



HAL
open science

La classification des connaissances et le web de données : opportunités pour la recherche

Marie Carlin

► **To cite this version:**

Marie Carlin. La classification des connaissances et le web de données : opportunités pour la recherche. Digital Libraries & Information Sciences (DLIS), 2018. hal-03700675

HAL Id: hal-03700675

<https://hal.science/hal-03700675>

Submitted on 21 Jun 2022

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.

LA CLASSIFICATION DES CONNAISSANCES ET LE WEB DE DONNÉES : UNE OPPORTUNITÉ POUR LA RECHERCHE

01/10/2018 MARIE CARLIN UN COMMENTAIRE

Marie Carlin est élève conservatrice à l'ENSSIB (dcb27). Titulaire d'un master en histoire moderne et du master technologies numériques appliquées à l'histoire de l'École des chartes, elle a travaillé huit ans au Centre de recherche du château de Versailles en tant que chargée de recherche en iconographie.

[Télécharger la version PDF](#)

Résumé

Nombreux sont les projets qui, depuis une petite décennie, s'approprient le web de données pour exposer leurs données, les ouvrir et les partager. Heureuse initiative, car le décloisonnement des données permet d'entrevoir de nouvelles possibilités de fouilles textuelles, de croisement et d'exploitation de la documentation dans des perspectives de recherche à une échelle jusqu'ici jamais atteinte. Le présent article se propose de présenter un projet qui se situe dans cette perspective d'ouverture et de mise à disposition de données structurées. Par la fusion de deux référentiels issus d'institutions différentes, mais travaillant sur des thématiques communes, le projet *Grand Siècle* a pour objectif de rendre accessible un référentiel sur l'Ancien Régime de façon pérenne et interopérable. Nous avons pris le parti de ne pas développer l'aspect technique de cette fusion, mais de se concentrer sur l'aspect méthodique et les choix scientifiques qui ont présidé à l'élaboration de ces référentiels. Les aspects plus techniques pourront faire l'objet d'un autre article.

Cette initiative s'inscrit dans un élan plus étendu, puisque plusieurs outils (OpenTheso ^[1], Ginfo ^[2]) et des vocabulaires dédiés (SKOS ^[3]) ont été mis à disposition pour permettre la publication de référentiels en utilisant les technologies du web de données. L'émergence de ces objets technologiques montre l'intérêt nouveau suscité par la normalisation, même s'il faut parfois s'accommoder de certaines limites. Car ces outils ne permettent pas – ce n'est pas leur

rôle – de répondre aux interrogations que soulève ce type d'exercice classificatoire. Dans une entreprise de classement, il faut toujours se poser la question de ce que l'on veut laisser à voir, comment le laisser voir et pourquoi. Puisque classer consiste à figer une certaine vision du monde, comment organiser l'information pour qu'elle soit le reflet le plus fidèle d'une réalité ? Comment rendre compte de tous les savoirs, de toutes les bifurcations, de toutes les subtilités d'une réalité dans un modèle, certes complexe, mais immobile ? Exercice impossible, tout le monde en conviendra.

Alors, à quels choix doit-on se plier, quels sacrifices doit-on concéder, quels solutions ou rectificatifs peut-on envisager pour modéliser le moins mal possible cette réalité derrière laquelle s'engage la course ?

Cet article ne se propose pas de formuler de réponses à des questions qui n'en appellent pas, mais plutôt d'expliquer une démarche intellectuelle en proposant de revenir sur la définition des objets classificatoire et leur évolution dans un premier temps, puis d'exposer les choix qui ont été engagés pour classer ces concepts et produire ce référentiel d'Ancien Régime dans un second temps.

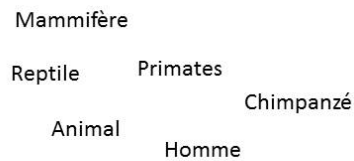
Les systèmes classificatoires

De l'absolue nécessité d'ordonner

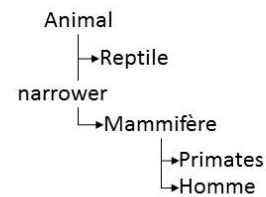
On pourrait définir un référentiel comme un outil servant à instrumenter les connaissances dans un champ délimité. Il organise la structuration et offre la normalisation d'un domaine de connaissance permettant à tous ceux qui s'y réfèrent de parler la même langue : grâce à des règles d'écriture (normalisation de la forme, syntaxe), de négociations de sens, une organisation hiérarchique ou sémantique entre les termes...

Les systèmes d'organisation des connaissances

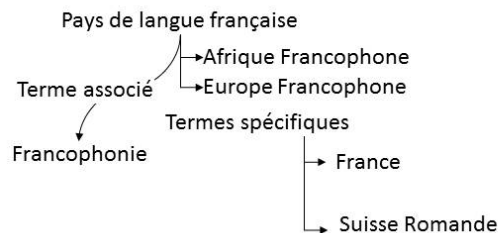
Vocabulaire contrôlé



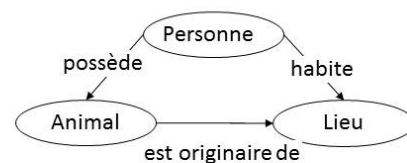
Taxinomie



Thésaurus



Ontologie



Les systèmes d'organisation des connaissances.

Diapo extraite de l'intervention de Gautier Poupeau sur le web sémantique (École nationale des chartes, avril 2016).

Il est heureux de constater que les grands systèmes classificatoires connaissent depuis une dizaine d'années une attractivité nouvelle, non plus seulement auprès des professionnels de la documentation qui en ont toujours eu usage, mais également auprès des grandes institutions productrices de contenu. Cet intérêt se mesure à la multiplication des grands programmes de normalisation et d'exposition des données institutionnelles, dont beaucoup sont un puissant levier pour la recherche actuelle^[4].

Ce retour peut sembler inattendu, car bien que n'ayant pas totalement disparu des systèmes de recherches d'information, les décennies des années 1990 et surtout 2000 leur avaient tenu mauvaise presse : trop lourds à organiser et à maintenir, chronophages, inutiles, dépassés... Tout un champ lexical de l'obsolescence alimentait alors le discours des détracteurs de ces outils, sans qu'on puisse cependant leur donner foncièrement tort : à quoi bon maintenir un instrument complexe de hiérarchisation des connaissances si quelques mots clés tapés dans un moteur de recherche suffisent à obtenir la réponse ? Quel besoin de relier sémantiquement des concepts ou de les unifier typographiquement si un correcteur orthographique me fait trouver le bon résultat malgré l'approximation ? Nous étions alors dans l'ère du nuage de mots-clés et de la folksonomie.

Cette « revanche des référentiels » tient en partie à leur efficacité en matière de recherche d'information mais aussi à la modification de la structure du support de l'information – à savoir le web – qui a depuis une décennie évolué vers un modèle moins cloisonné. Aujourd'hui, **l'une des extensions du web permet de penser une**

structuration qui relie non pas des [documents](#) (les pages HTML), mais bien les [données structurées](#) contenues dans ces documents (métadonnées, entités nommées...). Le modèle est donc très différent du web de document puisqu'on passe d'un système de silos de données isolés les uns des autres, à des données structurées qui communiquent sémantiquement entre elles pour construire un réseau global d'information^[5].

Dans ce contexte, on comprend mieux pourquoi la normalisation – raison d'être des grands outils classificatoires – est indispensable : **elle devient le gage de l'interopérabilité des données.**

Un référentiel de l'Ancien Régime

Les référentiels qui constituent le projet *Grand Siècle* ont été élaborés à partir de la mise en commun de données produites par deux institutions de recherche : le Centre de recherche du château de Versailles (CRCV)^[6] et l'Institut de recherche en Musicologie (IReMus)^[7].

Bien que travaillant sur des thématiques et des périodes assez différentes – l'Ancien Régime et la représentation du pouvoir pour le CRCV ; la musicologie et ses disciplines annexes sur une chronologie longue pour l'IReMus – il existe de nombreux carrefours de convergences entre ces deux institutions. Celles-ci mettent en place des programmes de recherche dont les valorisations prennent des formes assez similaires par le biais de publications (de sources, de colloques, de monographies) et l'élaboration d'outils en ligne (bases de données, banques d'images, publications électroniques...).

Ces valorisations sont l'occasion de générer une importante masse de données structurées, liées aux indexations et aux activités de recherche : index *locorum* et *nominum* pour les éditions papiers ; thésaurus de fonctions, d'institutions, de mots clé ; métadonnées pour les outils électroniques.

Ce sont sur ces données structurées que nous avons entrepris un travail de fusion et d'exposition, afin de produire un outil de référence pour quiconque s'intéresse aux structures de l'Ancien Régime.

Afin de mieux saisir la nature de ces données, il convient donc de préciser leur origine, et de s'attarder quelque peu sur les sources dont elles sont extraites.

IReMus

La première source de données provient du projet « Mercure Galant » porté par l'IReMus^[8]. Celui-ci propose l'édition électronique en XML-TEI d'articles relatifs à la musique, à la vie musicale, au théâtre et à la critique littéraire, publiés dans le célèbre

périodique. Cette édition en ligne est conçue comme une base de données documentaire : chaque texte est associé à des index et thésaurus qui recensent les noms de personnes citées, ceux des nombreuses institutions qui participent à la vie collective et artistique du temps, les lieux, les dates, les titres d'œuvres ainsi que les notions. Le bornage chronologique est très clairement circonscrit, et ne concerne que les années 1672 (date de création) à 1710 (mort du fondateur Jean Donneau de Visé).

Mercure Galant, mars, première partie
(tome 3), 1689

← rasilly

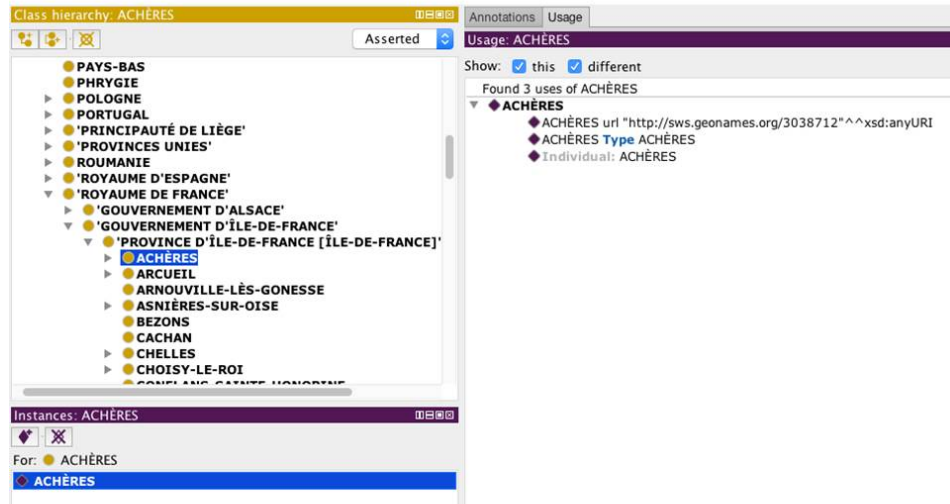
- └ Au Roy, sur l'azile donné au Roy d'Angleterre
- └ Sur l'Arrivée du Prince de Galles en France. Ode
- └ Sur les Conquestes de Monseigneur le Dauphin. Ode
- └ [Livre nouveau de M^r Bacilly]*
- └ A Monseigneur le Dauphin, sur son Retour des Conquestes d'Allemagne. Ode
- └ [Deux Sonnets de L.D.M.C.D.B.]*
- └ La Fontaine de Jouvence.
- └ [Paroles sur les Conquestes de Monseigneur]
- └ [Histoire de l'Accademie de Soissons]
- └ [Entretiens touchant l'entreprise du Prince d'Orange sur l'Angleterre]
- └ [Guerre des Turcs avec la Pologne, la Moscovie, & la Hongrie]
- └ [Ouvrage de Mr le Pays]

Au sortir de l'Eglise [des Prêtres Oratoires d'Orléans] le Roy monta en chaise pour aller coucher à Tours, après avoir fait mille honnestetez à M^r de Creil. Il y fut receu le soir au bruit du Canon par M^r le Marquis de **Rasilly**, Lieutenant general pour le Roy en Touraine, qui estoit allé au devant de ce Monarque à trois lieuës de la Ville, accompagné de la Noblesse, & suivy de ses Gardes & de la Maréchaussée. M^r l'Archevesque de Tours, & M^r de Miromenil, Intendant. estoient avec ce Marquis. Sa Majesté estant arrivée à la premiere porte de l'enceinte de la Ville, y receut les complimens & les presens des Echevins. La Bourgeoisie sous les armes formoit une double haye, & par tout où il passa il y eut des Illuminations aux fenestres. On luy avoit préparé un appartement au logis de M^r de **Rasilly**, & à la descente du Carosse il trouva Madame la Marquise de **Rasilly**, & Madame de Miromenil, qu'il salüa. Toutes les Compagnies en corps le vinrent complimenter ; après quoy on servit diverses tables pour Sa Majesté & pour les Seigneurs de sa suite. Le Mercredi 2. de ce mois, le Roy entendit la Messe au Convent des Jacobins avec la Musique de la Cathedrale, & partit de Tours à six heures du matin. La Bourgeoisie estoit sous les armes comme le jour precedent, & M^r de **Rasilly** conduisit Sa Majesté avec le mesme cortège jusqu'au mesme lieu où il l'avoit esté recevoir. [...]

Copie d'écran de la publication en ligne du Mercure Galant, mars 1689.

Chaque article est dépouillé et l'indexation des textes est reportée dans une ontologie élaborée grâce au logiciel Web Protégé. Nous verrons en temps voulu la structure de cette ontologie, il nous paraît pour l'instant suffisant de dire que cette ontologie se compose de 6 référentiels : personne ; lieu ; fonction ; institution ; congrégation ; mot-clé. Chaque item est renvoyé à la source dans laquelle il est cité et chaque nom de lieu est aligné sur une base de données géographiques grâce à ses coordonnées GPS.

Indexation des lieux dans Web Protégé



Indexation des items lieu dans Web Protégé, avec l'alignement vers le référentiel GeoNames.

CRCV

Le CRCV est un Groupement d'intérêts public (GIP) qui produit, autour de programmes de recherche triennaux, des outils de diffusion de la recherche (bases de données, publications papier ou électronique de colloques, de monographies, de sources). Les axes de recherches concernent l'étude des institutions curiales, la société de cour et la représentation et l'exercice du pouvoir en Europe à l'époque moderne, principalement aux XVII^e et XVIII^e siècles.

Au travers des programmes de recherche, le CRCV produit également des index (index *locorum* et *nominum* pour les éditions papiers) et des référentiels pour l'indexation des bases de données : référentiel de personne ; thésaurus de lieu ; fonction ; institution ; mot-clé. La structuration des données du CRCV reste celle classique d'un thésaurus (une structuration hiérarchique des termes) et n'inclue pas de relations sémantiques comme cela peut être le cas dans l'ontologie de l'IReMus.

Dans le cadre du projet qui nous intéresse, les référentiels ont été extraits de deux bases de données : la base de données *Hortus*^[9] ; la base de données *biographique*^[10].

Thésaurus des institutions de la base de données biographique du CRCV

TG Département de la Maison du roi
 TS Académie française
 TS Académie royale des inscriptions et belles-lettres
 EP Académie royale des inscriptions et médailles
 EP Petite Académie
 TS Académie royale des sciences
 EP Académie des sciences
 TS Bibliothèque du roi
 TS Bâtiments du roi
 TS Académie de France à Rome
 TS Académie royale d'architecture
 TS Académie royale de peinture et de sculpture
 TS Jardin du roi
 TS Maison de Mesdames
 TS Maison de Mesdames aînées
 TS Maison de Madame Adélaïde
 TS Maison de Madame Henriette
 TS Maison de Mesdames cadettes
 TS Maison de Madame Louise
 TS Maison de Madame Sophie
 TS Maison de Madame Victoire
 TS Maison de la dauphine
 TS Maison de Marie-Adélaïde de Savoie
 TS Maison de Marie-Anne-Victoire de Bavière
 TS Maison de Marie-Antoinette [dauphine]
 TS Maison de Marie-Josèphe de Saxe
 TS Maison de Marie-Thérèse-Raphaëlle d'Espagne

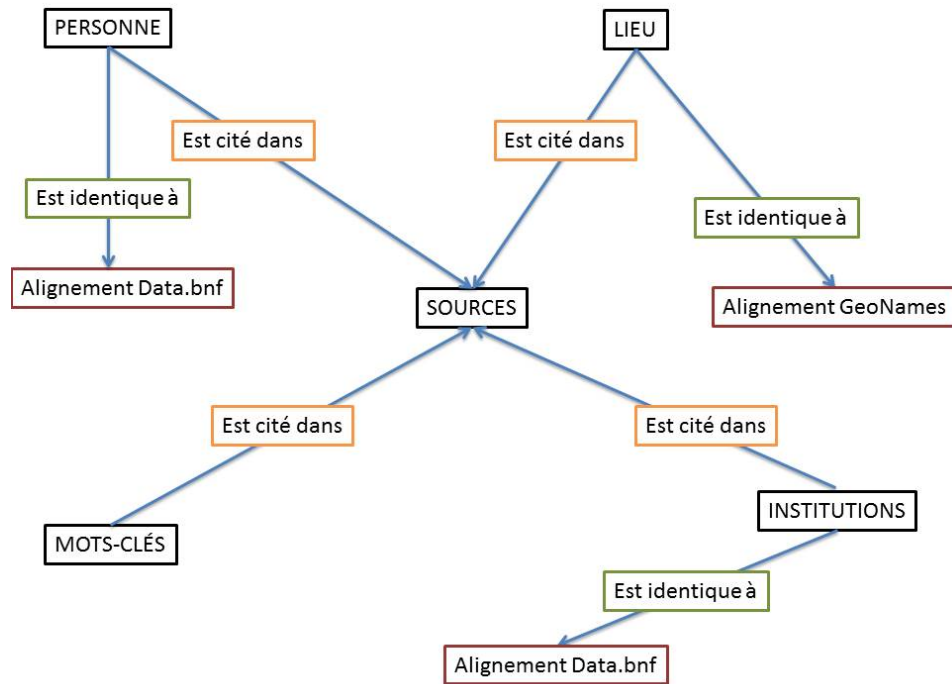
Fusionner ?

Nous le voyons donc, les données structurées produites par ces deux organismes de recherche sont analogues : des objets similaires (des thésaurus et référentiels), traitant de la même période, d'une population relativement homogène d'un point de vue socio-économique, sur une aire géographique relativement restreinte. Tout concourt à dire qu'une fusion de ces données puisse être envisageable, voire souhaitable.

Tout d'abord parce que les données produites dans le cadre de ces projets ne sont pas anecdotiques : elles sont le fruit de programmes de recherche et résultent d'un travail d'identification, de localisation et de validation par un collège d'experts. Elles sont donc intrinsèquement intéressantes, il serait donc utile de les faire partager à des fins de réutilisation par le plus grand nombre.

Ensuite parce que le schéma ontologique proposé par l'IReMus dans le cadre du projet « Mercure Galant » apparaît comme un enrichissement des données qu'il était intéressant de généraliser aux données du CRCV : un renvoi vers les sources dont chaque occurrence est extraite ainsi qu'une possibilité d'alignement des concepts vers d'autres référentiels (particulièrement intéressant pour le thésaurus des lieux) ajoute un complément intéressant au schéma de base.

Nous avons donc décidé de partir d'un schéma ontologique tel que présenter dans le schéma ci-dessous, et de nous concentrer sur des référentiels de personne ; lieu ; institution/fonction ; congrégation ; mot-clé.



Modélisation de l'ontologie Grand Siècle.

Organiser le monde

Un référentiel, même circonscrit à un domaine particulier de la connaissance, est une organisation structurée des savoirs : il fixe un état du monde. Les arbres de la connaissance de l'Encyclopédie et le scandale qu'ils ont alors soulevé sont là pour attester que classer n'est pas neutre. Aucun système de classification ne sera totalement satisfaisant, et c'est tant mieux, car sa structure traduit ce qui anime son créateur. « La classification est donc bien un exercice de pouvoir » affirme Robert Darnton^[11]. Ce à quoi nous ne pouvons qu'abonder.

Loin des enjeux qui entouraient l'édition des 17 volumes de l'Encyclopédie et sa nouvelle forme classificatoire des connaissances, nous avons dû toutefois procéder à des choix pour arrêter ce que nous voulions donner à voir.

Index nominus

Même dans le cas d'un référentiel *personne*, où l'on se contenterait de préciser les noms et dates de naissance et mort des individus, la question des règles de construction du terme générique génère des controverses.

Prenons le résultat de recherche pour le cardinal de Richelieu dans le référentiel IdRef (Identifiants et Référentiels pour l'Enseignement Supérieur et la Recherche), l'application développée et maintenue par l'Agence bibliographique de l'enseignement supérieur (Abes)^[12].

Richelieu, Armand Jean Du Plessis (1585-1642 ; cardinal duc de)^[13]

Puis prenons la même entité référencée sur data.bnf, le portail référentiel de la Bibliothèque nationale^[14] :

Armand Jean du Plessis **Richelieu** (cardinal duc de, 1585-1642)^[15]

Nous retrouvons les mêmes informations, placées dans un ordre différent, avec une variante typographique à Du Plessis/du Plessis^[16].

Même si l'agencement est différent, on comprend toutefois que le terme indexatoire est *Richelieu* : le cardinal duc de Richelieu est classé à *Richelieu*, son nom de terre, et non pas à *Du/du Plessis*, son patronyme. Cette confusion nom de terre/patronyme est commune, mais n'en reste pas moins une aberration. En effet, les noms de terre qui accompagnent un titre (duc, comte, marquis...) sont rattachés à un fief, une réalité matérielle circonscrite, et non à une personne. Ils peuvent donc, en cas de perte de la seigneurie, tout simplement disparaître (vente, défaite militaire, mariage dans le cas des femmes, décès...). De plus, ils ne sont théoriquement transmissibles qu'à l'aîné des fils.

Prenons un autre exemple, rencontré dans la base biographique du CRCV :

Adhémar de Montfalcon (Gabrielle-Pauline Bouthillier de Chavigny, comtesse d', 1735-1822)^[17]

La comtesse d'Adhémar de Montfalcon ne fut pas toujours appelée ainsi. Elle fut successivement connue sous les noms de : Mademoiselle de Beaujeu ; Gabrielle-Pauline Bouthillier de Chavigny (prénoms et patronyme) ; puis Gabrielle-Pauline Bouthillier de Chavigny, comtesse de Valbelle de 1752 à 1772 par mariage avec le comte Joseph Ignace Cosme Alphonse Roch de Valbelle ; et enfin Gabrielle-Pauline Bouthillier de Chavigny, comtesse d'Adhémar de Montfalcon de 1772 à son décès en 1822 par mariage avec Jean-Balthazar, comte d'Adhémar de Montfalcon (1731-1791). Comment déterminer quel est le nom « officiel » ?

On voit bien qu'il y a ici un certain nombre de variations possibles autour d'une seule et même personne, et que toutes ont une signification. Pourtant les grandes institutions ont pour règle de classer les noms de personne au dernier nom de terre connu. Personne n'aurait d'ailleurs idée d'aller chercher la marquise de Pompadour à l'entrée « Rochechouart », ou le cardinal de Richelieu à « Du Plessis ». Cela s'est installé comme une norme, et pourtant, il n'est pas difficile de démontrer la construction qu'il y a derrière cette classification : qui a constitué les grands catalogues du début du siècle dans les bibliothèques ? Sans doute des gens dont l'attachement à la terre et au titre avait valeur particulière.

Nous devons donc opérer un choix : respecter les règles onomastiques ou aligner nos pratiques sur les règles et usages des

grandes institutions productrices. Nous avons décidé de nous aligner sur la pratique la plus courante et de construire nos items sur ce modèle :

Nom de terre, Prénom(s) Patronyme (date-date), titre

Les structures

Structurer le réel

Ces exemples simples illustrent assez bien ce que nous avançons en introduction. Cette affirmation se renforce quand on aborde les structures de la société. Rien de moins évident que de fixer les forces qui animent et structurent une société, à plus forte raison une société qui nous est largement étrangère dans ses mentalités et son fonctionnement.

À l'image des processions générales, ces grands défilés de dignitaires qui servent à rendre visibles les corps les plus importants de l'ordre social, une classification est trompeuse. Elle ne peut être une représentation fidèle d'une société, car elle met en exergue certains éléments et en néglige d'autres. La préséance – élément central des processions – ne signifie pas pouvoir réel : richesse, statut et pouvoir ne vont pas de pair. Et surtout, chacun à sa définition du prestige : si l'on en croit la *Description de Montpellier* de 1768, le corps professoral (pourtant bien peu représenté et peu gratifié) est doté de bien plus de « dignité », ou de « qualité » que les membres du clergé qui ouvrent pourtant la marche dans un décorum extrêmement codifié et ostentatoire^[18].

Cela est vrai pour les corps représentés, mais qu'en est-il des absents ? Les non-participants sont-ils pour autant dépourvus de poids ? Comment donner à voir leur influence alors même qu'ils n'ont pas d'institutionnalisation formelle ? Donner à voir un organigramme de la société, est-ce représenter une société ?

Il est évident que cela ne suffit pas. Mais, matérialiser une société selon son mode d'organisation administratif, cela permet tout de même de donner à voir l'image qu'une société se fait d'elle-même, en tout cas ce qu'elle a envie que l'on voit d'elle-même.

Si on se penche sur les Menus Plaisir aux XVII^e et XVIII^e siècles, service de la Maison du roi en charge des fêtes et spectacles dans les Maisons royales, on constate que ce département était également employé à l'organisation des funérailles royales. La mort d'un prince est l'occasion d'une mise en scène théâtrale extrêmement précise avec ses acteurs, ses décors et un déroulé extrêmement codifié. Le matérialiser dans un arbre hiérarchisé permet de mieux saisir un aspect de cette société d'ancien régime.

Conscient des manques que peut comporter un tel outil, il peut toutefois être une première étape de réflexion pour saisir les structures, les manques, et les besoins de représentation d'une société d'Ancien Régime.

Exercer sous l'Ancien Régime

Nous avons donc décidé de nous orienter vers une structuration « officielle », c'est-à-dire reprenant les structures législatives et institutionnelles selon un schéma : administration centrale, succursale(s) et fonctions rattachées.

Nous avons un temps pensé à séparer les fonctions des institutions auxquelles elles étaient adjointes.

En effet, on peut être « sculpteur » ou « avocat » sans être rattaché aux Menus Plaisirs du roi ou au parlement de Bretagne. Cette logique de métier nous a semblé intéressante, car elle diversifie la recherche, et distingue bien la fonction de l'institution. Cette construction répond à une logique « personne » et non institutionnelle assez séduisante.
























Cependant, il est très vite apparu que cette classification présentait plus d'inconvénients que d'avantages.

En premier lieu, celui de ne pas pouvoir hiérarchiser l'information : proposer une liste alphabétique de fonctions dans un référentiel sur l'Ancien Régime ne fait pas beaucoup sens. De plus, comment savoir si, quand on le décontextualise, un terme doit s'interpréter comme une fonction ou comme un corps en représentation (une institution et non pas un individu) ? Par exemple, le terme « archer » peut se référer à la fonction d'archer dans une ville, ou alors le corps des archers de la ville défilant lors d'une parade et qui représente le plus souvent une fonction honorifique. Ce sont là deux interprétations très différentes, et qui n'ont pas le même sens.

De plus, ce référentiel a été pensé comme un instantané des structures d'Ancien Régime, il est donc plus judicieux de présenter une structure hiérarchisée, qui incarne la construction administrative, juridique, sociale de la société. Faire le choix de rattacher systématiquement les fonctions aux institutions présente ainsi plus fidèlement la société d'Ancien Régime : on exerce certes un métier, ou une charge, mais presque toujours dans le cadre d'une corporation de métier, d'une institution, d'une administration, d'une maison nobiliaire. D'aucuns objecteront que la société d'Ancien Régime n'est pas aussi rigide, et que beaucoup d'institutions et de fonctions étaient en réalité extrêmement intriquées. De plus, les fonctions peuvent être uniquement honorifiques, et non pas rattachées à une institution particulière. Nous leur donnerons raison. Mais nous pensons qu'il est plus simple, à ce stade du travail, d'exprimer ces complexités au sein même d'un thésaurus d'organisations, à travers une classe corporation de métier par

exemple, que de répéter deux fois l'information dans deux thésaurus distincts.

Nous avons pris le parti de hiérarchiser le plus possible les informations, afin de donner un instantané que nous espérons fidèle du formalisme des institutions de l'Ancien Régime. Les classes de premier niveau représentent les administrations ou organisations qui ont une unité, soit administrative, soit géographique, soit de corps. Ce qui génère des disproportions importantes, puisque, par exemple le terme « assemblée du clergé » ne comporte qu'une seule instance – « député » – alors que le terme « département de la Maison du roi » comporte 13 sous-classes et 7 niveaux de hiérarchie. Mais cela présente l'avantage de rendre visuellement les hiérarchies institutionnelles et les rôles de chacun.

- ▼  Congrégation
 - ▼  Clergé régulier
 - ▼  Chanoine régulier
 - ▼  Chanoine régulier de saint Augustin
 - ▶  Augustin
 - ▼  Augustine
 - ▼  Augustine hospitalière
 -  Augustine hospitalière de Beaulieu-les-Loches
 -  Augustine hospitalière de la Place royale de Paris
 -  Augustine hospitalière de la roquette de Paris
 - ▶  Chanoinesse régulière de saint Augustin
 -  Chanoine régulier de l'abbaye royale de Saint-Victor de Paris
 -  Chanoine régulier de l'ordre de Saint-Augustin, congrégation de Notre-Sauveur
 - ▶  Chanoine régulier de la maison royale de Sainte-Croix-de-la-Bretonnerie
 - ▶  Chanoine régulier de Prémontré
 - ▶  Chanoine régulier de Saint-Étienne de Dijon
 - ▶  Chanoine régulier de Saint-Martin de Plaimpied
 - ▶  Chanoine régulier de Saint-Ruf
 - ▶  Chanoine régulier de Sainte-Geneviève
 -  Chanoinesse augustine de l'abbaye Saint-Pantaléon de Toulouse
 -  Chanoinesse régulière anglaise de Paris
 - ▶  Haudriettes
 - ▶  Congrégations de clercs réguliers

Extrait du thésaurus des congrégations dans OpenTheso.

Toutes les ambiguïtés n'ont pas encore été levées, et certains cas n'ont pas pu être tranchés. Nous nous accordons encore un temps de correction afin de présenter un état le plus satisfaisant possible au vue des critères que nous nous sommes imposés. Toutefois, il convient de préciser que ce projet à un objectif participatif, et que nous prévoyons un module de suggestion de correction et de développer un mode participatif : nous sommes convaincus que l'apport de la communauté à la construction de ce référentiel est essentiel à son exemplarité.

Les lieux

Étrangement, les toponymes ont présenté moins de difficultés, ou du moins d'un autre ordre. La structure étant spatiale, la hiérarchie allait de soi : un monument appartient à une ville qui est localisée dans une province ou un gouvernement, qui sont eux même

rattachés à un pays ou un royaume. Les unités de sens et leur hiérarchisation sont moins complexes à mettre en ordre qu'une entité « personne » ou « institution ». D'autant que nous avons décidé de nous concentrer sur les lieux réels, sans indexer les lieux imaginaires. Se posait toutefois le problème de la diachronie. Les unités territoriales décrites ou mentionnées dans le *Mercurus Galant* ont une implantation correspondant à une réalité qui n'est plus celle qui a cours administrativement aujourd'hui. Les unités administratives ont évolué, les frontières se sont déplacées, les noms eux-mêmes ont pu changer. Il paraît impossible de lister chronologiquement tous les états d'une division territoriale sur plus de 300 ans.

Dans son dépouillement du *Mercurus Galant*, dont le bornage chronologique s'étend de 1672 à 1710, l'équipe de l'IReMus avait élaboré un thésaurus très fourni sur un état du monde en 1713. La date de 1713 est relativement importante en Europe puisqu'elle fixe, par les traités d'Utrecht qui marque la fin de la guerre de succession d'Espagne, un état de l'Europe stable, du moins jusqu'à la fin de la guerre de succession d'Autriche en 1748.

Il nous a donc paru intéressant de conserver ce thésaurus de 1713, et d'en aligner les concepts vers un référentiel de géo-référencement actuel, GeoNames^[19]. Cette base de données géographique géolocalise les lieux à partir des coordonnées géographiques, qui restent stables, quels que soient les changements statutaires qui s'opèrent au fil de l'histoire. Ainsi, chaque concept du thésaurus géographique élaboré dans le cadre du projet *Grand Siècle* est indexé sous sa forme ancienne 1713, et aligné vers sa dénomination actuelle grâce à ses coordonnées géographiques. **Ce système des alignements, rendu possible par les technologies du web de données, permet donc de traiter la diachronie de façon satisfaisante.**

Conclusion

À l'heure de la rédaction de cette présentation, l'outil dans lequel les données ont été migrées n'est pas encore en accès libre. Les thésaurus de personne, institution/fonction et congrégations ont été fusionnés et formalisés en XML-SKOS pour être intégrés dans l'outil OpenTheso^[20]. Cependant, de nombreux chantiers restent encore à consolider avant de permettre une complète éditorialisation qui interviendra, nous l'espérons, avant l'été 2019.

Nous pouvons toutefois avancer quelques chiffres, et présenter les perspectives que nous envisageons pour ce projet.

Le référentiel des personnes comprend à ce jour 4 500 entrées dont seulement 47 noms communs aux deux institutions. Ce qui est extrêmement peu, surtout au regard des critères qui ont présidé à la fusion des thésaurus. Il nous reste, pour cette partie du référentiel, à compléter les alignements vers les autres grands référentiels que sont data.bnf et IdRef, où là encore, nous avons été surpris par le peu

de concordance. Au-delà de l'étonnement, il faut savoir s'en réjouir : cette faible symétrie prouve que le travail d'identification est loin d'être terminé. Un outil tel que celui que nous construisons peut donc permettre de poursuivre la tâche.

Le thésaurus des lieux comprend 3 824 entrées, avec un niveau de granularité de 5 (pays ou royaume ; gouvernement ; province ; ville ; édifice). Les alignements vers le référentiel GeoNames sont complétés pour ce qui est du thésaurus de l'IReMus, il reste cependant à le faire fusionner avec le référentiel du CRCV, qui comprend environ 900 unités. Cette fusion des données ne peut pas être automatisées, car elle demande de niveler les informations : le référentiel du CRCV est un référentiel contemporain qu'il faut donc transposer en 1713. Tâche ingrate et particulièrement chronophage...

De plus, il est envisagé de produire un référentiel des différents lieux du château de Versailles, pour permettre l'identification des pièces et des renvois aux plans en accès libre sur la banque d'images du CRCV[21]. Ce projet soulève là encore de nombreuses questions, tant le château a connu des modifications au cours de son histoire : quel état choisir, comment traduire l'évolution architecturale d'un tel ensemble ? Nous en sommes pour l'instant à un stade embryonnaire, mais l'idée est séduisante.

Le thésaurus des institutions/fonctions comprend 2 500 entrées répartis dans 43 classes de premier niveau, et 2 364 termes spécifiques à un niveau de granularité de 7. La particularité de ce thésaurus tient à ce que les items ne sont pas tous issus d'une indexation, mais de recherches plus approfondies pour permettre une structure hiérarchique. Reste là aussi à effectuer les alignements vers les autres grands référentiels que sont data.bnf et IdRef et à valider certaines intuitions.

L'idée d'un outil collaboratif a rapidement été évoquée plus en amont. Il est bon de rappeler en conclusion l'esprit qui anime l'élaboration de cet outil, qui se veut avant tout un outil au service des chercheurs, et qui donc va nécessiter parfois leur implication. L'état actuel de ce référentiel est vu comme une première étape, un prototype ouvert à de plus grandes ambitions. Il pourrait être envisagé un élargissement des collaborations à d'autres collègues européens, et la fédération d'un groupe de travail autour de ce premier essai pour proposer un vaste référentiel à l'échelle de plusieurs pays. La publication, il est à souhaiter prochaine de ce premier état, pourrait être un test pour élargir les objectifs.

Notes bibliographiques

[1] <https://masa.hypotheses.org/99>

[2] <http://www.culture.gouv.fr/Divers/Harmonisation-des-donnees-culturelles/Referentiels/Les-vocabulaires-scientifiques-et-techniques/L-application-GINCO>

[3] <https://www.w3.org/2004/02/skos/>

[4] Data.bnf, <https://ukat.aim25.com/>

[5] « W3C Semantic Web Activity », World Wide Web Consortium (W3C), 7 novembre 2011, <https://www.w3.org/2001/sw/> (consulté le 26 août 2018).

[6] <http://chateauversailles-recherche.fr/>

[7] Institut de recherche en musicologie, UMR 8223, CNRS, <http://www.iremus.cnrs.fr/>

[8] <http://obvil.sorbonne-universite.site/projets/mercure-galant>

[9] <http://chateauversailles-recherche.fr/francais/ressources-documentaires/bases-de-donnees-en-ligne/base-hortus-121>

[10] <http://chateauversailles-recherche.fr/francais/ressources-documentaires/bases-de-donnees-en-ligne/base-biographique/dictionnaire-des-personnages>

[11] R. Darnton, *Le Grand Massacre des chats*, Paris, R. Laffont, 1985, p. 178.

[12] <https://www.idref.fr/autorites.jsp#>

[13] <https://www.idref.fr/027349217>

[14] <http://data.bnf.fr/>

[15] http://data.bnf.fr/fr/11921894/armand_jean_du_plessis_richelieu/

[16] La règle typographique veut que la majuscule ne s'applique qu'au patronyme, pas au nom de terre.

[17] <http://www.chateauversailles-recherche-ressources.fr/jlbweb/jlbWeb?html=notdictionnaire&ref=146>

[18] « Montpellier en 1768 d'après un manuscrit anonyme inédit », in *Archives de la ville de Montpellier*, J. Berthelé, Montpellier, 1909.

[19] <http://www.geonames.org/>

[20] Pour une présentation de l'outil, on pourra se reporter au billet publié dans le carnet du MASA : <https://masa.hypotheses.org/99>

[21] Projet VERSPERA : <https://verspera.hypotheses.org/>

LA CLASSIFICATION DES CONNAISSANCES ET LE WEB DE DONNÉES : UNE OPPORTUNITÉ POUR LA RECHERCHE

<https://dlis.hypotheses.org/2157>



Marie Carlin

Marie Carlin est élève conservatrice à l'ENSSIB (dcb27). Titulaire d'un master en histoire moderne et du master technologies numériques appliquées à l'histoire de l'École des chartes, elle a travaillé huit ans au Centre de recherche du château de Versailles en tant que chargée de recherche en iconographie.

CENTRE DE RECHERCHE DU CHÂTEAU DE VERSAILLES

INSTITUT DE RECHERCHE EN MUSICOLOGIE ONTOLOGIE

PROJET MERCURE GALANT RÉFÉRENTIELS ET AUTORITÉS

SYSTÈMES D'ORGANISATION DES CONNAISSANCES WEB DE DONNÉES