie : Positionnement et évolutions possibles pour A&A Partners

4 L'entreprise et ses solutions

4.1 Le contexte institutionnel

4.1.1 Les préconisations du SMF (Service des musées de France) du ministère de la Culture

Les musées bénéficiant du label "musée de France" partagent la même ambition définie par la loi n° 2002-5 du 4 janvier 2002 relative aux musées de France transcrite dans le code du patrimoine depuis 2004 :

a) Conserver, restaurer, étudier et enrichir leurs collections ;
b) Rendre leurs collections accessibles au public le plus large ;
c) Concevoir et mettre en œuvre des actions d'éducation et de diffusion visant à assurer l'égal accès de tous à la culture ;
d) Contribuer aux progrès de la connaissance et de la recherche ainsi qu'à leur diffusion." (Article L.441-1 du Code du patrimoine)

Les collections des musées de France sont imprescriptibles (L.451-3) et dans le cadre de la domanialité publique, inaliénables (L.451-5). Pour consacrer et garantir ce statut patrimonial spécifique, les collections des musées de France font l'objet d'une inscription sur un inventaire et sont récolées tous les dix ans (L. 451-2). Les modalités techniques et matérielles de la tenue de l'inventaire et du récolement décennal sont précisées dans un arrêté daté du 25 mai 2004[38]. Le Bureau de la politique documentaire et numérique des collections est chargé de collecter les notices d'inventaire des musées sur tout le territoire national et de les mettre en ligne sur le portail Joconde.fr qui constitue le Catalogue collectif des musées de France.

[38]http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDEMUSEES/societe-info.htm (consulté le 26/10/2017)

4.1.2 Les portails culturels

4.1.2.1 Le portail Joconde

L'expérimentation Joconde Labs[39] a permis d'enrichir les rubriques indexées des notices d'œuvres en les liant avec DBpédia[40], la base RDF alimentée par Wikipédia. Cette association avec Wikipédia présente plusieurs avantages :

- les données enrichies sont également multilingues dans la mesure où des articles Wikipédia équivalents existent dans les autres langues
- les articles sont mis à jours et révisés régulièrement par la communauté des contributeurs à Wikipédia
- la participation du public est possible via Wikipédia

4.1.2.2 Le moteur Collections du ministère de la Culture

Le moteur Collections est un point d'accès unique au patrimoine numérisé du ministère de la Culture, des établissements publics, des collectivités territoriales et des associations.

> *« Plus de 7 millions de documents et plus de 5 millions d'images donnant un accès direct à 70 bases de données, 627 expositions virtuelles et 177 sites Internet provenant tant du ministère que de sites partenaires. [41]»*

Avec l'INA (Institut National de l'Audiovisuel) et la BnF (Bibliothèque nationale de France) le moteur Collections est un agrégateur national qui alimente le portail numérique Européana.

4.1.2.3 La bibliothèque numérique Europeana

La bibliothèque numérique Europeana est lancée en 2008 par la Commission Européenne. Elle met en commun les références de ressources numériques des bibliothèques nationales des 27 états membres.

Son fonctionnement repose sur le principe de la syndication : les données ne sont pas intégralement recopiées mais agrégées dans Europeana. L'accès à l'information détaillée se fait en ouvrant la notice originale dans son contexte propre.

En 2010 Europeana s'ouvre à d'autres types de documents et notamment aux notices d'œuvres des musées. En France, le portail Joconde centralise la diffusion sur ce portail, ce qui altère la dimension horizontale d'Europeana qui vise à agréger dans toute leur richesse des ressources hétérogènes.

---

[39] http://jocondelab.iri-research.org/jocondelab/ (consulté le 26/10/2017)
[40] http://fr.dbpedia.org/ (consulté le 26/10/2017)
[41] http://www.culture.fr/Ressources/Moteur-Collections (consulté le 28/11/2017)

4.1.3 Le programme HADOC (HArmonisation pour la production des DOnnées Culturelles)

HADOC[42] est un programme (c'est-à-dire un ensemble de projets visant le même objectif) initié en 2008 par le ministère de la Culture. Il part du constat que les données produites par les différentes institutions contiennent trop d'informations implicites et spécifiques à chaque métier, voire à chaque producteur pour être valorisés dans le cadre de l'Open data. Sous la direction de Katell Briatte, ce programme propose des outils conceptuels et techniques pour améliorer l'exploitabilité des données dans les contextes d'échanges et participation à des portails culturels sémantiques.

4.1.3.1 Un modèle de référence bis ?

Le « Modèle harmonisé pour la production des données culturelles[43] » esquissé en 2012/2013 a pour ambition d'être directement compatible avec le modèle de référence du CIDOC tout en conservant les spécificités métiers de chaque domaine culturel.

> *« Contrairement à d'autres initiatives du champ patrimonial ou culturel, le Modèle harmonisé pour la production des données culturelles ne constitue pas strictement un modèle d'échange. Le choix de la modélisation en UML vise à en faire également un modèle d'implémentation pour les applications informatiques, adaptable aux contextes métier et aux contraintes locales. Toutefois, la compatibilité sémantique avec ces initiatives a constamment été recherchée, de manière à proposer dès la phase de production une structuration de l'information interopérable avec les modèles d'échange et de diffusion que constituent par exemple, dans le champ patrimonial, le Modèle conceptuel de référence du CIDOC (ou CIDOC CRM) et ses extensions, le modèle EDM ou le schéma XML LIDO ».*
> (BRIATTE 2013, p.9)

Le modèle défini ressemble à une simplification du CIDOC CRM permettant de cerner les entités utiles pour représenter un Bien Culturel tout en se présentant comme un modèle destiné à être adapté en fonction des besoins métiers. Si la démarche permet de bien identifier la problématique concrète de diversité des pratiques et des notions implicites, on peut se demander si cette étape de modélisation entre base de production et entrepôt sémantique ne rend pas le

---

[42] http://www.culturecommunication.gouv.fr/Divers/Harmonisation-des-donnees-culturelles/Modele-de-donnees/Enjeux-et-objectifs (consulté le 28/11/2017)

[43] http://www.culturecommunication.gouv.fr/content/download/74065/565300/version/1/file/MCC-HADOC-REF-modele_harmonise_donnees_culturelles.pdf (consulté le 28/11/2017)

processus plus complexe. En effet, le schéma ci-dessous limite à l'essentiel les relations mais reste très générique.

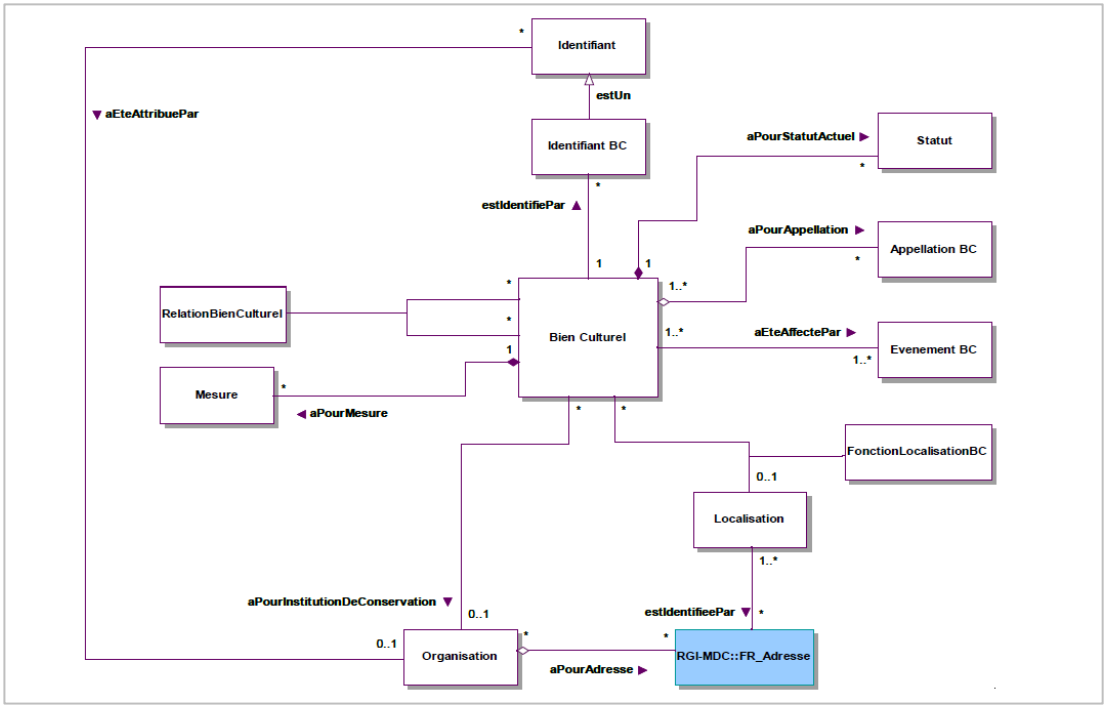

Figure 10 : Schéma sémantique simplifié du modèle HADOC (Katell Briatte)

4.1.3.2 MediHadoc

MEDIHADOC est un projet de mise au point d'un schéma de métadonnées pour le multimédia compatible avec le modèle de données HADOC. Il s'agit de gérer deux niveaux d'informations : celui du support image ou multimédia et celui du bien patrimonial constitué par l'image ou représenté par elle soit trois cas de figures :

- image illustrative
- image objet patrimonial
- sujet représenté par un bien culturel

Le schéma MediHadoc permet non seulement de décrire la ressource, mais également l'original.

4.1.3.3 GINKO : un outil pour harmoniser les vocabulaires scientifiques et techniques

Développé sous licence libre, le logiciel GINKO (Gestion Informatisée de Nomenclatures Collaboratives et Ouvertes) est l'un des projets du programme HADOC. Il doit permettre de rassembler des listes de termes utilisés localement pour l'indexation au sein de solutions propriétaires et les transformer en référentiels

terminologiques ré-utilisables par la communauté des gestionnaires de données. Son rôle est aussi de les rendre disponibles.

4.1.3.4 Identification des ressources et référentiels : le cas « ONOMA »

« ONOMA » (du grec ancien désignant un nom propre) répond à un besoin du domaine culturel de disposer d'un référentiel d'Acteurs (auteurs, créateurs, producteurs, personnalités intervenant dans le cycle de vie d'un Bien culturel).

> *« Le projet comprend la réalisation de l'outil de gestion des données de référence et la reprise des données existant dans les différents corpus [44]».*
> [14, Onoma : un référentiel d'acteurs du ministère de la Culture ancré dans le Web sémantique 2017]

Ce référentiel doit servir aussi bien en production pour faciliter la saisie, limiter les sources d'erreur et les redondances, qu'en consultation/recherche pour tirer parti de l'accès à une description unique et liée à toutes les ressources : complétude de l'information, rebond vers d'autres ressources.
Il doit enfin s'intégrer à l'écosystème du Linked Open Data.

4.2 La plateforme WebMuseo

La plateforme WebMuseo est une boîte à outils spécialisée dans la gestion et la publication web des collections. Un espace WebMuseo définit, au sein d'un modèle, un ensemble de caractéristiques (tables, liens de référencements entre les tables…) gérées par des outils communs à la plateforme ainsi que par des applications spécifiques au modèle.
Les solutions WebMuseo sont accessibles depuis n'importe quel navigateur, sans installation supplémentaire et depuis une connexion sécurisée par identifiant et mot de passe. Développées sur une plateforme commune elles sont nativement :

- multi-utilisateurs et sécurisées,
- accessibles selon des droits différenciés (à minima lecture / lecture +écriture),
- dotées d'un système de gestion de bases de données,
- structurées et optimisées pour l'échange de données web selon le protocole OAI-PMH

[44] http://www.culturecommunication.gouv.fr/Divers/Harmonisation-des-donnees-culturelles/Referentiels/Le-referentiel-des-acteurs-historiques/ONOMA (consulté le 28/11/2017)

4.2.1 WebMuseo Gestion

4.2.1.1 Périmètre fonctionnel

Le modèle standard WebMuseo *Gestion* permet de mettre en application l'ensemble du système descriptif préconisé par la *Méthode d'inventaire informatisé*[45] (disponible sur le site du ministère de la Culture) sans se restreindre à celui-ci. Les 1400 champs de la base de données couvrent l'ensemble des besoins de gestion des collections :

- **inventaire documentaire des collections** (compatible avec les champs et les thésaurus de la base nationale Joconde),
- **documentation annexe**
  - Expositions, Biographies, Evénements, Sites, Thésaurus et listes
  - Archives, Œuvres bibliographiques
  - Photographies, Enregistrements sonores, Vidéos, Ressources numériques
- **gestion physique des œuvres** (Localisations, Sorties, Constats d'Etat, Restaurations),
- **valorisation des collections :** exports Joconde, gestion de la diffusion, entrepôt OAI-PMH,
- **édition des registres légaux** (Inventaire et Dépôts),
- **récolement décennal** (Plans, Campagnes, Fiches, Calcul des statistiques et génération des annexes de PV).

4.2.1.2 Modèle de données

La base de données s'articule autour d'un enregistrement central permettant l'établissement et la mise à jour de l'inventaire documentaire des collections : la notice d'œuvre. Celle-ci est organisée selon trois grands types de statut : Bien affecté, Bien reçu en dépôt, Bien reçu en prêt, correspondant à trois tables annexes. Le réseau des tables offre une infinité de points de vue sur la collection : il est ainsi possible de filtrer la base par Acquisition, par Emplacement, par Créateur, etc.

- Trois tables annexes (une par statut) sont dédiées à la gestion des statuts juridiques des objets. Elles sont accessibles dans la section «Statuts ».
- Les autres tables annexes permettent de détailler un domaine documentaire (Biographies, Œuvres bibliographies, Photographies, etc.) et sont le cas échéant référençables dans d'autres tables.
- Les cinq tables de gestion physique (Localisations, Fiches de Sorties, Constats et Restaurations, Récolements) référencent la notice. Elles peuvent également se référencer entre elles et faire appel aux tables documentaires annexes.

---

[45] http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDEMUSEES/methode.htm (consulté le 6/11/2017)

- Les tables Registres permettent de tenir et d'éditer les registres légaux. Un niveau de droit spécifique en sécurise l'accès.

De nombreux thésaurus et listes (comprenant les thésaurus du système descriptif mis en place par le Service des musées de France) sont intégrés au sein de tables annexes exploitées notamment par la table « Objets de la collection ».

4.2.1.3 Modules métiers validés par le SMF

Les modules Registre, Récolement décennal et Export Joconde ont été développés en collaboration avec le SMF (Service des musées de France).

4.2.1.3.1 Extraction et impression des registres

L'application « Registres » de WebMuseo Gestion permet de répondre aux exigences légales relatives à la tenue de l'inventaire des collections des musées de France (au sens de la loi de la loi 2002-5 du 4 janvier 2002).

A chaque fiche documentaire peut correspondre une fiche Registre. Un registre est un enregistrement créé à la suite de la compilation annuelle des fiches registres.

Seul le document imprimé sur papier à partir du PDF généré par WebMuseo possède la valeur juridique d'un registre d'inventaire ou d'un registre des dépôts.

4.2.1.3.2 Récolement décennal des collections

Le récolement décennal est une obligation légale à laquelle sont tenus les musées de France. Les différents textes regroupant les modalités du récolement décennal informatisé sont disponibles sur la page http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDEMUSEES/recolement-informatise.htm du portail Joconde.

La solution WebMuseo Gestion permet de traiter cette exigence de gestion physique et administrative des collections quelles que soient les méthodes mises en œuvre : fiches papier, saisie hors connexion sur tablette ou saisie directe dans WebMuseo, récolement en partant de l'objet ou de l'inventaire.

WebMuseo Gestion permet non seulement de renseigner des fiches de récolement, mais également d'assurer le suivi des campagnes et de générer automatiquement et à volonté les statistiques nécessaires à la production des Procès Verbaux de Récolement.

De plus, des fiches régies (Localisations, Constats d'état, Tâches) sont créées automatiquement à partir des informations saisies dans la fiche de récolement.

Un enregistrement « Plan » permet de documenter le plan de récolement programmant les différentes campagnes.

4.2.1.3.3 Export vers Joconde.fr

Les notices d'œuvres WebMuseo Gestion sont exportables vers la base Joconde via un assistant qui vérifie la présence de contenu au niveau des champs obligatoires et met aux normes les images associées exportées (taille et format de fichier).

Une fois l'export validé, il est communicable au service des musées de France sous forme de fichier zippé. Une URL dédiée à son téléchargement est également disponible.

Une même notice peut être exportée plusieurs fois pour mise à jour. La procédure est identique qu'il s'agisse d'un premier export ou d'une mise à jour. Un même export peut concerner indifféremment des notices relevant des deux cas. Les exports de mise à jour sont différentiels : seules les données enregistrées depuis le dernier versement sont extraites. L'historique des différents exports d'une notice reste toujours disponible.

Si des corrections doivent être apportées à une notice Joconde, elles doivent l'être depuis la base de données avant d'être exportée et transmise aux gestionnaires de Joconde (notion de notice unique).

4.2.2 WebMuseo Publication

4.2.2.1 Catalogue en ligne

La publication des notices d'œuvres gérées sous WebMuseo *Gestion* au sein d'un espace WebMuseo *Publication* s'appuie sur le protocole d'échanges **OAI-PMH** (Open Archive Initiative – Protocol for Metadata Harvesting).

Dans ce cas, l'espace WebMuseo *Gestion* joue le rôle du ***fournisseur de données*** et l'espace WebMuseo *Publicati*on celui du ***fournisseur de service***. C'est donc depuis l'application WebMuseo *Publication* qu'est déclenché le moissonnage des notices d'œuvres saisies sous WebMuseo *Gestion.*

4.2.2.2 Valorisation des collections

WebMuseo permet de créer des expositions virtuelles selon un principe modulaire très simple : celui de la vitrine. Se référençant les unes les autres, les vitrines virtuelles permettent de concevoir les parcours les plus variés dans les collections. Au niveau de la navigation, un fil d'Ariane permet au visiteur de se repérer et de remonter la chaîne des vitrines visitées.

Techniquement, une vitrine virtuelle est un enregistrement permettant la présentation et la valorisation des notices publiées. Elle est notamment composée :

- d’un contenu spécifique (illustrations, titre*, auteur, texte, résumé, liens, fichiers à télécharger, etc.),
- de liens référençant des notices d’œuvres,
- de liens référençant d’autres vitrines.

4.3 L'accompagnement des utilisateurs

4.3.1 Formation à l'utilisation

Le catalogage des collections met en jeu trois niveaux de connaissance : l'expertise historique et scientifique de l'objet, la connaissance du langage documentaire et la maîtrise du logiciel de gestion de base de données. L'action de formation à l'utilisation de WebMuseo Gestion est prioritairement ciblée sur la maîtrise fonctionnelle du logiciel. L'objectif premier est de permettre à l'utilisateur de bien comprendre la structure relationnelle de la base de données. Par exemple, la rubrique Créateur n'est pas un champ texte dans le lequel il faudrait ré-écrire à chaque fois le nom d'un artiste, mais une clé permettant de créer un lien vers la fiche Biographie de l'artiste.

Toutefois, la formation s'appuie autant que possible sur des exemples concrets de notices d'œuvres et des cas d'usage musée : recherche par date ou thématique, création de dossiers, suivi des sorties d'œuvres (ou prêts).

4.3.2 Migration de données

La reprise des données produites sur d'anciens supports (autre logiciel de gestion, tableurs Excel, logiciel générique de base de données tels qu'Access ou 4D) est une étape décisive dans la livraison d'une solution de gestion.

Quatre scénarios sont possibles en fonction de deux variables : anciens supports toujours en production ou pas ; solution WebMuseo installée sur les serveurs du client ou hébergée par A&A Partners.

La solution WebMuseo Gestion intègre dans sa version standard des thésaurus, et notamment ceux proposés par le Service des Musées de France (ministère de la Culture). Les principales rubriques d’indexation sont alimentées par des tables liées contenant les descripteurs (Dénominations, Domaines, Techniques, Matériaux, Utilisations, Représentations, etc.). La conformité des données sources avec ces thésaurus est prise en compte lors de la migration.

Le schéma ci-dessous présente le scénario de migration le plus courant : reprise de données avec interruption de service minimum (1 semaine) et hébergement par A&A Partners.

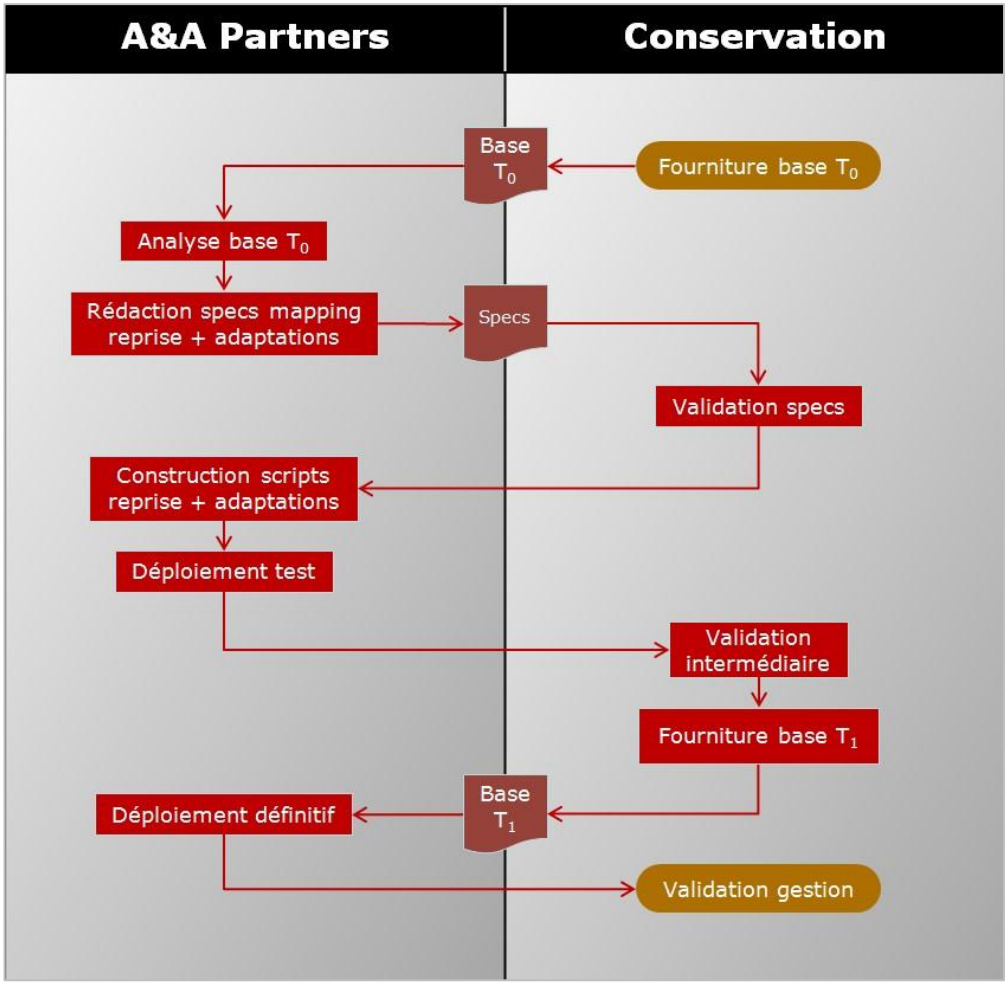

Figure 11 : Processus de migration des données (EP, A&A Partners)

4.3.2.1 Fourniture de la base T0

La prestation de migration démarre par la fourniture exhaustive des données prévues dans leur état en début de projet. Ce jeu de données forme un tout appelé "Base T0". En parallèle de la mise au point de la migration, l'équipe du musée peut poursuivre la saisie selon les mêmes partis pris, de façon à limiter le différentiel avec les données qui seront fournies dans leur état définitif (Base T1). Entre T0 et T1, le chef de projet côté musée s'assure que ces conditions de saisie sont bien comprises et appliquées par son équipe et celles de ses partenaires.

4.3.2.2 Analyse de la base T0 et rédaction des spécifications

La base T0 est analysée afin de formaliser des règles de reprise au sein d'un document de spécifications : concordance entre les rubriques, ventilation des données, filtres selon thésaurus, corrections automatisées, rattachement des illustrations. Le cas échant, des adaptations du modèle de données standard peuvent y être précisées.

4.3.2.3 Validation des spécifications et Développement des scripts de reprise

La validation des spécifications par le chef de projet côté musée permet d'engager le développement des scripts de reprise à partir des sources T0. Ces scripts sont re-jouables à volonté pour mettre au point la migration sans altérer les sources.

4.3.2.4 Déploiement test

Lorsque les scripts sont achevés, un premier espace WebMuseo de test est déployé. Il contient l'intégralité des données reprises.

La partie marginale des données n'ayant pu être ventilée, corrigée ou reformatée est sauvegardée en tâches liées aux notices. La table de gestion des tâches offre également une vision des anomalies selon un **journal des rejets** organisé par rubriques.

De plus, chaque notice contient un champ "Données sources" rappelant le texte brut de la notice originale. La façon dont les données ont été ventilées dans WebMuseo reste ainsi toujours directement observable.

La livraison du premier espace de tests est accompagnée d'une présentation du résultat et d'une initiation à l'utilisation de l'application, sur place ou à distance par conférence téléphonique avec partage d'écran (via TeamViewer par exemple). Les agents validateurs sont ainsi en mesure de contrôler la reprise et les adaptations réalisées.

4.3.2.5 Validation intermédiaire

Lors de la phase de vérification, les agents validateurs s'assurent de la conformité du résultat avec les spécifications. Ils font part de leurs remarques et peuvent suggérer des améliorations. La migration peut être rejouée par A&A Partners jusqu'au déploiement d'un espace parfaitement conforme. A chaque nouvelle version, le contrôle des notices témoins sélectionnées en amont permet de s'assurer de la non-régression de la migration entre deux versions.

4.3.2.6 Fourniture base T1

La validation de la migration T0 entraine la fourniture des données définitives (T1) et l'abandon complet de la saisie sur les anciens supports. Les données T1 doivent être une simple et ultime version d'actualisation des sources fournies au démarrage de la prestation (même format et même structuration des données). Elles sont importées selon les scripts mis au point pour T0.

4.3.2.7 Déploiement définitif

L'espace WebMuseo déployé en production est une dernière fois vérifié. Au besoin de légers correctifs de scripts induits par l'actualisation des sources peuvent être apportés. Sa validation marque la fin de la prestation de migration. Les comptes utilisateurs sont ouverts et la formation des utilisateurs finaux dispensée afin que le travail puisse reprendre sur la solution WebMuseo. Les rejets sauvegardés en tâches peuvent alors être corrigés manuellement par les utilisateurs concernés.

4.3.3 Développements spécifiques

La société A&A Partners propose des solutions de gestion et de valorisation "sur étagère", modulables, paramétrables et pouvant être enrichies selon des développements spécifiques.

L'offre standard est composée de différents modules, empilables : inventaire documentaire des collections, gestion physique des œuvres, publication en ligne des notices, etc.

Afin de réduire les délais de livraison et d'optimiser les coûts de maintenance, le développement d'outils ou d'interfaces spécifiques est idéalement précédé d'une présentation de nos solutions standard et du recueil des besoins compte tenu du socle fonctionnel présenté.

4.3.3.1 Présentation du socle fonctionnel

La présentation de l'application dans sa version standard permet de familiariser les utilisateurs avec les fondamentaux de la plateforme WebMuseo. Quel que soit le modèle il est toujours adossé à un socle fonctionnel : système de données relationnel, outils de recherche et de navigation communs, modules spécialisés standards : documentation des collections, suivi des sorties d'œuvres, catalogage de la bibliothèque, etc.

4.3.3.2 Recueil du besoin

Au-delà de ce premier contact avec l'environnement fonctionnel, la présentation initiale offre la possibilité de lister les cas d'usage et d'évaluer l'écart entre le besoin et la solution existante.

La demande est formalisée par le chef de projet A&A Partners au sein d'un document de spécifications compréhensible (et au besoin explicité) des clients et suffisamment précis pour guider les ingénieurs chargés du développement. La mise

au point de ce document charnière peut faire l'objet de plusieurs allers-retours entre le chef de projet MOE (A&A Partners) et le responsable du projet MOA (Client).

4.3.3.3 Développement, vérification et validation des adaptations

Les développements sont effectués sur la base des spécifications validées. Toutefois, un premier prototype, pas forcément complet, est soumis le plus rapidement possible à l'équipe projet du musée. Cette livraison partielle et rapide permet d'introduire un peu d'agilité dans un processus soumis aux contraintes du forfait. La vérification en tant que telle démarre à la livraison complète de la prestation. Le chef de projet A&A Partners peut jouer un rôle de conseil dans cette phase de test : méthodologie, élaboration de cahier de test, réponses aux questions.

5 Les échanges de données un enjeu métier

5.1 Exports unilatéraux : publication, archives

5.1.1 Mise en ligne des collections muséales

Les solutions WebMuseo sont conçues et documentées pour optimiser les évolutions fonctionnelles et les échanges de données avec des systèmes tiers. Pour cela les ingénieurs d'A&A Partners adossent leurs développements sur des protocoles et des standards documentaires internationaux (OAI-PMH, SRU, Dublin Core, DarwinCore, etc.)

Via le moissonnage, les données de gestion sont recopiées dans l'espace de publication. Lors de cette opération les données sont optimisées pour la publication. Les informations sensibles ou confidentielles (régie des collections, adresse des anciens propriétaires, valeur d'assurance, etc.) ne sont pas moissonnées. A l'inverse des sous-rubriques de la base de données de gestion peuvent être autonomisées ou concaténées. La rubrique Créateur, sous-rubrique de Création/Exécution donne lieu à une rubrique donnant accès à l'index des auteurs. Elle est également utilisée pour composer le contenu de la rubrique de publication Création/Exécution dans un format texte riche. La publication n'est pas une réplique de la gestion, mais une extraction selon une mise en concordance (ou *mapping*[46]) des champs de la base de publication avec celle de la base de gestion.

[46] https://fr.wiktionary.org/wiki/mapping (consulté le 14/11/2017)

A chaque fois qu'un moissonnage est déclenché, la base de publication est mise à jour (nouvelles notices créées, notices existantes modifiées).

Chaque notice de l'espace de publication contient un lien vers sa notice de référence au sein de l'espace de gestion. La mise à jour ne concerne pas la suppression des notices. Les notices supprimées de l'espace de gestion ne seront pas supprimées de l'espace publication via le moissonnage.

5.1.2 Export enrichi pour archivage (Ressources EFEO au CINES)

Un dispositif d'export spécifique et sémantiquement enrichi a été mis au point pour l'archivage pérenne des ressources WebMuseo de l'EFEO (Ecole Française d'Extrême-Orient) vers l'OAIS (Open Archival Information System) du CINES (Centre Informatique National de l'Enseignement Supérieur).

5.2 Import de données et mises à jour : l'exemple d'e-ReColNat

5.2.1.1 Présentation du projet eReColNat

E-ReColNat est un projet national visant à «*faciliter l'exploitation des données de ressources biologiques issues des collections des différents muséums de sciences naturelles, et donc permettre, à terme, d'accélérer l'inventaire de la biodiversité mondiale*[47]». Afin de mener à bien cette ambition le projet met en place une base de données centralisée destinée à échanger des données avec les différentes institutions partenaires du projet.

Cette base de données a pour objectif d'unifier ces ressources dans un schéma simplifié assurant la normalisation et l'interopérabilité des données. Dans cette optique, les flux de données doivent être assurés vers la base centrale e_ReColNat et au départ de celle-ci vers les bases de données des partenaires du projet.

[47] https://www.recolnat.org/fr/le-programme (consulté le 12/11/2017)

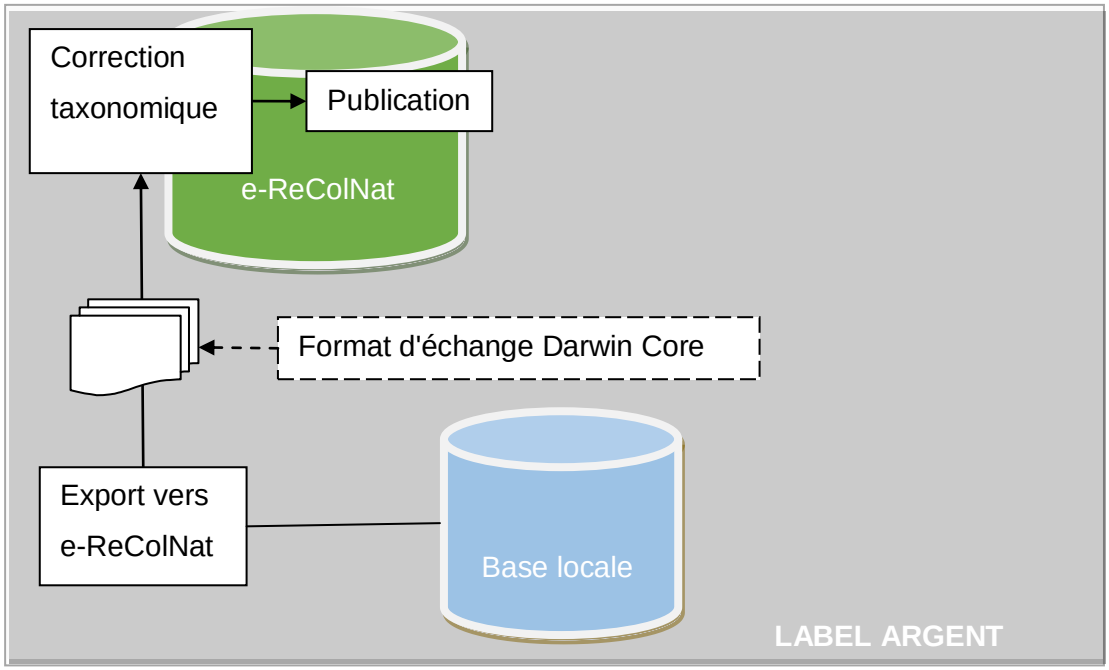

Figure 12 : Connecteur e-ReColNat label ARGENT : export des notices (JP)

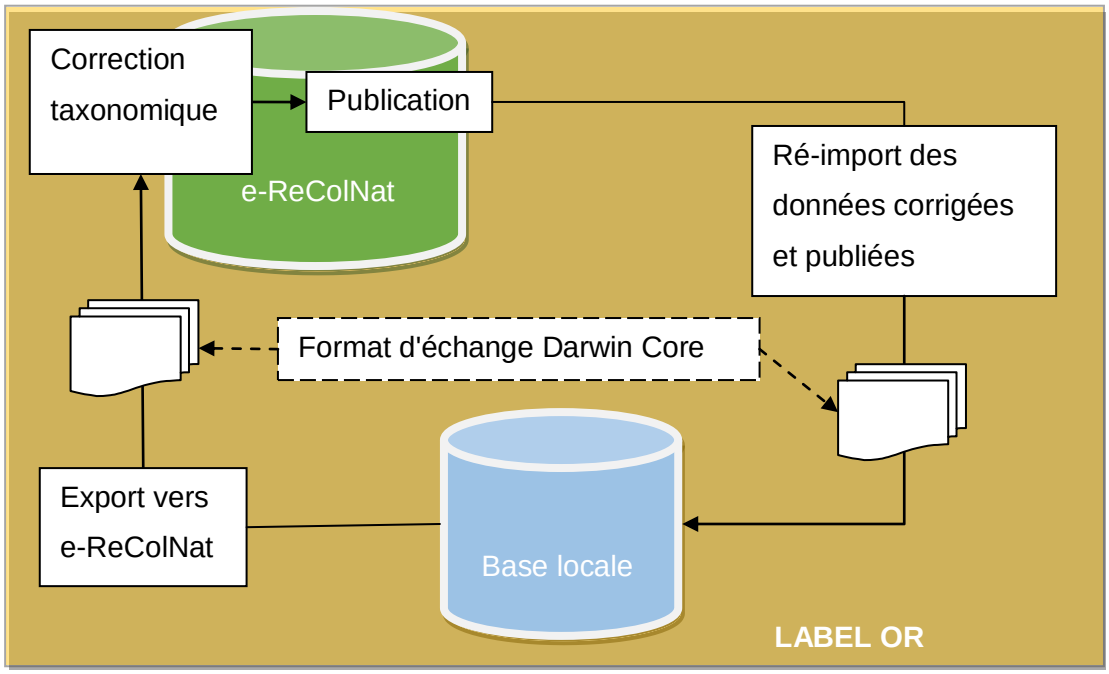

Figure 13 : Connecteur e-ReColNat label OR : export et ré-import des notices (JP)

Le muséum exporte un lot de notices de spécimens vers e-ReColNat sous la forme d'une archive Dublin Core. A la réception e-ReColNat identifie les primo-exports et

mises à jour en fonction de l'UUID (Universal Unique Identifier[48]) de chaque notice. Avant publication sur la base en ligne, les notices sont vérifiées et mises en cohérence au niveau taxonomique[49] par les gestionnaires de la base E-ReColNat. Les notices de spécimen publiées sont donc validées. Elles sont disponibles pour être ré-importées dans le logiciel de gestion du muséum producteur des données.

5.2.1.2 Périmètre du projet

L'objectif d'A&A Partners est d'adapter les structures actuelles de WebMuseo au format Darwin Core utilisé par le projet e-ReColNat afin de couvrir l'ensemble du spectre disciplinaire des sciences naturelles (botanique, zoologie, paléontologie, géologie, minéralogie et anthropologie). Cette adaptation doit permettre l'échange de fichiers tant à destination de la base e-ReColNat qu'en provenance de celle-ci (Label OR) mais également grâce au standard Darwin Core l'échange avec d'autres bases de données internationales telles que le GBIF[50] (Global Biodiversity Information Facility).

---

[48] https://fr.wikipedia.org/wiki/Universal_Unique_Identifier

[49] Nous optons pour l'usage du terme « taxonomie », plutôt que « taxinomie »
KEMBELLEC, Gérald, 2012. Bibliographies scientifiques : de la recherche d'informations à la production de documents normés [en ligne]. Université Paris VIII Vincennes-Saint Denis. [Consulté le 5 décembre 2017]. Disponible à l'adresse : https://tel.archives-ouvertes.fr/tel-00771553/document

[50] http://www.gbif.fr/

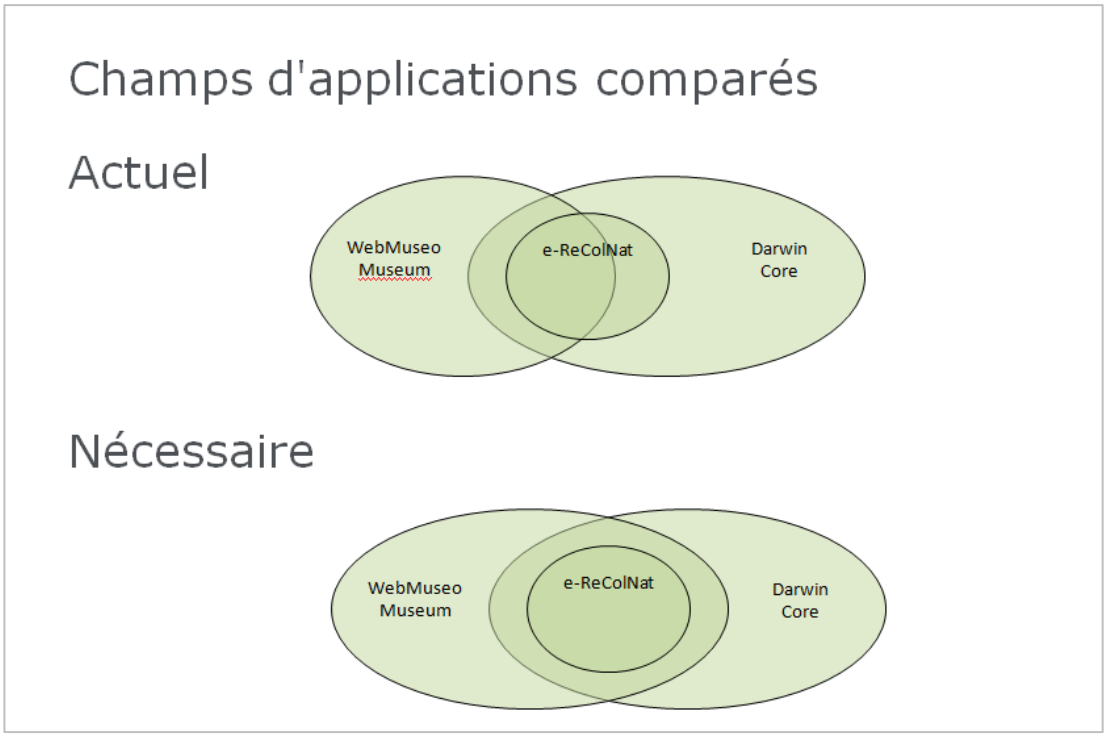

Figure 14 : Champs d'applications WebMuseo Gestion "Museum", e-ReColNat et Darwin Core (EP, A&A Partners)

5.2.1.3 Détermination et référentiel taxonomique

5.2.1.3.1 La détermination comme identification d'un spécimen

La détermination d'un spécimen correspond à son identification scientifique : rattachement à un taxon et relevé de déterminants complémentaires tels que le stade, la phase ou le sexe.

Elle est produite à une certaine date, par un ou plusieurs déterminateurs. L'identification peut être complète ou partielle. Le déterminateur peut indiquer la source de classification qu'il utilise et se référer à des publications scientifiques précises. Enfin, il peut arriver que le spécimen ait servi de référence ou ait été cité lors de la typification de l'espèce à laquelle il appartient. Son statut particulier (holotype, syntype, lectotype, etc.) est dans ce cas rappelé lors de la détermination. La détermination peut être révisée après nouvelle étude du spécimen. L'évolution du référentiel taxonomique ne justifie pas l'attribution d'une nouvelle détermination.

5.2.1.3.2 La taxonomie comme classement arborescent

Dans WebMuseo, un taxon ne contient pas directement l'intégralité de sa lignée. Chaque entrée du référentiel est seulement liée à l'entrée supérieure, quel que soit

son rang (Espèce, Genre, Famille, etc.). Ainsi la lignée taxonomique est composée dynamiquement à partir de l'enchainement des «parents».

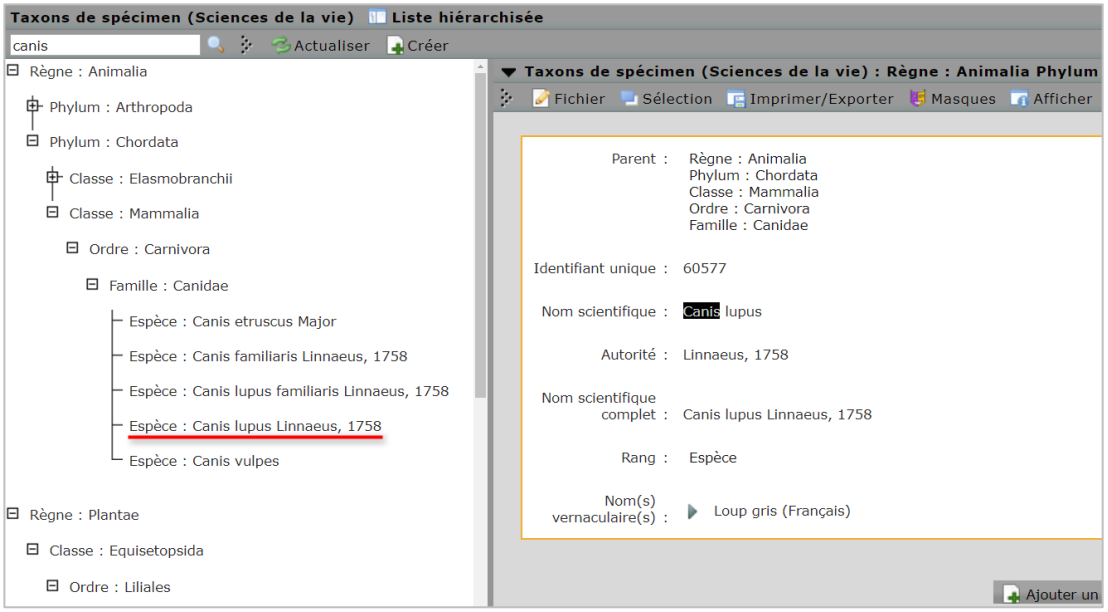

Figure 15 : Arbre taxonomique de l'espèce Canis lupus, Linnaeus, 1758 dans WebMuseo (JP)

Le rang permet de typer les taxons mais n'induit pas mécaniquement la hiérarchie qui est prise en charge par la relation parent/enfant. Dans l'exemple ci-dessus,

- le taxon de rang Espèce **"Canis Lupus Linnaeus, 1758**"
- **a pour parent** le taxon de rang Famille **"Canidae**"
- qui lui même **a pour parent** le taxon de rang Ordre "**Carnivora**"
- qui lui même **a pour parent** le taxon de rang Classe "**Mammalia**"
- qui lui-même **a pour parent** le taxon de rang Phylum "**Chordata**"
- qui lui même **a pour parent** le taxon racine de rang Règne "**Animalia**".

Règnes, Phylum, Classe, Ordre, Famille, Genres et Espèces sont stockés dans la même table "Taxons du spécimen" même s'ils sont typés par Rang. Cela permet une souplesse dans la précision de la détermination. Selon l'état des connaissances (en général ou propres au déterminateur) il est possible d'établir une détermination à un niveau plus ou moins générique (celui de la Famille par exemple).

5.2.1.4 L'historisation du référentiel taxonomique : principal enjeu de l'import

Après enrichissement du modèle de données de WebMuseo Gestion Museum, l'export vers e-ReColNat reste un export classique nécessitant la mise en correspondance vers un format d'échange. C'est la phase de ré-import des notices

corrigées qui amène à une réflexion approfondie sur la mise à jour du référentiel taxonomique.

La correction des notices au moment de leur publication sur la base nationale peut donc potentiellement concerner n'importe quel échelon taxonomique[51]. En effet, au gré du débat scientifique et des nouvelles découvertes, il arrive régulièrement que certaines espèces voient leur classification taxonomique évoluer, le plus souvent sur un ou deux degrés seulement. Dans ce cas comment actualiser les données dans l'arbre taxonomique de WebMuseo ?

5.2.1.5 Mise à jour de la taxonomie : temps réel ou historisation ?

En optant pour une solution "temps réél", le référentiel n'est pas corrigé mais complété. Il reflète à chaque instant la situation du moment sans conserver une mémoire d'une classification passée différente.

Lors d'une redescente d'e-ReColNat, une fiche Taxon peut provoquer la mise à jour d'une partie du référentiel sans qu'il en garde la trace. Les déterminations liées aux taxons reclassifiés héritent implicitement de cette mise à jour.

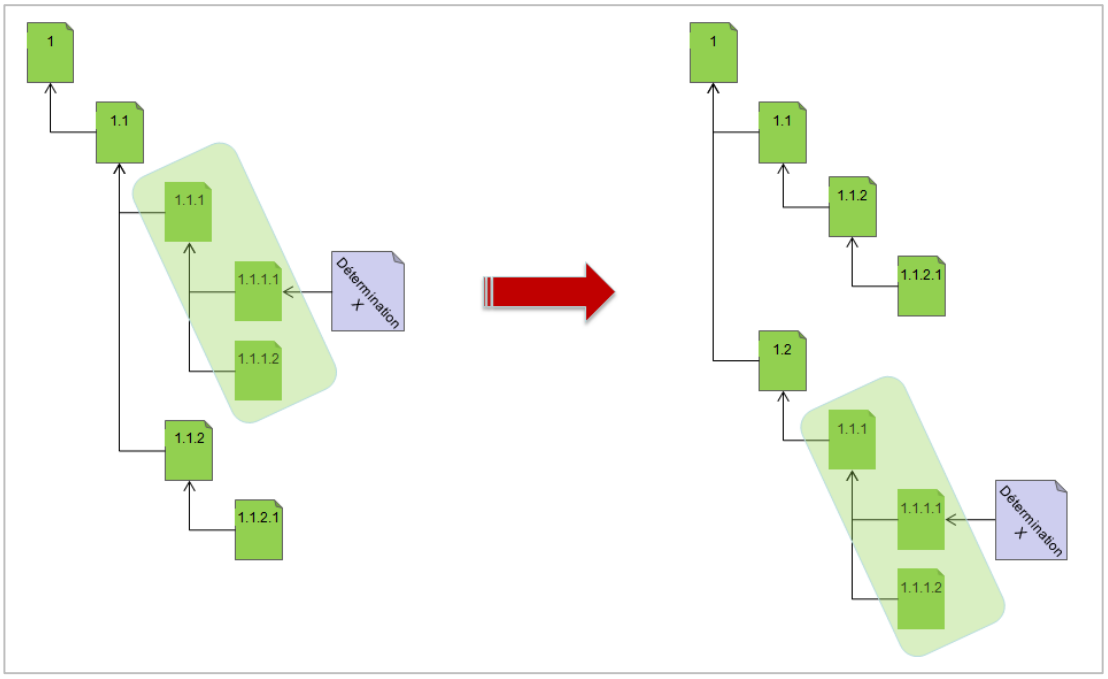

Figure 16 : Arbre taxonomique en temps réel (EP, A&A Partners)

Une détermination ancienne pourrait ainsi se retrouver renseignée avec une taxonomie récente (cf. Figure 16). Pour pallier à cette incohérence, il serait possible

[51]Dans le format Darwin Core la taxonomie est décomposée par rang, jusqu'à sept niveaux, dont le projet e-ReColNat ne retient que six.

d'envisager un référentiel qui conserverait structurellement la trace des anciennes classifications. Un lien "Ancien parent" (cf. Figure 17) permettrait de conserver la mémoire de l'évolution du référentiel taxonomique.

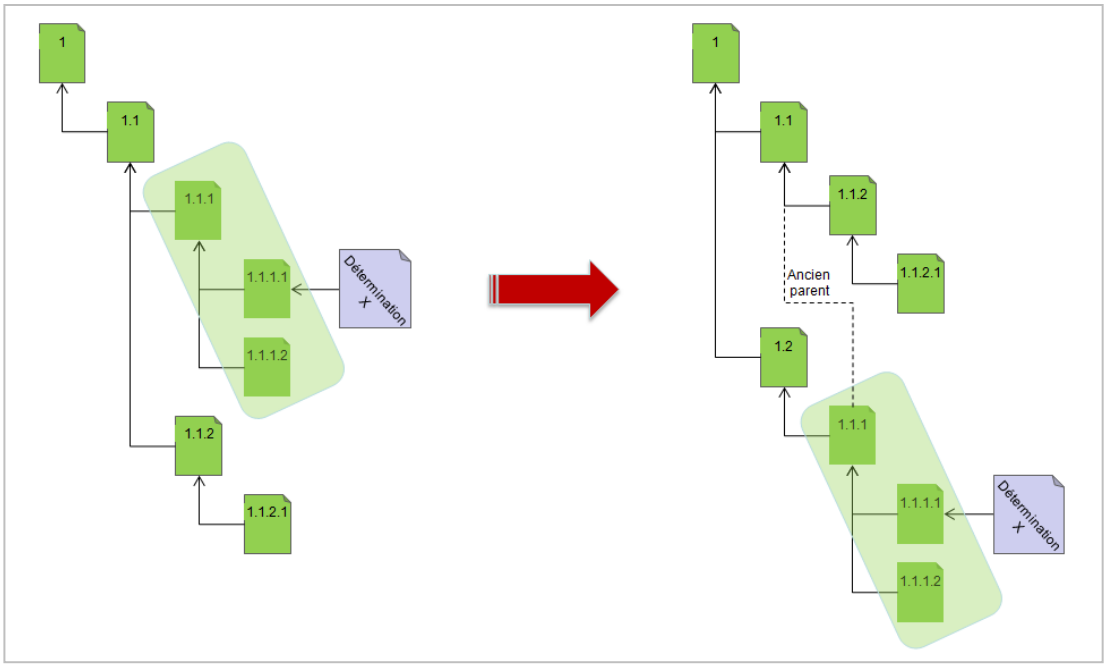

Figure 17 : Arbre taxonomique historisé (EP, A&A Partners)

Le taxon re-hiérarchisé conserve le lien vers son ancien parent. La sous-rubrique "Ancien parent" permet des recherches tenant compte des anciennes classifications. Cependant depuis la détermination, il faut penser à ouvrir le détail du référentiel taxonomique pour atteindre l'ancien taxon. D'autre part cette solution nécessite une gestion complexe du référentiel.

5.2.1.6 Une alternative efficiente : le verrouillage des déterminations

Comme la première solution de mise à jour, cette solution postule que le référentiel reflète à chaque instant la situation du moment. Plutôt que de faire porter l'effort d'historisation sur le référentiel, une solution alternative peut être de verrouiller la détermination. Celle-ci n’est plus simplement liée à une espèce, mais à chaque niveau de la lignée au moment de la détermination. Un script automatique permettrait de créer les liens à la création de la détermination de façon à conserver une photographie de l'arborescence à un instant T. Ainsi lors d’une redescente d’e-ReColNat, une fiche Taxon peut provoquer la mise à jour d’une partie du référentiel, sans qu’il en garde la trace. Les déterminations liées aux taxons reclassifiés héritent implicitement de cette mise à jour mais conservent en même temps le souvenir de l’ancienne lignée.

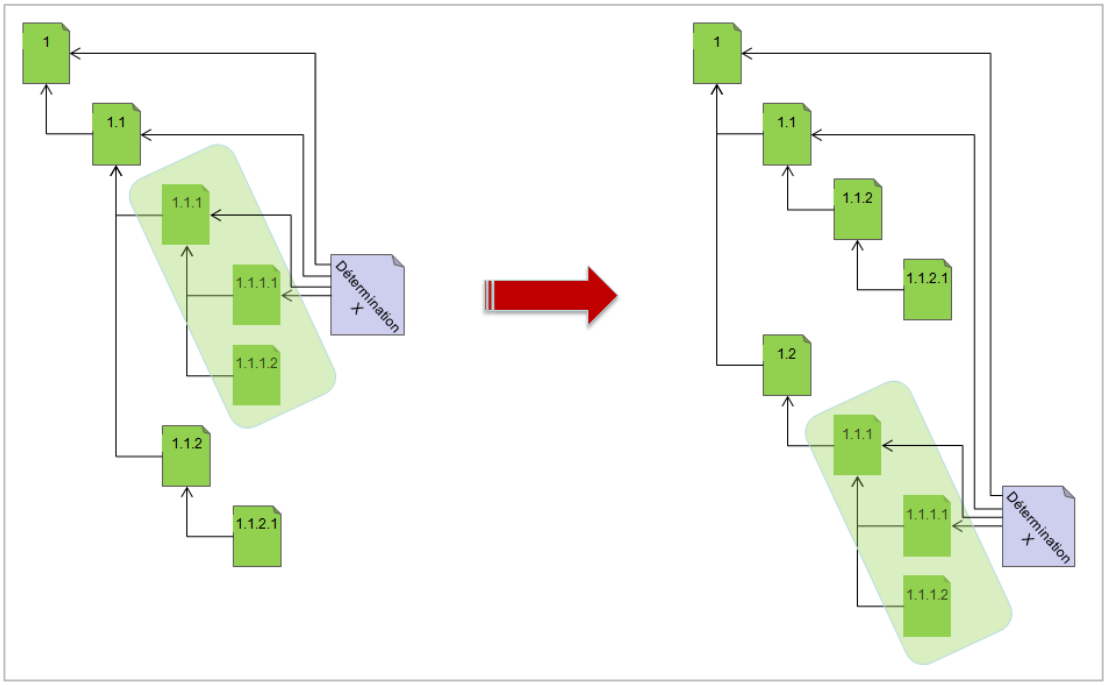

Figure 18 : Verrouillage de la détermination (EP, A&A Partners)

6 Compatibilité et propositions d'évolutions

6.1 Un existant favorable à l'exposition sémantique

6.1.1 Eléments favorisant la qualité des données

Les solutions Webmuseo sont mises au point et enrichies de manière à favoriser l'intégrité (cohérence, fiabilité, pertinence) des données qu'elles contiennent.

6.1.1.1 Cohérence du modèle relationnel

Les modèles de données relationnels WebMuseo (élaborés sur un SGBD PostgreSQL) sont conçus avec le souci qu'une information ne devrait pas être répétée dans plusieurs champs, mais liable à tous les contextes dans lesquelles elle est susceptible d'être utile. Ainsi les dates de naissance et de décès d'un auteur sont présentes dans sa fiche « Biographie » mais ne sont pas à saisir à nouveau dans la notice d'œuvre. Ces datations étant néanmoins utiles à la consultation de la notice, une option d'affichage des dates est paramétrable par le catalogueur depuis chaque fiche « Biographie ».

6.1.1.2 Aide contextuelle

Pour chaque champ de la solution, une « info des champs » accessible depuis une icône définit le type d'information attendu.

De plus, dans le modèle de gestion de collections muséales, les champs contrôlés par le SMF (Service des musées de France) se distinguent par une icône « ? » renvoyant à la page de la méthode d'informatisation (en ligne sur le site Joconde) détaillant les préconisations de saisie (informations gérées, mise en forme).

6.1.2 Compatibilité native du modèle WMG avec le CIDOC CRM

6.1.2.1 Le champ date riche

Les modèles de gestion de collections supportés par la plateforme WebMuseo proposent un type de champ "date riche" permettant d'exprimer une date quel que soit son type (date, siècle, période, intervalles, etc.) ou son degré de précision (siècle, quart de siècle, année, année/mois, année, mois, jours). Il est possible d'y associer, le cas échéant, un affixe d'incertitude, matérialisé par un point d'interrogation. Lors d'un filtre, c'est la valeur absolue de la requête qui sera prise en compte : l'ensemble des datations pertinentes sera retourné quel que soit son type. Par exemple une recherche avec l'option « est inclus(e) dans » sur les œuvres créées au 19e siècle prendra en compte toutes les œuvres créées entre 1800 et 1899, quel que soit le degré de précision de la datation.

Ce fonctionnement en valeur absolue offre de multiples possibilités d'exploitation en export. Par exemple dans le cadre d'un export vers la base Joconde, un champ date riche permet de générer les champs dates monotypes « Périodes » et « Dates ».

Le principe fondamental de ce format riche rejoint celui de l'entité 52_Laps de temps du CIDOC CRM. En effet, chaque date est interprétée quelle que soit son expression comme un intervalle entre deux bornes temporelles. Une date en jours/mois/année est un intervalle de 24 heures. Ce choix technique rejoint la définition de la classe E52 Laps de temps du CIDOC CRM :

> *« [La classe E52 Laps de temps] regroupe les étendues temporelles abstraites, au sens où l'entend la physique galiléenne, qui ont un début, une fin, et une durée. E52 Laps de Temps n'a pas d'autre connotation sémantique. »*
> [15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.43]

Quel que soit le degré de précision de la datation, la date enregistrée selon de le format riche de WebMuseo sera précisément bornée dans le temps, ce qui est plus restrictif que les attendus de l'entité 52_Laps de temps :

> *« Notre connaissance de l'histoire étant imparfaite, on peut au mieux considérer les instances de E52 Laps de Temps comme des approximations des laps de temps réels au cours desquels se sont*

déroulées les entités temporelles. Les propriétés de E52 Laps de Temps visent à permettre d'exprimer précisément ces approximations. Un cas extrême d'approximation pourrait, par exemple, définir une instance de E52 Laps de Temps dont le début, la fin et la durée ne sont pas connues. Si elle était utilisée comme instance de E52 Laps de Temps commune à deux événements, elle indiquerait malgré tout qu'ils ont eu lieu simultanément, même si l'on ne savait rien d'autre. »
[15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.43]

6.1.2.2 Les cas d'usage correspondant au E10_Transfert de la Garde

Le concept de transfert de garde de l'entité E10 permet d'exprimer les données administrées depuis les outils de gestion des sorties d'œuvres de WebMuseo. Ce module permet au musée de gérer les prêts qu'il consent à un tiers pour une exposition ou une restauration en atelier par exemple.

Lorsque le musée acquiert une œuvre (par don, legs ou achat) ou lorsqu'il reçoit un bien en dépôt, le transfert de garde est implicite. Il pourrait être exprimé dans la sémantique CIDOC au moment d'un export vers un entrepôt RDF (ou triplestore).

La notion de transfert de garde est également adaptée à la notion de vol d'un bien muséal selon la législation française : l'objet volé reste de façon imprescriptible la propriété de l'institution muséale. Le concept de transfert de garde appliqué à ce cas de figure permet d'exprimer que l'objet n'est plus sous la responsabilité de son propriétaire.

De plus, comme le modèle de référence du CIDOC, WebMuseo Gestion distingue la notion de transfert de garde de celle de re-localisation. En effet, dans WebMuseo les localisations successives peuvent être historisées par objet. Ainsi dès la deuxième localisation il serait possible de déclarer un E9_Déplacement.

Tout objet peut servir de cadre de référence pour déterminer un E53 Lieu. Le modèle prévoit la notion de «section» d'un E19 Objet Matériel comme moyen valable de déterminer un E53 Lieu.

6.1.2.3 Les Biographies

La table Biographies de WebMuseo permet de décrire aussi bien une personne physique qu'une personne morale. En ce sens elle couvre un périmètre en adéquation avec l'entité E39_Agent :

« Cette classe regroupe les personnes, soit individuellement soit en groupe, qui ont la faculté d'exécuter des actions intentionnelles dont elles peuvent être tenues responsables.

L'ontologie ne tente pas de modéliser les actions non intentionnelles de tels agents. Il convient que les personnes individuelles soient documentées comme des instances de E21 Personne, et que les groupes soient documentés comme des instances soit deE74 Groupe, soit de sa sous-classe E40 Personne Morale. »
[15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.37]

Cette définition conceptuelle plaide pour un maintien d'une table Biographie distincte de celle des emplacements, même si elle peut ressembler à une redondance dans le cas des personnes morales. Ainsi il est justifié qu'un musée soit référencé selon même intitulé en tant que Personne morale en Biographies et en tant que lieu au sein de la table des Emplacements.

6.1.3 Notices d'œuvres et représentation des connaissances

Dans la méthode d'inventaire informatisée préconisée par le SMF, on observe une distinction nette entre ce qui relève de la forme matérielle de l'œuvre (description, techniques de fabrication, matériaux, inscriptions, dimensions) qui sont présentées sous forme de données objectives, et ce qu'elle exprime (sa représentation, sa destination) qui offre plus de place à l'interprétation du catalogueur. On retrouve ici le rapport consubstantiel fond/forme, support/message décrit par Bruno Bachimont :

« Un contenu est donc matériel et possède les propriétés physiques de son support : un contenu immatériel n'existe pas. Mais le contenu ne se réduit pas à son support [..] un objet [...] véhicule [...] une signification qui s'adresse à un interprète. »
[28, BACHIMONT 2017, p.22]

Le CIDOC CRM modélise non pas l'objet, mais les connaissances produites à son sujet, quelle que soit la façon dont elles sont exprimées. Il s'attache à séparer la forme (textuelle) de sa signification selon la logique linguistique de séparation lemme/lexème :

- *« le lemme est l'unité autonome du lexique au plan sémantique : un même lemme contient un ou plusieurs lexèmes, lesquels dépendent du contexte d'emploi ;*

- *le lexème (ou unité lexicale) est le constituant lexical (prononcé ou écrit), autonome ou non, d'un lemme ».*[52]

Le contenu des rubriques d'une base de données musées doit être exprimé comme une suite de caractères (lexème / E41 Appellations) renvoyant à une réalité conceptuelle (lemme / E55 Types). Un lemme peut être exprimé selon plusieurs lexèmes.

> *« E55 Type : Cette classe regroupe des concepts désignés par des termes provenant de thésaurus et de vocabulaires contrôlés, utilisés pour caractériser et classifier des instances de classes. Les instances de E55 Type représentent des concepts, par opposition aux instances de E41 Appellation qui sont utilisées pour nommer des instances de classes. »*
> [15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.46]

6.2 Principales notions à modéliser

6.2.1 Précisions sur la modélisation

Les flèches en rouge représentent la relation « est une sous-classe de ». Celles en noir représentent les propriétés.

Les liens représentés tiennent compte des héritages de propriété. Ainsi, un objet fabriqué (E22) peut avoir des dimensions (P43) en tant que chose (E70).

La modélisation part du principe que les collections documentées sous WebMuseo Gestion sont par définition des objets fabriqués (E22). En fonction du type de collection il sera possible de reporter le modèle sur une collection d'objets biologiques (E20) en supprimant les propriétés liées à la dimension conceptuelle, notamment la propriété « P129 a pour sujet » et appliquant des ajouts pertinents.

Les propriétés terminées par le suffixe i sont des propriétés inversées.

6.2.2 Caractéristiques générales des objets

6.2.2.1 Identification

Comme toute entité, un objet fabriqué peut être identifié (P1) par une appellation (E41). En tant que chose fabriquée (E71) un objet fabriqué peut avoir un titre (E35).

---

[52] https://fr.wikipedia.org/wiki/Lex%C3%A8me_(linguistique)

En tant que chose matérielle fabriquée (E24), un objet matériel fabriqué peut être support d'une image (E38).

> *« Cette classe regroupe les distributions de formes, de tons et de couleurs que l'on peut trouver sur des surfaces telles que des photographies, des peintures, des gravures et des sculptures ou directement sur des médias électroniques.*
>
> *Le degré auquel les variations de forme et de couleur affectent l'identité d'une instance de E38 Image dépend d'une finalité donnée. On peut dire que l'original de la Joconde au Louvre porte la même instance de E38 Image que des reproductions sous forme de diapositives, cartes postales, affiches ou T-shirts, bien que tous ces objets puissent être différents par la taille et le support et présenter des variations de ton et de couleur. Les images d'un «jeu des sept différences» ne sont pas les mêmes, en considération de leur contexte, quelque similaires qu'elles paraissent à première vue. »*
> [15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.37]

Les numéros de référence de l'objet peuvent être portés par l'entité « E42 Identifiant ». Les identifiants peuvent être typés (Numéro d'inventaire, Numéro de dépôt, Anciens numéro, Numéro de catalogue, etc.) via « E55 type ».

6.2.2.2 Mesures

L'objet fabriqué est directement relié à l'entité autonome dimension via la propriété P43 « a pour dimension ». Le chemin plus complet composé des triplets [*E22 objet fabriqué, P39i a été mesuré par, E16 Mesurage]* et *[E16 Mesurage, P40 a relevé comme dimension, E54 Dimension]* permettrait d'exprimer des précisions sur l'action de prise des mesures. Cette possibilité n'est pas modélisée car WebMuseo Gestion ne va jusqu'à ce niveau de détail dans sa version standard.

6.2.2.3 Matières

En tant que chose matérielle (E18) un objet fabriqué peut référencer un ou plusieurs matériaux (E57).

6.2.2.4 Concepts portés par l'objet : l'exemple de l'Iconographie/représentation

La propriété « P129 a pour sujet » permet de rendre compte de la représentation portée par une œuvre.

En tant qu'objet propositionnel, (E89) l'objet fabriqué peut avoir n'importe quelle entité pour sujet. Pour représenter les caractéristiques générales d'un objet de

collection le schéma modélise la possibilité de référencer des personnes (physiques ou morales) et des éléments du thésaurus Garnier rattaché à l'entité type (E55).

NB : Le modèle CIDOC CRM permettrait également de modéliser l'utilisation/destination d'un objet de collection ethnographique.

6.2.2.5 Objets composites

En tant qu'objet matériel (E19) l'objet fabriqué peut être composé de plusieurs parties. Cette caractéristique peut être prise en charge par le triplet *[E22 Objet fabriqué, P57 a pour nombre de parties E60 nombre]*.

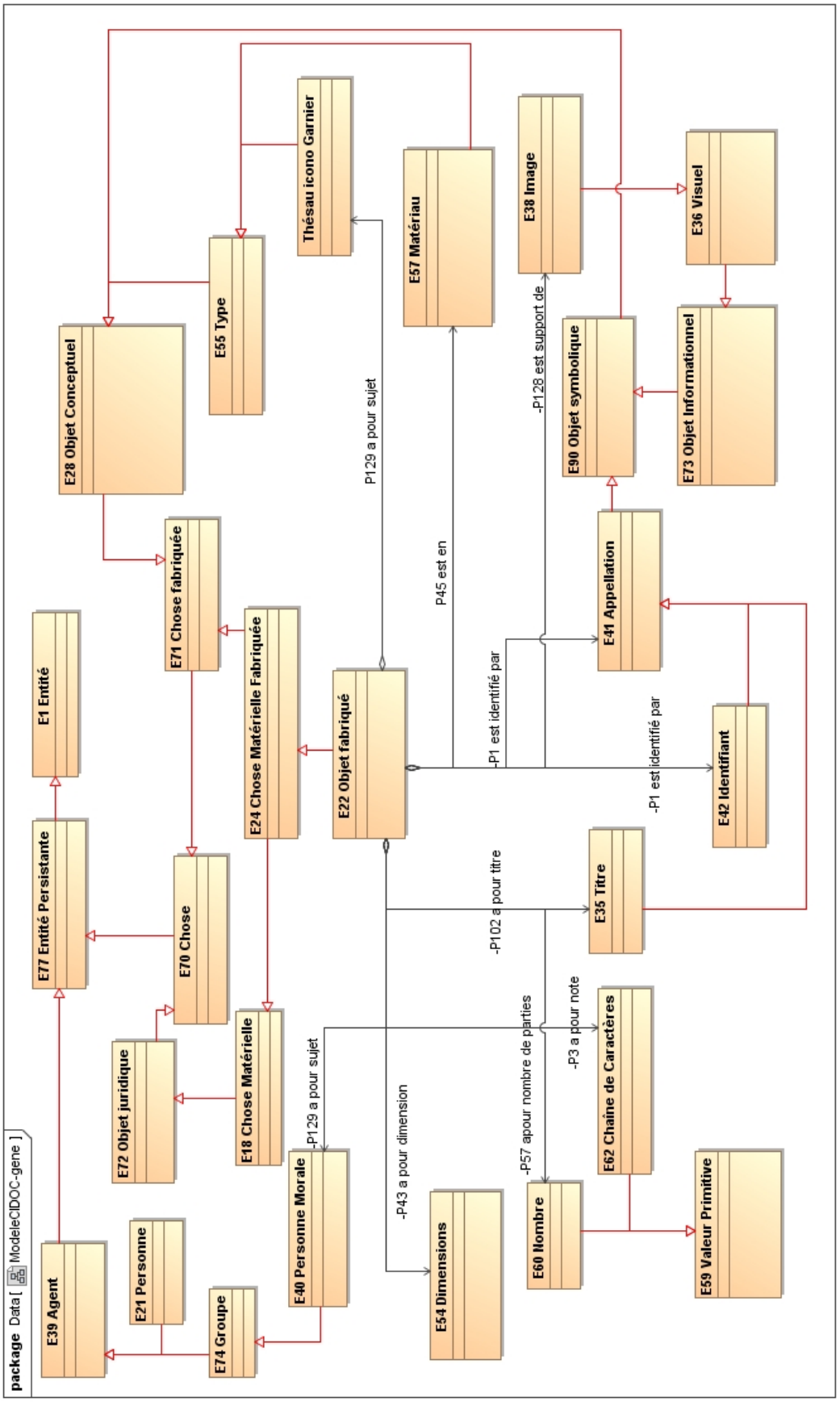

Figure 19 : Modélisation des caractéristiques générales des objets (JP)

6.2.3 Etapes de la vie d'un objet : de sa conception à sa conservation patrimoniale

6.2.3.1 Création / Exécution

Dans le modèle CIDOC CRM, le début d'existence d'un objet fabriqué est le fruit d'une modification matérielle (E11) et/ou d'une activité de création (E65).

Afin de rester au plus proche de la structuration des données source, le graphe est organisé autour de l'entité « E65 Création » qui permet d'agréger :

- le ou les auteurs *[E65 Création, P14 a été réalisé par, E39 Acteur]*
- le ou les événements associés *[E65 Création, P117 survient pendant, E5 Evénement)*
- le ou les laps de temps de création *[E65 Création, P4 a pour laps de temps, E52 Laps de temps]* et par extension la date de création (Cf. Figure 21 : Modélisation possible d'un laps de temps)
- le lieu de création *[E65 Création, P7 a eu lieu dans, E53 Lieu]*
- la ou les technique(s) utilisée(s) *[E65 Création, P32 a employé comme technique générique, Thésau Techniques (sous classe de E55 type)]*

Dans WebMuseo Gestion les techniques sont associées aux matériaux. Les deux référentiels sont néanmoins stockés dans des tables différentes et exploités dans des sous-rubriques spécifiques. Il sera donc possible de les différencier lors d'un export selon un schéma CIDOC CRM.

L'activité de l'activité de modification (E11) pourrait être également modélisée le cas échéant. Par exemple, il serait imaginable de documenter au sein d'une solution photothèque les retouches graphiques d'une image en tant que modification d'un fichier original.

C'est à l'activité de création que doivent être liées les techniques employées.

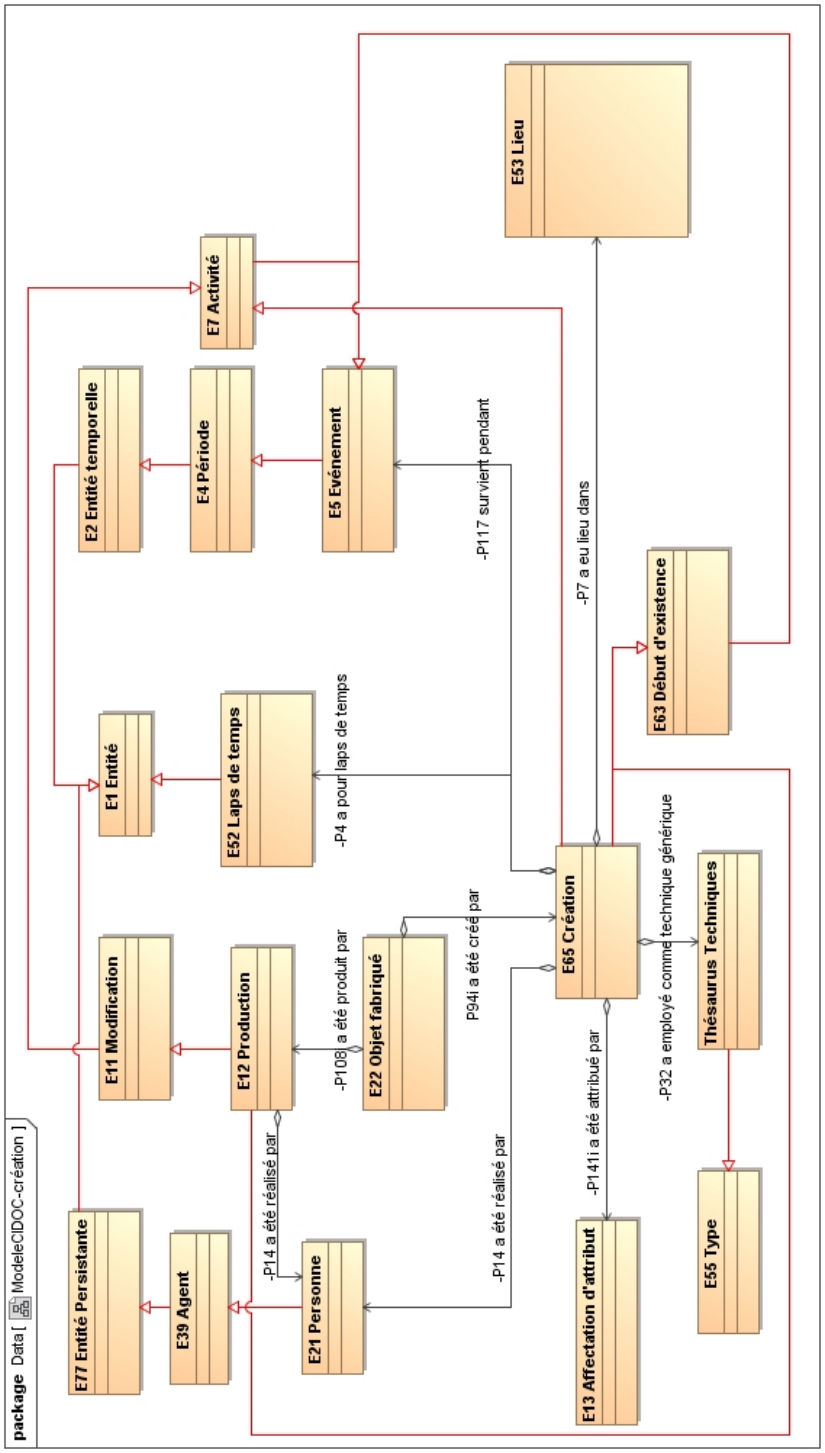

Figure 20 : Modélisation de la création de l'œuvre (JP)

Dans le système CIDOC CRM, un laps de temps peut être identifié par une date et calculé en durée maximum et minimum, via « E54 Dimension » et/ou « E61 Primitive Temporelle ». Dans le cas d'une donnée issue du champ date de création de WebMuseo la durée minimum sera par défaut égale à la durée maximum.

Il ne serait possible de distinguer deux durées au sein d'un laps de temps que si des qualificatifs exprimant les notions de couverture minimum et maximales étaient associés aux dates de création ce qui n'est pas l'usage.

En effet les dates de création ne sont généralement pas exprimées selon des durées maximale ou minimale. En revanche, il est souvent fait mention d'indication d'incertitude. Ainsi, les qualificatifs d'incertitude associables à la date riche de WebMuseo (point d'interrogation, vers, circa, etc.) pourraient être repris et dédoublés pour qualifier le début et la fin du laps de temps.

Lorsque plusieurs dates de création sont renseignées pour rendre compte des étapes de création et de fabrication d'une œuvre (une estampe ou une sculpture par exemple), le laps de temps commence par la date la plus ancienne et se termine par la plus récente. En effet, l'entité temporelle « E65 Création » ne peut être liée qu'à un seul laps de temps qui correspond à sa durée théorique réelle.

> *« La Conférence de Yalta (E7) a pour laps de temps le laps de temps correspondant à la Conférence de Yalta (E52) ».*
> [15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.64]

Le laps de temps peut cependant s'insérer ou être inséré (P86 ou P86i) dans d'autres laps de temps et peut être identifié par plusieurs « E49 Appellation de temps » et par conséquent par plusieurs « E50 Date ».

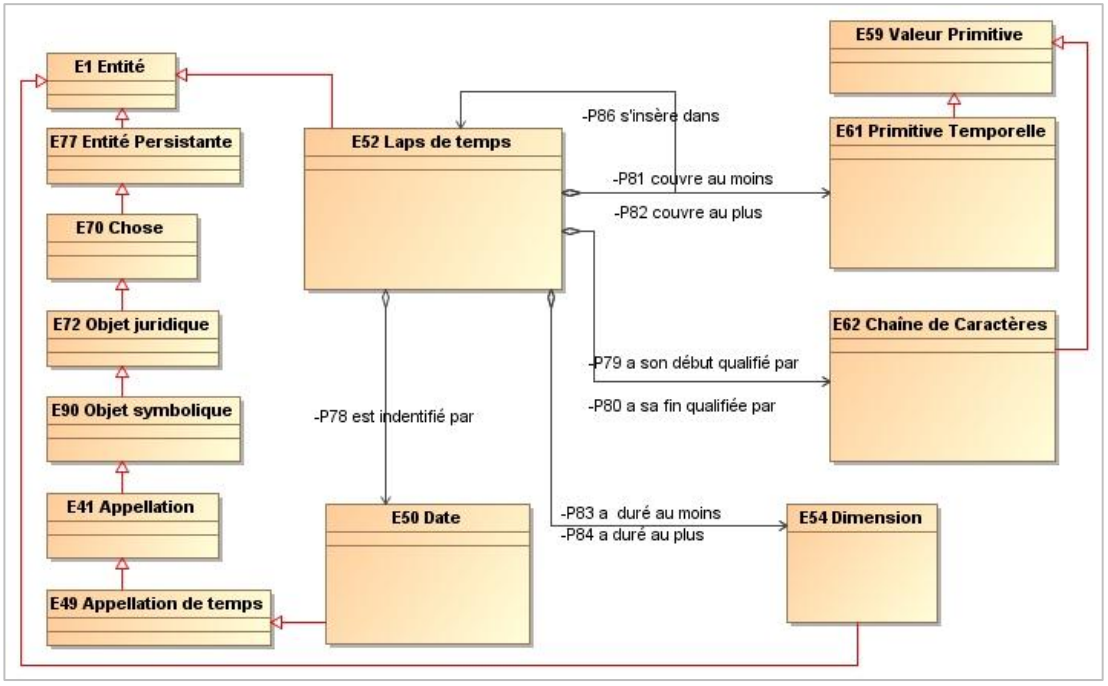

Figure 21 : Modélisation possible d'un laps de temps (JP)

6.2.3.2 Acquisition et Anciennes appartenances

Le modèle WebMuseo Gestion est destiné à documenter les collections d'un établissement muséal. La notion d'acquisition est réservée à l'entrée du bien dans la collection au terme d'un processus en trois temps :

1. acquisition par l'institution propriétaire (une ville par exemple),
2. affectation à un établissement (musée),
3. inscription au registre des collections (par le conservateur).

Dans les faits, l'acquisition et l'affectation sont concomitantes. Une fois inscrit au registre d'inventaire, le bien acquiert un statut exceptionnel de propriété et devient inaliénable et imprescriptible.

L'historique des précédents transferts de propriété du bien peut être documenté dans une rubrique complémentaire « Anciennes appartenances ». Celle-ci recouvre en partie l'acquisition lorsqu'elle référence le dernier ancien propriétaire qui est dans le contexte de l'acquisition muséale le « Donateur, Testateur, Vendeur ».

Du point de vue du CIDOC CRM, l'acquisition est une activité de transfert de propriété qui permet notamment de retracer l'historique des anciens propriétaires et donc de représenter la rubrique musée « Anciennes appartenances ». Pour caractériser la dernière acquisition musée, il est possible d'utiliser le thésaurus SMF des modes d'acquisition comme sous-classe de « E55 Type ».

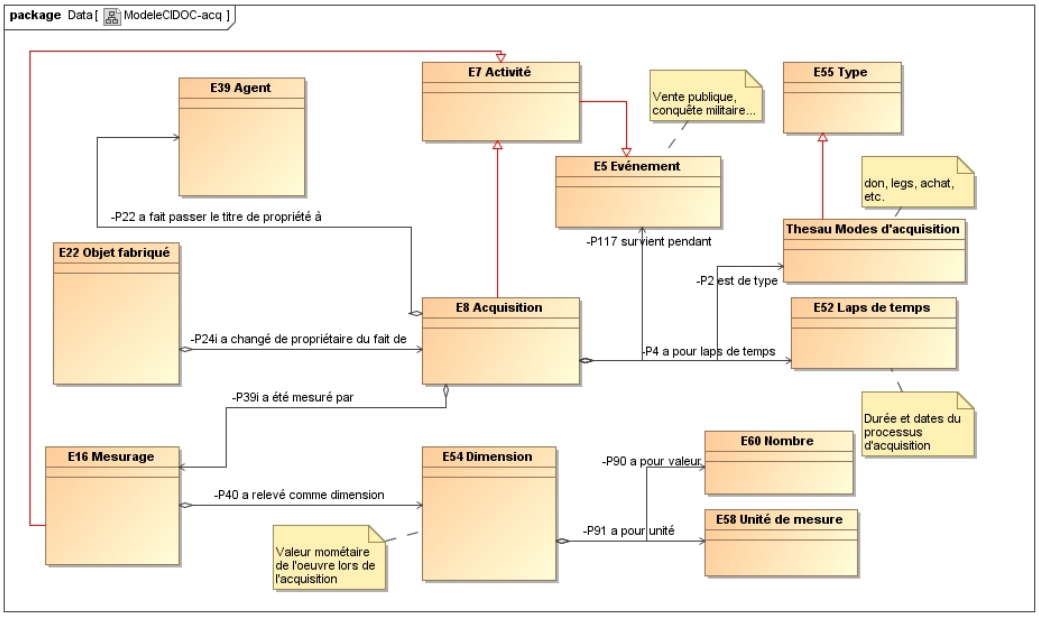

Figure 22 : Modélisation des acquisitions (JP)

6.2.3.3 Entrée et gestion physique dans l'institution muséale

Dans le système CIDOC CRM les actions de régie sont, en tant qu'activité (E7) localisables (E54), datables (E55) et liables à un auteur et/ou des participants (E39 Agent). Ces caractéristiques sont partagées avec les fiches de régie de WebMuseo gestion (Localisations, Constats d'état, Restauration, Fiches de sorties).

6.2.3.3.1 Transfert de garde

Le concept de transfert de garde du CIDOC CRM couvre les activités de gestion muséale d'entrée physique d'un bien dans la collection (couplées avec celles d'acquisition), de réception d'un bien en prêt ou en dépôt, d'envoi d'un bien en prêt ou en dépôt. Les types de transfert (dépôt, prêt, livraison d'acquisition) peuvent être utilisés via les types (E55).

L'entité « Transfert de garde E10 » permet de désigner un destinataire de la garde (« P29 a confié la garde ») et, en tant qu'événement, de référencer des participants.

Un transfert de garde peut intervenir à l'occasion d'une exposition. La description de l'exposition (son responsable, son lieu, ses dates) peut être prise en charge par l'entité « E5 Evénement ». La distinction entre dates de sorties et dates de l'exposition possible proposée dans WebMuseo Gestion ainsi peut être exprimée.

Le transfert de garde est souvent accompagné d'une assurance spécifique couvrant le laps de temps durant lequel l'objet emprunté sera sous la responsabilité d'un tiers.

La modélisation ne va pas plus loin que la granularité du modèle de gestion WebMuseo. Il serait en effet également possible de structurer en CIDOC CRM spécifiquement le début et la fin de la garde.

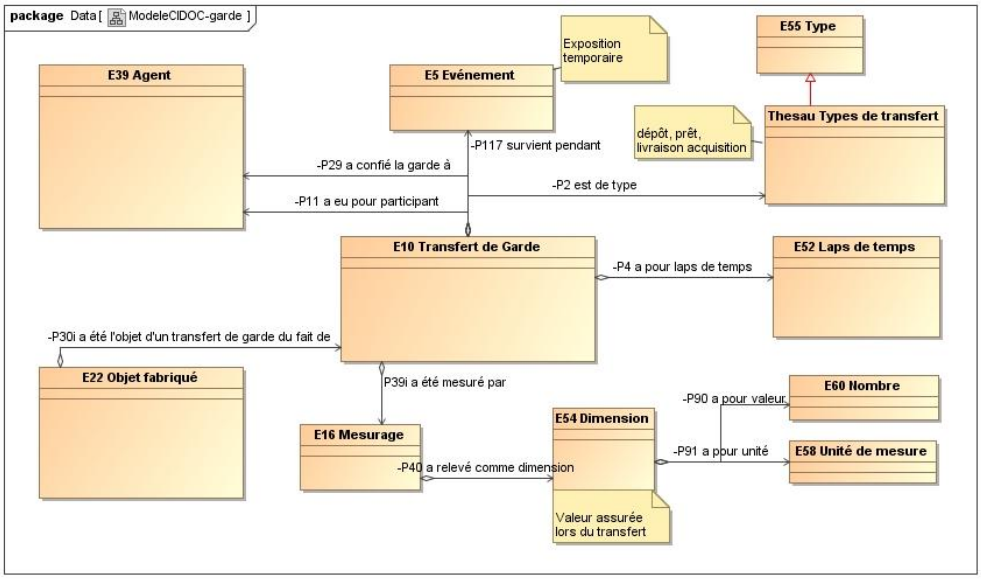

Figure 23 : Modélisation des transferts de garde (JP)

6.2.3.3.2 Localisation

La Figure 24 met en évidence les deux types de localisation attendus d'un musée et qui sont prises en charge par le modèle WebMuseo Gestion : localisation permanente et historique de déplacements.

La notion de localisation actuelle du CIDOC CRM ne garantit pas la présence effective du bien à l'emplacement indiqué. L'information « peut dépendre du moment où la propriété a été instanciée[53] ». Elle correspond à la rubrique calculée de WebMuseo gestion « Dernière localisation ».

Dans WebMuseo Gestion, le bien (E22) peut être directement localisé sans faire l'objet d'un constat de localisation daté et signé. L'emplacement est lié en tant

[53] Voir note d'application de la propriété P55, p80 dans *NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014*

qu'emplacement physique réservé à l'objet dans l'institution propriétaire (salle d'exposition ou réserves) sans notion de confirmation de sa présence.

> *« La localisation permanente] indique le E53 Lieu actuellement réservé à un objet, tel que la localisation de stockage permanent ou une localisation d'exposition permanente. L'objet peut être temporairement retiré de sa localisation permanente, par exemple lorsqu'il est utilisé dans des expositions temporaires ou prêté à une autre institution. Il peut arriver qu'un objet ne se trouve en fait jamais dans sa localisation permanente. »*
> [15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.80]

WebMuseo Gestion permet de rendre compte à tout moment de la localisation d'un objet via une fiche de localisation qui fonctionne comme un constat de localisation, daté et signé. Une nouvelle localisation peut donc confirmer la localisation précédente et l'actualiser. Le CIDOC CRM n'historise pas les localisations mais les déplacements (E9). Il permet donc d'exprimer le différentiel entre deux localisations via les deux propriétés « P27 a été déplacé de » et « P26 a été déplacé vers ».

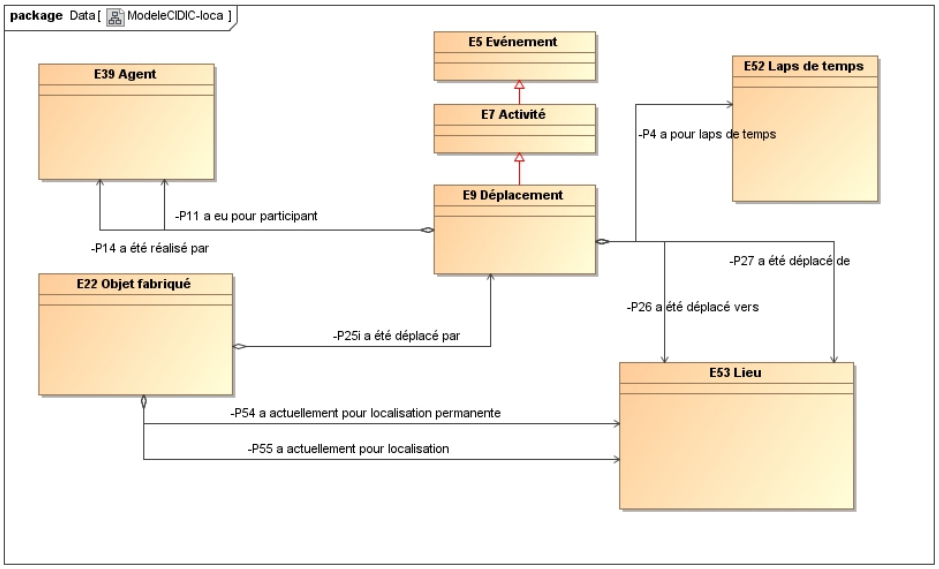

Figure 24 : Modélisation de la localisation (JP)

6.2.3.3.3 Constats d'état

Dans un WebMuseo une fiche de constat d'état permet de faire le lien entre une notice et un aspect général (bon état, état moyen, etc.) à un moment donné (date de constat), à un endroit (lieu du constat) et selon une personne (auteur du constat). Si le constat n'est pas réalisé en direct mais d'après un document, il est possible de l'attacher en fichier joint.

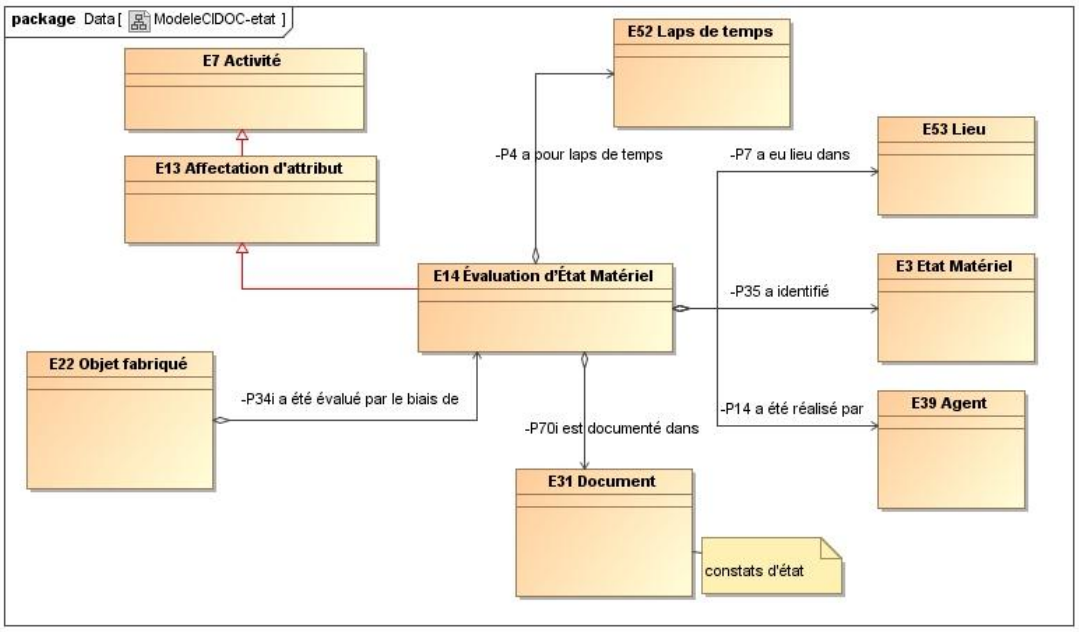

Figure 25 : Modélisation des constats d'états (JP)

6.2.4 Documentation

WebMuseo Gestion permet de décrire de façon assez détaillée les fonds documentaires du musée. Chaque fiche de documentation (ouvrage bibliographique, archive, photographie, enregistrement sonore, vidéo) peut être associée à la notice d'œuvre. La modélisation proposée ici prend le parti de traiter ces informations comme étant des références annexes aux objets de la collection, sachant qu'en cas de besoin, le CIDOC CRM permettrait d'aller plus loin dans le détail de chaque entité documentaire. En effet, les propriétés décrites plus haut au sujet des objets de la collection peuvent s'appliquer aux documents et images.

6.2.4.1 Bibliographie, archives et médias

L'entité « E31 Document » permet de décrire tout type de fonds documentaire quel que soit leur mode d'expression : écrit pour les ouvrages et articles, image, sons, audiovisuel.

> *« Cette classe regroupe les articles immatériels identifiables qui énoncent des propositions à propos de la réalité.*
>
> *Ces propositions peuvent être exprimées par du texte, du graphisme, des images, des audiogrammes, des vidéogrammes ou d'autres moyens semblables. Les bases de données documentaires sont considérées comme un cas particulier de E31 Document. »*
> [15, NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel 2014, p.34]

En plus de représenter les objets de la collection l'entité « E22 Objet fabriqué » peut en tant qu'héritier de « E19 Objet Physique » être le support du contenu d'un document via la propriété « P128 a pour support ».

6.2.4.2 Evénements et Expositions

Dans WebMuseo Gestion, les événements auxquels l'objet participe en tant qu'élément de collection muséale sont caractérisés comme des expositions ou des campagnes de restauration et sont intégrés au module de régie des collections. Les événements représentés ou évoqués par l'objet sont documentés dans une table spécifique. Toutes ces dimensions peuvent être prises en charge par l'entité « E5 Evénement » en s'appuyant sur les propriétés « P62 figure » et « P12i a été présent à ».

6.2.4.3 Médias

En plus de représenter les objets de la collection l'entité « E22 Objet fabriqué » peut, en tant qu'héritier indirect de « E18 Chose matérielle fabriquée », être le support d'une image (tirage photographique par exemple) via la propriété - « P65i est présenté par ».

La propriété « P62i est figuré sur » permet d'indiquer que l'image est une illustration de l'œuvre.

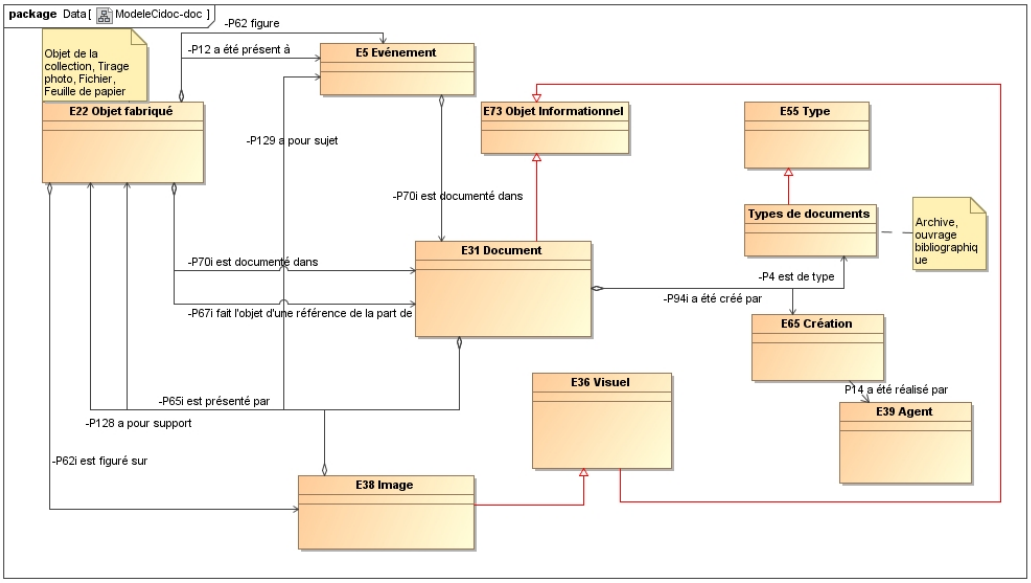

Figure 26 : Modélisation de la documentation associée (JP)

6.3 Limites de la modélisation en CIDOC CRM

Le modèle relationnel WebMuseo Gestion est facilement transposable dans la sémantique du CIDOC CRM, rendant possible l'agrégation des contenus structurés. La sémantique implicite des rubriques d'indexation peut être encapsulée avec les valeurs. Cependant les rubriques contenant du texte « libre » (Description, Précisions sur la représentation, etc.) devraient être associées à des dispositifs de traitement automatisé du langage naturel[54] pour être sémantiquement exploitable.

6.4 Enseignements vis-à-vis de la gestion du cycle de vie

Si le cycle de vie de l'objet peut être exprimé selon la sémantique du CIDOC CRM ce dernier pourrait également permettre de conserver une trace de l'évolution de la connaissance de l'objet. Les interprétations successives ou parallèles au sujet d'un objet pourraient être archivées en s'appuyant sur les entités héritant de la classe « E7 Activités ».

6.5 Evolution vers des fonctionnalités d'éditorialisation

Approfondir la dimension sémantique n'implique pas tant des considérations techniques que des opportunités d'évolutions fonctionnelles. Comme nous l'avons vu, le rôle d'A&A Partners sera de fournir les données au(x) format(s) RDF plus que de produire directement des solutions sémantiques. Les fonctionnalités d'un outil de gestion métier ne couvriront par définition jamais l'étendue des possibilités du CIDOC CRM.

Une caractéristique du modèle de référence et du web de données liées en général pourrait être renforcée : la dimension interprétative et donc relative des assertions. Au niveau juridique l'éditorialisation des données se situe à la frontière entre le droit des bases de données et le droit d'auteur.

[54] https://fr.wikipedia.org/wiki/Traitement_automatique_du_langage_naturel (consulté le 6/12/2017)

Conclusion

Le CIDOC CRM naît d'un souci ancien d'interopérabilité entre les documentations des collections culturelles de natures et d'approches différentes : collections de musées, de muséums, de bibliothèques et d'archives. La création de ponts systématiques entre ces gisements d'information apparaît très tôt comme une chance pour leur valorisation et enrichissement mutuels.

Cependant, cet objectif de bien informationnel culturel et commun suppose une dimension coopérative qui est souvent entravée par la compétition existant entre les institutions culturelles[55]. Les attributs de l'interopérabilité (standards d'échange, thésaurus, modèles de représentation des connaissances) constituent en soi un enjeu stratégique. Comme nous l'avons vu dans le §2.1 le SAAM (Smithsonian Museum of American Art) a d'abord constitué son propre modèle sémantique sur la base de plusieurs standards. La priorité du SAAM dans le cadre du web sémantique était d'être compatible avec les acteurs les plus imposants, le Getty Research Institute et le Rijksmuseum. Ce n'est que lorsque la question de la collaboration au sein d'un collectif de musées américains s'est posée que le SAAM a aligné son ontologie sur le CIDOC CRM. Le fait que ce modèle soit reconnu comme norme ISO est un élément qui favorise son utilisation, y compris par des institutions ayant la capacité d'imposer des standards de fait.

Cet exemple illustre également le rôle de « glue [56]» sémantique que les concepteurs du CIDOC CRM revendiquent. La dimension très générique du modèle est sa force et sa faiblesse. Sa faiblesse car au premier abord les entités apparaissent forcément trop génériques vis-à-vis de la spécificité de chaque collection et de chaque domaine. Mais c'est fondamentalement sa force dans la mesure où il ne s'agit pas pour les établissements culturels de changer de modèle (encore moins de logiciel) mais de transposer leurs données dans une grammaire sémantique

[55] (Juanals, Minel 2017)

[56] « *The primary role of the CRM is to serve as a basis for mediation of cultural heritage information and thereby provide the semantic 'glue' needed to transform todays disparate, localised information sources into a coherent and valuable global resource.* » Nick Crofts (OLDMAN, CRM Labs 2014)

commune. On n'adopte pas le CIDOC CRM, on l'utilise, de façon plus ou moins fine en fonction du projet et de la structure des données sources.

Ainsi, des descriptions d'objets patrimoniaux (archives, œuvres) qu'elles soient sommaires ou détaillées peuvent être liées à des discours (articles scientifiques, récits d'historiens). Les recherches et expérimentions que nous avons décrites aux chapitres 3.2 et 3.3 vont dans ce sens. Ces initiatives pourraient, comme c'est le cas du SyMoGIH élaboré par Laboratoire de recherche historique Rhône-Alpes (LARHRA), s'orienter vers une participation au LODLAM (Linked Open Data in Libraries Archives and Museums) via une transposition en CIDOC CRM. L'enrichissement mutuel changerait alors d'échelle. Le projet d'alignement du SyMoGIH sur le CIDOC CRM nous a permis de mettre en évidence la problématique des héritages de propriétés au sein d'une ontologie. Pour exprimer sa dimension hypothétique, il faut modéliser l'acteur « SyMoGIH » en tant qu'assertion historique, c'est-à-dire une information sourcée et datée.

La maîtrise des processus d'échange de données est un enjeu majeur de l'activité de la société A&A Partners, fournisseur de solution web de gestion, publication et syndication de collections en ligne. Spécialisée dans le secteur des musées, la plateforme WebMuseo a été développée dans le respect des standards du web et notamment du Dublin Core. De fait, chaque application WebMuseo est potentiellement un entrepôt de données et un service web de moissonnage OAI-PMH.

L'installation d'une solution de gestion pour un nouveau client passe, dans la majeure partie des cas, par une reprise de l'existant. Cette étape nécessite une gestion rigoureuse du projet et une implication des clients dans les phases de vérification et de validation des livraisons de test. L'enjeu est de faire correspondre les concepts couvrant les champs sources avec ceux de la base cible. Il faut donc s'intéresser au sens que les rubriques sources ont pour le client. Par exemple une colonne Prix dans un tableau Excel peut correspondre à une estimation de valeur, une valeur d'assurance, ou un prix d'achat.

La solution WebMuseo Gestion est un outil de production des données qui permet de les exploiter en vue d'usages métiers spécifiques : extraction et impression du

registre légal (18 colonnes), export à destination du catalogue collectif des musées de France Joconde, moissonnage OAI-PMH d'un espace de publication. Dans ces cas le flux de données est toujours unilatéral : de la base WebMuseo Gestion vers un objet informationnel fixe (registre, archive OAIS) ou susceptible d'être mis à jour (catalogue en ligne). La réalisation d'un connecteur d'export/import e-ReColNat (§5.2) a fait émerger la problématique de la mise à jour des référentiels embarqués dans WebMuseo et de la pérennité des descriptions précédemment saisies.

Gestion des référentiels, mises à jour des données, intégrité de l'information sont trois préoccupations essentielles du développement du web de données liées auxquelles la société A&A Partners est déjà sensibilisée.
Comment gérer la multitude des référentiels et des bases d'autorités ? Les avancées du programme HADOC en matière d'alignement des thésaurus en SKOS et mise en ligne de référentiels tels qu'ONOMA seront suivis avec attention.
Comment assurer la mise à jour des *triplestrore* ? La réflexion autour de l'exposition des données au sein de *triplestore* sera naturellement accompagnée du souci de leur mise à jour. Les modélisations imaginées au §6.2 cherchent le bon équilibre entre une granulosité satisfaisante et un systématisme favorisant la dimension opérationnelle des exports.
Comment conserver la traçabilité des données agrégées ? Lors des exports de triplets ré-utilisables l'attention sera portée sur les possibilités d'encapsulation du contexte de l'archive.

La transposition sémantique de l'indexation des collections suppose la production de données de qualité. Une stratégie d'interprétation en aval via des outils de fouilles de données permettrait d'élargir le spectre aux données semi-structurées (au sein de rubriques texte) ou non structurées (fichiers texte attachés aux notices.) Les liens vers les référentiels communs découleraient non pas d'un choix explicite mais d'une analyse automatisée des contenus.

Sur le plan fonctionnel, cette perspective pourrait inspirer le développement d'outils d'éditorialisation permettant de faire cohabiter des assertions différentes voire contradictoires.

Bibliographie

Web de données : méthodes et techniques

[1] T. FRANCART, « Partager et réutiliser des données structurées sur le web », *Blog Sparna*, 03-déc-2017.

[2] T. FRANCART, « Penser, modéliser (pour le web de données) (1/2) », *Blog Sparna*, 05-déc-2014.

[3] T. FRANCART, « Penser, modéliser (pour le web de données) – 2/2 », *Blog Sparna*, 26-mai-2014.

[4] T. FRANCART, « Recherche d'informations : du plein-texte aux ontologies », *Blog Sparna*, 24-août-2014.

[5] J. DELAHOUSSE, « Sur l'alignement et la mise en correspondance de terminologies », *Leçons de choses*

[6] P. VANDENBUSSCHE, B. VATANT, et J. CHARLET, « Linked Open Vocabularies, un écosystème encore fragile », présenté à Atelier Qualité et Robustesse pour le Web de Données in 23es journées Francophones d'Ingéniérie des Connaissances (IC), Montpellier, 2012.

[7] STEINBERG, Karine, 2016. Qualité des données de santé disponibles en France et de leurs modèles - Comment la garantir pour répondre aux enjeux de la gestion des connaissances médicales ? [en ligne].. INTD-CNAM-Institut national des techniques de la documentation. [Consulté le 6/12/2017]. Disponible à l'adresse : https://memsic.ccsd.cnrs.fr/mem_01476178/document

[8] MENON, Bruno, 2016. Comprendre les standards du web de données. I2D – Information, données & documents [en ligne]. 1 juillet 2016. Vol. me 53, n° 2, pp. 32-34. [Consulté le 6/12/2017]. Disponible à l'adresse : https://www-cairn-info.proxybib.cnam.fr/revue-i2d-information-donnees-et-documents-2016-2-p-32.htm

Interopérabilité des données culturelles

[9] P. BOURDENET, « Interopérabilité des systèmes d'information documentaire ». Techniques de l'Ingénieur, 10-mai-2016.

[10] J.-M. NOYER et M. CARMES, « Le mouvement " Open Data " dans la grande transformation des intelligences collectives et face à la question des écritures, du web sémantique et des ontologies. » 2012.

[11] B. JUANAIS et J.-L. MINEL, « La construction d'un espace patrimonial partagé dans le Web de données ouvert », *Communication[En ligne]*, vol. 34, n° 1, août 2016.

[12] B. JUANALS et J.-L. MINEL, *Enjeux numériques pour les médiations scientifiques et culturelles du passé*. Presses universitaires de Paris Nanterre, 2017.

[13] B. DUHAMEL, « Les technologies du Web sémantique et du record linkage au service de data.bnf.fr et du Linked Open Data culturel : étude sur les nouveaux paradigmes informationnels », Université Lille 3, Lille, 2014.

[14] *Onoma : un référentiel d'acteurs du ministère de la Culture ancré dans le Web sémantique*, 2017. [en ligne]. INHA, Galerie Colbert, [Consulté le 29 août 2017]. Disponible à l'adresse : https://www.inha.fr/fr/agenda/parcourir-par-annee/en-2017/juin-2017/onoma-un-referentiel-d-acteurs-du-ministere-de-la-culture-ancre-dans-le-web-semantique.html

Présentation du CIDOC-CRM

[15] NF ISO 21127 - Information et documentation - Une ontologie de référence pour l'échange d'informations du patrimoine culturel, 2014.

[16] LE BOEUF, Patrick, 2013. *De la sémantique des inventaires aux musées en dialogue : la modélisation CIDOC CRM* [en ligne]. 4 avril 2013. [Consulté le 28 août 2017]. Disponible à l'adresse : https://hal-bnf.archives-ouvertes.fr/hal-00807664

[17] D. OLDMAN et CRM Labs, « The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER ». Professor Donna Kurtz, In association with CLAROS4 (Oxford University), CultureBroker5, CultureCloud, Delving BV6, Institute of Computer Science, F.O.R.T.H (Heraklion)7, and ResearchSpace (British Museum8), juill-2014.

[18] M. RIZZA, C. BARBANT, P. L. BOEUF, et S. Fargier-Demergès, « 1. Le document au cœur de l'organisation muséale », *Documentaliste-Sciences de l'Information*, vol. 51, n° 2, p. 30-43, juill. 2014.

[19] P. LE BOEUF, « Modélisation conceptuelle de l'information bibliographique et muséologique : CIDOC CRM et FRBR 00 », oct-2009. [En ligne]. Disponible sur: http://portaildoc-intd.cnam.fr/Record.htm?idlist=150&record=19177228124919954009. [Consulté le: 19-juin-2017].

Applications du CIDOC-CRM

[20] OLDMAN, Dominic, MAHMUD, Joshan et ALEXIEV, Vladimir, 2013. *The Conceptual Reference Model Revealed. Quality contextual data for research and engagement: A British Museum case study* [en ligne]. 2013. ResearchSpace Project. [Consulté le 11 octobre 2017]. Disponible à l'adresse : https://confluence.ontotext.com/display/ResearchSpace/BM+Mapping

[21] SZEKELY, Pedro, KNOBLOCK, Craig A., YANG, Fengyu, ZHU, Xuming, FINK, Eleanor E., ALLEN, Rachel et GOODLANDER, Georgina, 2013. Connecting the Smithsonian American Art Museum to the Linked Data Cloud. In : *The Semantic Web: Semantics and Big Data - 10th International | Philipp Cimiano | Springer* [en ligne]. [Consulté le 10 octobre 2017]. Disponible à l'adresse : https://www.isi.edu/~szekely/contents/papers/2013/eswc-2013-saam.pdf

[22] P. TCHIENEHOM, « Projet ModRef : Migration de Données vers des Triplestores CIDOC-CRM », présenté à INFORSID 2017, Toulouse, 2017.

[23] A.-V. SZABADOS, K. BRIATTE, et R. LETRICOT, « Utiliser l'Ontologie CIDOC-CRM pour l'information relative au patrimoine culturel », in *THATCamp Paris 2012 : non-actes de la non-conférence des humanités numériques [en ligne]*, Paris, 2012.

Ingénierie des connaissances

[24] C.-C. BUTEZ et F. BERETTA, « Un SIG collaboratif pour la recherche historique Partie 1 : Naissance et conception d'un système d'information géo-historique collaboratif : Conception d'un atlas historique numérique et d'une palteforme de travail collaborative à partir de la méthode SyMoGIH », *Géomatique Expert*, n° 91, mars 2013.

[25] PRINGUET, Virginie, 2017. *Vers un atlas de l'art dans l'espace public : la modélisation d'un musée réticulaire* [en ligne]. Rennes 2. [Consulté le 10 octobre 2017]. Disponible à l'adresse : http://www.theses.fr/2017REN20006

[26] BERETTA, Francesco et VERNUS, Pierre, 2012. Le projet SyMoGIH et la modélisation de l'information : une opération scientifique au service de l'histoire. *Les Carnets du LARHRA* [en ligne]. 2012. N° 1, pp. 81-107. [Consulté le 11 octobre 2017]. Disponible à l'adresse : https://halshs.archives-ouvertes.fr/halshs-00677658/document

[27] STOCKINGER, Peter, LALANDE, Steffen et BELOUED, Abdelkrim, 2015. Le tournant sémiotique dans les archives audiovisuelles. *Les Cahiers du numérique* [en ligne]. 10 septembre 2015. Vol. 11, n° 3, pp. 11-38. [Consulté le 24 octobre 2017]. Disponible à l'adresse : https://hal.archives-ouvertes.fr/hal-01292376

[28] BACHIMONT, Bruno, 2017. *Patrimoine et numérique: technique et politique de la mémoire*. Bry-sur-Marne, France : INA. ISBN 978-2-86938-190-2.

[29] QUANTIN, Matthieu, LAROCHE, Florent et KEROUANTON, Jean-Louis, 2016. Récit historique et objet technique : outil de valorisation mutuelle. Cahiers d'histoire du CNAM [en ligne]. juillet 2016. Vol. 5, pp. 93-120. Disponible à l'adresse : https://hal.archives-ouvertes.fr/hal-01359096