

# Conception de Bases de Données Prosopographiques en Histoire - Un Etat de l'Art

## Design of prosopographic databases in history - A state of the art

Jacky Akoka<sup>1</sup>, Isabelle Comyn-Wattiau<sup>2</sup>, Cédric du Mouza<sup>3</sup>

<sup>1</sup> lab. CEDRIC, CNAM & TEM-Business School, Paris, jacky.akoka@cnam.fr

<sup>2</sup> ESSEC Business School, wattiau@essec.edu

<sup>3</sup> lab. CEDRIC, CNAM, Paris, dumouza@cnam.fr

**RÉSUMÉ.** La prosopographie a pour but l'étude de similitudes de trajectoires des personnes dans leur parcours biographique. Elle s'appuie sur une base de données contenant des informations relatives à des personnes d'un milieu spécifique définies chronologiquement et géographiquement. L'objectif de cet article est de proposer un état de l'art des bases de données prosopographiques mettant l'accent sur ses concepts spécifiques et la recherche afférente à la modélisation de ces concepts. Nous présentons différentes définitions de la prosopographie. Puis nous décrivons la méthode prosopographique qui a connu une certaine évolution du fait du recours aux bases de données informatisées. Un panorama de ces bases de données prosopographiques est présenté en soulignant leurs avantages et leurs limites. Les principaux concepts de ces systèmes sont décrits, notamment la personne qui est le concept central, l'onomastique qui permet l'étude des noms des personnes et des institutions et les notices biographiques qui recensent les événements auxquels les personnes sont liées. La description de ces événements pose le problème du temps et de sa représentation. Toutes les informations ainsi considérées sont entachées d'un niveau d'incertitude et d'imprécision, voire de contradiction. Les informations sont issues de sources historiques dont la crédibilité est primordiale. Ces concepts font parfois l'objet d'une modélisation. Nous présentons les principaux modèles conceptuels prosopographiques, inclus les approches ontologiques. La conclusion ouvre vers les axes de recherche à poursuivre dans ce domaine.

**ABSTRACT.** Prosopography aims to study similarities in the trajectories of people in their biography. It is based on a database containing information relating to people from a specific environment defined chronologically and geographically. The objective of this article is to propose a state of the art of prosopographic databases emphasizing its specific concepts and the research relating to the modeling of these concepts. We present different definitions of prosopography. Then we describe the prosopographic method which has undergone a certain evolution due to the use of computerized databases. An overview of these prosopographic databases is presented, highlighting their advantages and limitations. The main concepts of these systems are described, including the person who is the central concept, the onomastics which allows the study of the names of people and institutions and the biographical notes which list the events to which people are linked. The description of these events raises the problem of time and its representation. All the information thus considered is associated with a level of uncertainty and imprecision, even contradiction. The information comes from historical sources whose credibility is essential. These concepts are sometimes modeled. We present the main conceptual prosopographic models, including ontological approaches. The conclusion opens towards the lines of research to be pursued in this field.

**KEYWORDS.** prosopography, databases, conceptual modeling, ontology, historical data sources, event, time, uncertainty.

**MOTS-CLÉS.** Prosopographie, bases de données, modèle conceptuel, ontologie, source d'information historique, événement, temps, incertitude.

### 1. Introduction

La prosopographie est une démarche du champ historique relative à l'étude des caractéristiques communes d'un groupe d'acteurs historiques au moyen d'une analyse collective de leurs trajectoires [Stone, 1971]. Le terme *prosopographie* provient des mots grecs *prosopon* (personne) et *graphein* (description). Il apparaît pour la première fois au XVIII<sup>e</sup> siècle et désigne alors une liste d'individus. Plus tard, les historiens l'utilisent pour désigner la mise en série de notices biographiques.

Bien que la prosopographie ne concerne pas uniquement une catégorie sociale [Charle, 2001], elle centre néanmoins son analyse sur l'individu en fonction de la catégorie sociale dont il fait partie. Les données biographiques décrivent les caractères d'une population dont les membres partagent un ou plusieurs traits communs, comme le métier, l'origine sociale ou géographique, la formation, etc. Le but est de mettre en évidence les caractéristiques communes d'un groupe d'acteurs à une époque déterminée à partir de données biographiques.

De nature essentiellement descriptive, la prosopographie cherche à déterminer la structure sociale d'un collectif. A cet effet, elle procède à l'accumulation de données structurées sous la forme de fiches individuelles relatives à chacun de ses membres. Il existe une variété de vocabulaires qui se rattachent à la prosopographie, tels que les études de carrières, de trajectoires, de cycles de vie [Hyvönen, 2020], de biographies collectives, de groupes, de masse, d'histoires de vie et d'études longitudinales [Lemercier & Picard, 2012]. La prosopographie permet d'étudier des trajectoires individuelles dans leur complexité et leurs interactions. Elle permet aussi de tester des hypothèses sur une population donnée.

A la différence des études biographiques, l'étude prosopographique suppose la définition préalable de la population qui fait l'objet de l'étude. Certains auteurs définissent la prosopographie comme étant, « la recherche des éléments communs et des écarts différentiels présentés par des biographies particulières » [Maurin, 1982]. Les premières études prosopographiques concernent principalement l'histoire romaine. Un exemple en est la *Prosopographia Imperii Romani*<sup>1</sup>.

La prosopographie s'appuie généralement sur une base de données contenant des informations relatives à des personnes d'un milieu spécifique définies chronologiquement et géographiquement [Bulst, 1991]. Les historiens étudient généralement de grands groupes d'individus décrits dans des documents incomplets. La fiabilité et la qualité de la documentation d'origine (démographique, économique, administrative, religieuse, familiale, etc.) sont cruciales. De plus, les historiens sont confrontés à la rareté relative des sources. Représenter le temps et les dimensions d'incertitude liées aux personnes, aux lieux, aux factoïdes (francisation du terme anglais factoid) et aux sources, constitue une vraie difficulté. La prosopographie est confrontée à des informations souvent incomplètes, imprécises et contradictoires. Elle suit des méthodes et des démarches qui lui sont propres.

L'approche prosopographique s'est incontestablement développée grâce à l'apport de l'informatique, et plus particulièrement grâce aux bases de données. Réussir à traiter un nombre très important de fiches individuelles et répondre aux questions des prosopographes requiert le recours à la démarche et à l'outil informatique. Différents outils ont été utilisés depuis de nombreuses années, depuis des tableurs, des fichiers classiques jusqu'aux bases de données. On assiste à une évolution importante concernant ces bases de données. Des bases simples comme Foxbase, 4ème Dimension, et FileMaker ont d'abord été utilisées. Aujourd'hui, la tendance générale est l'utilisation des bases de données relationnelles. Plus récemment, des projets prosopographiques utilisent des bases de données XML ou JSON. Toutes ces bases doivent permettre de faire face à des défis tels que la gestion de l'incertitude, la crédibilité des sources et le passage à l'échelle.

*L'objectif de cet article est de proposer un état de l'art des bases de données prosopographiques mettant l'accent sur les concepts spécifiques de la prosopographie et la recherche afférente à la modélisation de ces concepts.*

Le reste de l'article est organisé comme suit. La section 2 est dédiée à la description de la démarche prosopographique. Nous dressons dans la section 3 un panorama de quelques bases de

---

<sup>1</sup> <http://pir.bbaw.de/>

données prosopographiques représentatives. Une analyse de leurs avantages et de leurs limites y est présentée. Les principaux concepts utilisés dans la modélisation des bases de données prosopographiques sont décrits et discutés dans la section 4. La section 5 présente les principaux modèles conceptuels des bases de données prosopographiques, incluant les principales ontologies. Enfin, la section 6 conclut l'article et propose des pistes de recherche future.

## 2. Méthode prosopographique

La méthode prosopographique a pour objectif principal l'étude de similitudes de trajectoires des personnes dans leur parcours biographique. Les objectifs de recherche les plus courants sont [Verboven *et al.*, 2007] : l'étude de la stratification sociale, la mobilité sociale, les processus de prise de décision, le fonctionnement ou le dysfonctionnement des institutions. Plus généralement, la méthode prosopographique cherche à déterminer les facteurs généraux qui aident à expliquer la vie des individus. Les cinq étapes qui caractérisent la méthode prosopographique impliquant la création d'une base de données sont [De Ridder-Simoens, 1991] :

1. la définition de la question qui permet d'indiquer ce qui est recherché et ce qui est à mettre en exergue. Le type de question détermine le type de données biographiques ;
2. la définition du groupe étudié tant au niveau géographique que chronologique et thématique ;
3. l'identification des sources disponibles et leur analyse critique ;
4. l'établissement d'un modèle de fiche ou notice, destiné à répertorier les données biographiques de chaque personne du groupe étudié ;
5. le regroupement des données recueillies ainsi que leur interprétation dans le cadre d'un contexte.

On assiste à une évolution des méthodes auxquelles la recherche prosopographique a recours depuis la fin du 20e siècle, notamment grâce au développement des bases de données informatisées. Ces dernières permettent la structuration autour d'événements et de liens. Certaines bases sont structurées en «épisodes». Ces derniers représentent des éléments de carrière caractérisés par l'identifiant des personnes concernées, des dates ou encore des lieux. L'apport de l'informatique dans les méthodes en histoire a permis à la prosopographie de réaliser les premières études en histoire médiévale et moderne [Loriga, 2010 ; Fossier & Genet, 1983]. Les bases de données relationnelles, en particulier, ont joué un rôle important en permettant de reconstituer des carrières individuelles situées dans des phénomènes collectifs [Dedieu, 2004 ; Dedieu & Chaparro Sainz, 2013].

La numérisation, et notamment les techniques de visualisation, permet de relier les fiches prosopographiques à des graphes relationnels ou à des systèmes d'information géographiques. En termes de traitement, certaines techniques permettent d'aller au-delà des comptages ou des typologies qui constituent l'essentiel des opérations réalisées dans la plupart des projets prosopographiques actuels ou passés. Parmi ces techniques, citons : l'analyse factorielle des correspondances, la classification automatique, la régression multivariée, l'analyse de réseaux, et l'analyse de séquences.

La méthode de recherche prosopographique applique les étapes suivantes [Verboven *et al.*, 2007 ; Lemercier & Picard, 2012] :

1. détermination de la population cible, incluant ses limites géographiques, chronologiques et thématiques. Un axe pivot et des critères d'inclusion et d'exclusion peuvent faciliter la définition du périmètre ;

2. formulation des hypothèses de travail spécifiques et des questions historiques sur la population cible, conduisant à un questionnaire spécifique ;
3. conception d'une base de données prosopographique ;
4. alimentation de la base de données en utilisant les sources primaires et la littérature ;
5. interrogation et analyse des données de la base ;
6. synthèse des résultats, par combinaison et interprétation de ces données, en analysant les sources et la littérature, en regard du contexte historique ;
7. présentation des résultats.

Ainsi, la méthode prosopographique s'appuie sur une base de données comprenant la description de la population ciblée. Grâce aux technologies de l'information, les historiens ont largement recours aux systèmes de gestion de bases de données. Dans la suite, nous décrivons ces bases de données prosopographiques.

### 3. Les bases de données prosopographiques

Nous présentons dans cette section un panorama de quelques bases de données prosopographiques représentatives et nous dressons un état des avantages et des limites qui les caractérisent.

#### 3.1. Panorama des bases de données prosopographiques

L'avènement de l'informatique dans le domaine de la recherche en histoire a offert de nombreuses possibilités en termes de stockage et d'interrogation d'informations. L'apparition des ordinateurs personnels couplée à l'amélioration des techniques de saisie, de stockage et d'interrogation a favorisé l'émergence des bases de données prosopographiques. Plusieurs répertoires recensent des projets prosopographiques et les bases de données associées. Citons par exemple celui mis à disposition par l'université d'Oxford qui recense de nombreux projets répartis sur trois périodes : ancienne, médiévale et moderne<sup>2</sup>.

A titre d'exemple, mentionnons le projet COL&MON<sup>3</sup> (Collégiales et monastères de la réforme carolingienne au Concile de Trente (816-1563)). Les bases de données constituées par ce projet doivent permettre la reconstitution des frontières diocésaines et fournir des données moissonnables.

Récemment est également apparu un projet EquipEX dirigé par la Bibliothèque nationale de France (BnF) et l'Institut de recherche et d'histoire des textes (IRHT), Biblissima<sup>4</sup>, dont l'objet est de donner accès au patrimoine écrit du Moyen Âge conservé ou connu. Pour réaliser ce projet, Biblissima utilise une ontologie qui lui permet de mettre en relation d'importants gisements de données sur les œuvres, les lieux et les personnes.

Les projets Prosopography of Anglo-Saxon England (PASE)<sup>5</sup> et Prosopography of the Byzantine World (PBW)<sup>6</sup> sont deux autres exemples d'initiatives s'inscrivant dans ce contexte de diffusion et

---

<sup>2</sup> <https://prosopography.history.ox.ac.uk/directory.htm>

<sup>3</sup> <https://colemon.huma-num.fr/>

<sup>4</sup> <http://www.biblissima-condorcet.fr/fr/projet/presentation>

<sup>5</sup> <http://pase.ac.uk/>

de partage des connaissances auprès d'une communauté. Ces projets reposent sur une approche de la prosopographie structurée (base de données relationnelles et ontologie) proposée par l'équipe du Centre for Computing in the Humanities du Department of Digital Humanities (CCH / DDH) du King's College de Londres, fondée sur le modèle factoïde. Le projet PASE est développé par le CCH / DDH en association avec le Département d'Anglo-Saxon, Norvégien et Celte de l'Université de Cambridge. PASE a produit une base de données qui fournit des informations structurées sur tous les habitants recensés de l'Angleterre de la fin du VIe au XIe siècle. Elle repose sur un examen systématique des sources écrites disponibles pour la période. La prosopographie du monde byzantin (PBW) est un projet visant à créer une base de données prosopographique d'individus nommés dans des sources textuelles de l'empire byzantin et des régions avoisinantes entre 642 et 1265. Le projet consiste en une collaboration entre l'Académie britannique et l'Académie des sciences et des sciences humaines du Brandebourg.

Studium Parisiense<sup>7</sup> est une base de données en ligne développée dans le cadre d'un Advanced Project de l'ERC (European Research Council). Elle concerne les maîtres, les étudiants et les suppôts des écoles et de l'université de Paris sur la période XIe-XVIe siècle. Actuellement implantée comme une base de données semi-structurées XML, elle est en cours de migration vers une base de données graphe afin de bénéficier d'un langage d'interrogation puissant permettant l'extraction semi-automatique d'informations.

La Digital Prosopography of the Roman Republic (DPRR)<sup>8</sup> est le résultat d'un projet de trois ans du King's College de Londres financé par Arts and Humanities Research Council (AHRC). L'objectif principal du projet est de faciliter les recherches prosopographiques sur l'élite de la République romaine, sa structure, son échelle et l'évolution de sa composition. À cette fin, une base de données complète et consultable de tous les membres connus de la société romaine des couches supérieures a été créée à partir de quatre sources de données existantes. Elle rassemble des informations sur les carrières individuelles, les fonctions exercées, le statut personnel, les dates de vie et les relations familiales. Toutes les données créées par l'équipe du projet DPRR sont exprimées explicitement dans le format RDF, standard du Linked Open Data.

Le projet Mamluk Prosopography<sup>9</sup> (MPP – 2016/2020) est une base de données sur les sciences humaines en libre accès, destinée à l'étude des groupes sociaux d'élite, des réseaux et des pratiques socioculturelles du sultanat syro-égyptien de la fin du Médiéval (XIIIe-XVe siècles). Ce projet ambitionne le développement d'une structure de base de données relationnelle complexe qui permet la migration, la mise à niveau et la mise à jour de jeux de données préexistants, la création d'applications Web modulaires pour les utilisateurs et la mise en œuvre d'outils d'exploration de données et de visualisation pour des analyses sociales, temporelles, spatiales et sémantiques.

Enfin, China Biographical Database (CBDB)<sup>10</sup> est une base de données relationnelle librement accessible comprenant des informations biographiques sur environ 427000 personnes, principalement du VIIe au XIXe siècle.

---

<sup>6</sup> <https://pbw2016.kdl.kcl.ac.uk/>

<sup>7</sup> <http://lamop-vs3.univ-paris1.fr/studium/>

<sup>8</sup> <http://romanrepublic.ac.uk/>

<sup>9</sup> <https://research.flw.ugent.be/nl/projects/mpp-mamluk-prosopography-project-project-specific-database-open-access-digital-humanities>

<sup>10</sup> <https://projects.iq.harvard.edu/cbdb>

La liste précédente avait pour objectif de montrer la diversité des projets et des types de bases de données existant dans le monde de la recherche historique. Pour autant, ces réalisations présentent des caractéristiques communes que l'on peut qualifier d'avantages ou de limites comme décrit ci-après.

### **3.2. Avantages et limites des bases de données prosopographiques existantes**

Les données prosopographiques sont très structurées (nom, date de naissance, lieu d'un fait, etc.), ce qui les rend particulièrement adaptées à un stockage dans une base de données structurées (relationnelle) ou semi-structurée. [Mathisen, 1988] recensait déjà une liste des avantages inhérents à une base de données dédiée à la prosopographie : vitesse d'accès à l'information, précision de l'information, multiplicité et diversité des accès, facilité d'accès, de révision et de mise à jour, portabilité, compatibilité avec d'autres bases biographiques et prosopographiques, facilité à générer des rapports.

Pourtant le succès de ces bases ne fut pas immédiat. Ainsi, R. Mathisen présente vingt projets de bases prosopographiques en cours de création [Mathisen, 1988]. Plus d'un an après, seule la base Florentine Catasto est achevée, une autre a été intégrée dans un projet futur (Religious Woman), deux étaient encore en développement (Vienna Family et Late Antique Database) et les autres étaient abandonnées. Si la majorité des avantages liés à l'utilisation d'une base de données cités ci-dessus ont pu être vérifiés dans les faits, la raison principale de ces échecs découle des nombreuses difficultés qui freinent le développement des bases de données prosopographiques. Les principales sont :

1. la difficulté pour les développeurs de la base de données à concevoir un modèle qui puisse capturer l'ensemble des informations, structurées ou non, présentes dans les données prosopographiques par nature très riches ;
2. la base de données n'est qu'un outil, point de départ pour la recherche. Les interventions humaines sont toujours nécessaires, non seulement lors de la création de la base, mais surtout lors de son utilisation. Cela inclut non seulement la vérification de la validité et de l'adéquation des données obtenues, mais également une analyse judicieuse de ces données ;
3. l'utilisation de la base, pour créer ou interroger les données, doit être possible sans connaissances informatiques particulières ;
4. la création d'une base de données ne permet pas de résoudre les problèmes d'onomastique, qui exigent une expertise humaine. Détecter automatiquement des données d'individus « proches » dans une base de données prosopographique est aujourd'hui encore un problème difficile ;
5. l'absence de standard lorsqu'on crée une base de données prosopographique limite les échanges possibles entre différentes bases.

Chaque projet prosopographique tend à inclure la conception et le développement d'une base de données dédiée, requérant un effort et des moyens considérables. Même en cas de succès, la maintenance d'une telle base est problématique. Comme expliqué plus haut, la méthode prosopographique comporte une phase de modélisation qui inclut une représentation des principaux concepts communs à tous les projets prosopographiques. Ces concepts sont décrits ci-après.

## **4. Les principaux concepts de la modélisation prosopographique**

Le but de cette section est de présenter les principaux concepts utilisés dans la modélisation des bases de données prosopographiques. La personne est le concept central, conduisant à la problématique de son identification [François et Muller, 2015]. Une science connexe,

*l'onomastique*, est dédiée à l'étude des noms des personnes et des institutions. Les notices biographiques recensent les *événements* auxquels les personnes sont liées d'une façon ou d'une autre. La description de ces événements pose le problème du *temps* et de sa représentation. Toutes les informations ainsi considérées sont entachées d'un niveau *d'incertitude* et d'imprécision, voire de contradiction. Les informations sont issues de *sources* historiques dont la crédibilité est primordiale [Prost, 1996].

#### 4.1. *L'onomastique*

L'onomastique concerne l'étude des noms propres, qu'il s'agisse de personnes, d'institutions ou de lieux. L'historien est confronté à l'orthographe des patronymes, qui peut être très variable dans les documents anciens. Un patronyme sera parfois orthographié différemment pour un même individu ou une même famille dans la rédaction d'un acte. L'orthographe des patronymes peut être considérée comme relativement arbitraire, puisqu'elle résulte généralement de la fixation d'une prononciation à un moment donné dans un endroit particulier.

L'ISO 3166-2:2007 fournit un code d'application universelle pour la représentation des noms des principales divisions administratives des pays et territoires figurant dans l'ISO 3166-1.

Les travaux de l'onomastique s'appuient sur le traitement de la langue naturelle et bénéficient des travaux relatifs aux entités nommées [Hengchen *et al.*, 2015].

L'étude des noms propres intègre une analyse commune à toutes les langues et une partie spécifique à une langue donnée. Des outils, comme les ontologies, sont élaborés pour prendre en compte cette complexité, par exemple Prolexbase [Tran, 2006].

L'onomastique et la prosopographie sont des outils interdépendants pour l'historien et très proches de la généalogie [Keats-Rohan, 2007]. Les systèmes de noms ne sont pas statiques, la généalogie est requise pour suivre l'évolution de leur orthographe. Certains projets combinent la recherche onomastique et la prosopographie, par exemple le projet *Lexicon of Greek Personal Names*<sup>11</sup>.

Les principales difficultés dans l'étude des noms sont liées notamment 1) aux lacunes inhérentes aux documents historiques (certaines portions de texte ayant pu disparaître), 2) à l'homonymie notamment dans certains milieux où la variété des noms est faible, 3) aux ambiguïtés terminologiques, liées à la polysémie [Carrié, 2003].

Dans les bases de données prosopographiques, les historiens sont amenés à identifier de manière unique les personnes et à enregistrer les différentes variantes des noms de ces individus dans les différentes sources les mentionnant. A titre d'exemple, dans *Studium Parisiense*<sup>12</sup>, un nom principal a été retenu, parfois *Anonymous*, et des variantes de ce nom sont aussi accessibles.

#### 4.2. *Les événements*

Un événement est « quelque chose qui arrive », planifié ou non et d'intérêt. Les informations relatives à cet événement doivent être décrites. Un événement peut être défini comme un sextuple  $\langle A, O, T, V, P, L \rangle$  où A est l'action, O les objets (acteurs ou entités), T la période pendant laquelle se déroule l'événement, V l'environnement de l'événement, P des assertions décrivant l'événement (préconditions, post-conditions, conditions intermédiaires) et L des expressions de langage décrivant l'événement [Liu *et al.*, 2010].

<sup>11</sup> <http://www.lgpn.ox.ac.uk/>

<sup>12</sup> <http://lamop-vs3.univ-paris1.fr/studium/>

En prosopographie, il est crucial de pouvoir identifier les événements afin de les relier à d'autres événements, personnes, lieux, périodes et documents. [Westermann & Jain, 2007] distinguent différents types d'événements, discrets ou continus. Un certain nombre de normes ont été élaborées pour représenter les événements [Shaw & Larson, 2008].

EventsML-G2<sup>13</sup> est un modèle standard pour représenter les événements et faciliter leur échange et leur partage dans le secteur des médias. Il permet de représenter tous les faits liés à un événement. Dans EventsML-G2, les heures d'événement sont spécifiées en tant que plages de dates. Cependant, les événements qui ont une heure de début et aucune heure de fin sont considérés comme instantanés.

Dans le CIDOC-CRM<sup>14</sup>, les événements ont des durées, et ces dernières peuvent être nulles pour les événements instantanés. Dans certains cas, il peut être nécessaire de pouvoir exprimer des temps relatifs (pour les événements qui se produisent avant, pendant ou après une période de temps ou un autre événement). La plupart des normes ne supportent pas explicitement l'expression de relations temporelles relatives.

[Shaw & Larson, 2008] comparent ainsi sept standards de représentations des événements. Les standards permettent aussi de relier les événements entre eux par des relations temporelles mais aussi d'associer des événements à des lieux géographiques ou des juridictions politiques.

Plusieurs ontologies décrivant des « événements » ont été proposées. Ces ontologies ont des concepts similaires. Ce sont principalement des ontologies de type OWL (Web Ontology Language) qui est un langage de représentation des connaissances construit sur le modèle de données de RDF (Resource Description Framework). C'est le cas de CIDOC-CRM [Crofts *et al.*, 2003] qui constitue une « ontologie » de l'information relative au patrimoine culturel.

D'autres ontologies centrées événement ont été proposées, notamment ABC [Lagoze & Hunter, 2001] qui, à l'instar de CIDOC-CRM [Doerr, 2003], a comme objectif de permettre l'interopérabilité des formats de métadonnées permettant de décrire des contenus multimédia des musées et des bibliothèques. L'ontologie Event EO [Raimond & Abdallah, 2007] constitue l'ontologie musicale la plus utilisée notamment dans la communauté Linked Data. L'ontologie SEM [van Hage *et al.*, 2011] représente des attaques maritimes dans des lieux parfois imprécis. DOLCE+DnS Ultralite (DUL) combine une axiomatisation simplifiée des concepts de DOLCE avec les patrons de représentation de Descriptions and Situations [Gangemi & Mika, 2003]. L'ontologie F Event est fondée sur DUL.

### 4.3. Le temps

L'un des problèmes auxquels est confronté tout historien est la question de la représentation des données temporelles. L'expression du temps peut prendre plusieurs formes. Cette forme peut être : (i) précise (par exemple le concert a lieu le 1/12/2017 à 20 heures), (ii) vague (par exemple j'ai discuté avec elle hier). Selon le niveau de granularité, le caractère vague peut disparaître comme dans l'expression « je vais à la piscine trois fois par semaine ». Par conséquent le temps (ou la manière de l'exprimer) peut être une source d'imprécision et d'incertitude.

Dans tout travail prosopographique, le chercheur est confronté à deux types de données : des données « uniques » qui s'expriment par une date et des données évolutives qui se traduisent par un intervalle de dates [Bernaudeau, 2008]. Ce type de données est particulièrement applicable à des

<sup>13</sup> <http://www.iptc.org/EventsML/>

<sup>14</sup> <http://www.cidoc-crm.org/>

« états », comme l'état civil, qui varient au cours du temps. Il existe aussi des données incomplètes qui se présentent sous la forme de dates approximatives, ou contradictoires, auxquelles on associe généralement un degré de confiance qui dépend de la fiabilité de la source.

Les modèles temporels considèrent deux types de données. Le premier type est "l'instant" qui permet d'identifier le moment d'origine d'un événement. Le second type est "l'intervalle" qui identifie la durée de l'événement. Il existe plusieurs modèles temporels, notamment l'ISO 19108: 2002, qui se concentre sur les concepts temporels dans le contexte des systèmes d'information géographique. [Allen & Ferguson, 1994] proposent un modèle temporel fondé sur des intervalles de temps, modélisés par une paire de points temporels ordonnés, le premier point étant inférieur au second.

Les bases de données intègrent l'aspect temporel en s'appuyant sur un modèle de données temporel qui a conduit à une version spécifique du langage SQL (TSQL2) [Snodgrass, 2012]. Le modèle temporel MADS [Parent *et al.*, 1998] prend en compte les relations temporelles entre intervalles en considérant l'algèbre relationnelle d'Allen [Allen, 1983]. Le modèle temporel de Perceptory [Bédard *et al.*, 2004] suit deux principes fondamentaux : l'existence et l'évolution.

Le modèle temporel dans AROM-ST [Moisuc *et al.*, 2005] propose plusieurs types de temps, notamment les types instantané, intervalle, multi-instants et multi-intervalles. Ces derniers sont des types composés, qui constituent des collections d'instants ou d'intervalles.

Plusieurs approches ont été proposées pour la modélisation du temps sous la forme conceptuelle Entité-Relation : TERM, RAKE, MOTAR, etc. [Gregersen & Jensen, 1999]. Les auteurs fournissent un aperçu complet de ces modèles ER temporels ainsi que leur comparaison.

Les premiers travaux qui ont souligné l'importance du concept de temps dans les ontologies ont été initiés par la communauté web sémantique [Bry & Marchiori., 2005]. Diverses approches ont été proposées pour représenter l'information temporelle dans RDF [Manola *et al.*, 2004] et OWL [McGuinness & van Harmelen, 2004]. L'ontologie OWL-Time [Hobbs & Pan, 2004], développée au sein du consortium W3C, est consacrée aux concepts et aux relations temporelles telles que définies dans la théorie d'Allen. Par exemple, dans leur ontologie, en ce qui concerne les assertions temporelles imprécises, [Matousek *et al.*, 2007] proposent la catégorisation suivante :

1. assertions précises (par exemple, le 12 décembre 2019 à 12H30),
2. assertions avec une granularité plus élevée (par exemple, le 12 décembre 2019 qui peut être un instant ou une période de 24 heures),
3. assertions incomplètes (par exemple, le 12 janvier à 12H30 mais l'année est omise),
4. assertions incertaines avec une spécification absolue de l'incertitude (par exemple, entre le 12 janvier 2019 et le 13 janvier 2019),
5. assertions incertaines avec une spécification relative de l'incertitude (par exemple, aux alentours du 12 janvier 2019),
6. assertions faisant référence à d'autres assertions contenant des propriétés temporelles (par exemple, durant le règne de François 1<sup>er</sup>),
7. assertions avec des informations inconnues ou manquantes (par exemple, à cette époque).

Ainsi, dans la plupart de ces catégories, le temps est entaché d'incertitude.

#### 4.4. L'incertitude

Plusieurs conceptions de l'incertitude existent au sein des disciplines de sciences sociales. Une description détaillée de ce concept est offerte par [Wakeham, 2015]. Comme [Tannert *et al.*, 2007], il différencie l'incertitude subjective de l'incertitude objective. L'incertitude subjective se produit lorsqu'un individu éprouve des doutes sur le résultat possible de certaines actions. L'incertitude objective suppose un monde qui ne soit connu que dans une certaine mesure. Dans cette dernière catégorie, on distingue l'incertitude épistémologique où quelque chose qui est potentiellement connaissable reste inconnu, de l'incertitude ontologique où quelque chose reste inconnu parce qu'il est inconnaisable [van Asselt & Rotmans, 2002]. Notons que le concept d'incertitude est souvent associé (sinon confondu) à d'autres termes comme le risque [Taleb, 2010], l'ambiguïté [Dequech, 2000] ou l'ignorance [Smithson, 2009].

L'incertitude est généralement considérée comme une mesure subjective de la certitude de quelque chose. Un état d'incertitude est créé lorsqu'il est impossible de déterminer si une information est vraie ou fausse. A titre d'exemple, citons l'expression, « L'archevêque peut avoir 92 ans », qui laisse un doute sur l'âge de l'archevêque. Le National Research Council des États-Unis définit l'incertitude comme « un concept général qui reflète notre manque de certitude à l'égard de quelque chose ou de quelqu'un, allant d'un manque de certitude à un manque presque total de conviction au sujet d'un résultat [Crofts *et al.*, 2003]. L'incertitude indique un manque de confiance dans un objet, dans un événement ou dans une personne.

Un état de l'art sur les théories et les pratiques dans la gestion de l'incertitude est présenté dans [Zio & Pedroni, 2013]. [Li *et al.*, 2012] décrivent quatre types de méthodes de gestion de l'incertitude. La première, la théorie des probabilités, est la méthode la plus largement utilisée, en particulier pour l'incertitude ontologique. La deuxième méthode, celle de la logique floue, est dédiée à la théorie épistémologique. Elle traite principalement de l'incertitude due à l'imprécision (vagueness) découlant des étiquettes linguistiques de la langue naturelle (human linguistic terms). Elle s'appuie sur des classes avec des limites mal définies, comme par exemple *chaud*, *frais*, *froid* et *brûlant*. La troisième théorie, théorie de l'écart-info (info-gap), appartenant à la catégorie de l'incertitude épistémologique, reflète l'écart d'information entre ce que l'on sait et ce que l'on a besoin de savoir. La dernière théorie, théorie de l'incertitude dérivée, traite de l'incertitude épistémologique subjective de l'homme.

Il existe de nombreuses classifications du concept d'incertitude [Thunnissen, 2003]. Chacune met un accent particulier sur un aspect de l'incertitude liée à un domaine. Par exemple, en sciences de gestion, l'incertitude est classée en aléatoire, épistémique, paramètre, modèle et volitif (volitional) [Bedford & Cooke, 2001].

Smithson propose une classification de l'incertitude fondée sur une taxonomie de l'ignorance [Smithson, 1989]. Il distingue deux types d'ignorance : l'erreur et l'inapplicabilité. Il considère l'incertitude comme un type d'erreur. Le concept d'incertitude est un type particulier d'incomplétude. Cette classification a inspiré plusieurs autres classifications, en particulier celle d'URREF [Costa *et al.*, 2012], dans laquelle le terme incertitude englobe une variété d'aspects de la connaissance imparfaite, y compris l'ambiguïté, l'incomplétude, l'imprécision (vagueness), le caractère aléatoire et l'incohérence. L'ambiguïté caractérise une situation où l'information manque de sémantique complète. C'est le cas lorsqu'on affirme que, "Il est indéterminé si le chiffre d'affaires est mensuel ou trimestriel". Par conséquent, l'ambiguïté peut être définie comme une expression qui a plus d'un sens. L'incomplétude reflète un manque d'information. L'imprécision (vagueness) correspond à une situation caractérisée par une connaissance incomplète des faits et des événements considérés. L'expression, "l'archevêque est dans ses premières années de sacerdoce" en est un exemple. L'aléatoire caractérise l'absence d'ordre ou de patron (pattern). Généralement, le caractère

aléatoire exprime l'absence de prévisibilité d'un événement. Il peut être considéré comme représentant la variabilité des phénomènes naturels observés. L'incohérence caractérise une situation où deux ou plusieurs informations ne peuvent pas être vraies en même temps. Par exemple, "l'archevêque est âgé de 85 et de 93 ans".

Presque toutes les données prosopographiques contiennent des types d'incertitude. Les données en entrée peuvent être ambiguës ou peu claires. La connaissance des événements peut être discutable et peut prendre la forme de plausibilité, de probabilité et / ou de possibilité.

#### 4.5. Les sources historiques

Les historiens fondent leurs travaux sur la base de sources d'information. Ces dernières sont de trois types : les sources primaires, les sources secondaires et les sources tertiaires. Une source primaire est un document original donnant des informations directes sur le sujet de la recherche. Ce document date le plus souvent de l'époque où le ou les événements rapportés ont eu lieu. Les sources primaires sont des lettres, des textes, des images, des rapports, des registres, etc.

Les sources secondaires désignent des documents qui utilisent des sources primaires, et souvent la consultation d'autres sources secondaires, dont ils constituent une analyse, une synthèse, une explication ou une évaluation. Les biographies ou les ouvrages de recherche en Histoire sont des exemples de sources secondaires.

Une source tertiaire est une sélection et une compilation de sources primaires et secondaires. Les bibliographies, les catalogues de bibliothèques, les répertoires, les listes de lectures conseillées et les articles proposant des tours d'horizon sont des exemples de sources tertiaires.

La qualité des analyses historiques dépend directement de la qualité de ces sources d'information. Cette qualité recouvre de nombreux aspects, comme l'a montré [Zaveri *et al.*, 2016] dans le cadre des données liées ouvertes. Étendant le modèle de [Strong *et al.*, 1997], la qualité comprend les quatre dimensions : la qualité intrinsèque, la qualité de la représentation, la qualité contextuelle, l'accessibilité mais inclut aussi la confiance et le caractère dynamique. Ces six dimensions ne sont pas exclusives l'une de l'autre. Ainsi l'actualité (timeliness) est fonction du contexte et de la confiance.

Plus généralement, la crédibilité d'une source d'information historique se définit comme un degré de confiance accordée par les historiens en la capacité de cette source à fournir une information fiable [Akoka *et al.*, 2020]. À l'heure où les historiens disposent de plus en plus de sources numérisées, la mesure automatique de cette crédibilité est un enjeu majeur.

La crédibilité d'une source dépend de plusieurs facteurs. Avant tout, elle dépend de l'auteur de cette source : est-il connu comme fiable et précis dans sa production ? Cela est souvent lié à sa place dans la société (rang, classe, fonction, etc.), la date de production de la source par rapport à la date de l'événement relaté, la motivation pour l'auteur à produire cette source, ses convictions personnelles, etc. Cela dépend aussi de la nature du fait décrit : s'agit-il de faits rapportés ou de faits observés/constatés par l'auteur ? La crédibilité de la source peut être attestée directement par des experts du domaine, ou parce qu'elle est citée dans d'autres sources considérées comme sérieuses.

La crédibilité d'une source secondaire dépend de son auteur : est-il reconnu comme fiable par les experts du domaine ? Quelle est la thèse qu'il cherche à défendre et reste-t-il objectif ? Quelle est la crédibilité des sources qu'il cite ? Cette source secondaire est-elle citée souvent, par des experts reconnus ?

Il n'existe pas de méthode systématique d'évaluation de la crédibilité des sources historiques. En revanche, on trouve beaucoup de guides méthodologiques à l'usage des chercheurs ou des étudiants

en histoire. Ces guides abordent, entre autres, le sujet de la qualité des sources historiques et fournissent des directives utilisables dans le processus d'évaluation. Par exemple, l'université du Kentucky propose une démarche d'évaluation des sources historiques qui diffère selon que ces dernières sont primaires ou secondaires [UKY, 2019]. L'évaluation d'une source primaire (manuscrit, lettre, journal intime, mémo, autobiographie, etc.) s'appuie sur l'auteur, l'audience, la logique, le cadre de référence, le lien avec d'autres sources, etc. L'évaluation d'une source secondaire (livre historique, article de recherche historique, etc.) est fondée sur l'analyse de la structure, de la thèse défendue, de l'argumentation, etc. Le terme le plus utilisé est celui de fiabilité (en anglais *reliability* ou *trustworthiness*). Là encore, on trouve des aides à l'évaluation de la crédibilité comme OPCAM pour Origin, Perspective, Context, Audience et Motive ou TOMACRU pour Type, Origin, Motive, Audience, Content, Reliability et Usefulness [Vest, 2007].

La crédibilité d'une information est un concept dont la définition n'est pas standardisée. Le terme lui-même possède plusieurs synonymes et concepts voisins, notamment la réputation, l'objectivité et la vérifiabilité. [Akoka *et al.*, 2020] ont contribué à affiner ces définitions dans le contexte des informations historiques. Les auteurs proposent un modèle conceptuel regroupant les éléments principaux décrivant les sources d'information et les aspects pouvant contribuer à l'évaluation de leur crédibilité. Il regroupe à la fois les concepts utilisés pour l'analyse des sources historiques (intention, perspective, audience, contexte, auteur ou origine) et les mesures de qualité de l'information (crédibilité d'une source, réputation d'un auteur) qui peuvent en être déduites. Ce modèle est utilisé comme fondement pour définir une approche automatique d'évaluation de la crédibilité.

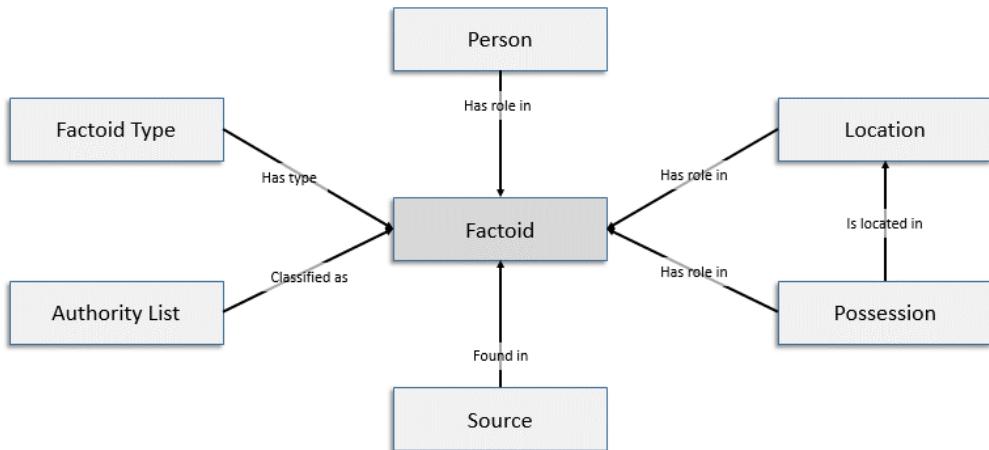
Les bases de données prosopographiques décrites plus haut ont été construites suite à une modélisation logique (relationnelle, XML, etc.). Les différents concepts formant le socle de ces bases de données prosopographiques ont été décrits ainsi que les approches ontologiques afférentes. L'intégration de ces différents concepts permet de construire des modèles conceptuels permettant une vision abstraite plus riche de la prosopographie.

## 5. La modélisation des bases de données prosopographiques

On distingue généralement trois niveaux d'abstraction dans la modélisation des données. Le premier niveau concerne la modélisation conceptuelle qui est indépendante des architectures matérielles et logicielles utilisées. Le second niveau est relatif à la modélisation logique qui intègre les architectures logicielles. Enfin, le dernier niveau d'abstraction correspond à la modélisation physique qui comprend les architectures logicielles et matérielles inclus les préoccupations de performance nécessitant le choix et la gestion des index et des « clusters ».

Les projets prosopographiques ont recours généralement à la phase de modélisation logique, le plus souvent en adoptant le modèle relationnel. Grâce à ce dernier, les historiens disposent de moyens pour représenter les objets qui composent l'univers prosopographique. Studium Parisiense s'appuie sur un modèle logique simple qui permet la saisie de textes libres dans des champs peu structurés. Il rend ainsi aisées la saisie et la recherche en plein texte. Toutefois, les technologies liées au big data laissent entrevoir des opportunités nouvelles pour construire des bases de données notamment graphes ou documents, potentiellement plus adaptées au contexte de la prosopographie. Dans cette optique, il est important de passer en revue les modèles conceptuels prosopographiques existants.

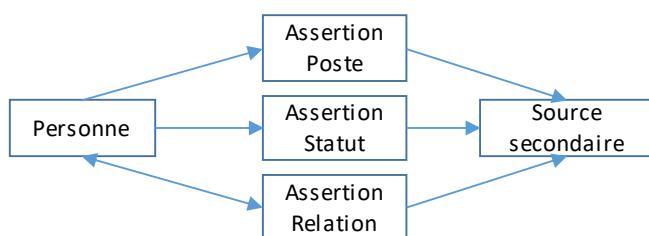
Le modèle conceptuel prosopographique le plus courant est celui du factoïde [Bradley & Short, 2002 ; Pasin & Bradley, 2013]. Un factoïde est un endroit dans une source qui dit quelque chose sur une ou plusieurs personnes.



**Figure 1.** Le modèle factoïde<sup>15</sup>

Les types de factoïdes correspondent à des traits, des états et des événements. Les traits ne changent pas au fil du temps, tandis que les états ne sont vrais que pour une période donnée. Les événements peuvent entraîner un changement d'état ou d'un trait. Par exemple, dans PASE [Pasin & Bradley, 2013], le statut et les informations personnelles sont des traits, alors que la possession et le bureau sont des états. Enfin, une transaction est considérée comme un événement. Dans le modèle factoïde, une source S atteste que le fait F peut être énoncé à propos de l'entité E. Cela peut être couplé avec un ensemble d'informations temporelles T. Il peut exprimer une relation entre l'entité (sujet) E et d'autres entités (objets) O1, O2, etc. Un factoïde est une assertion faite par une source, qui peut être niée par une autre. Ainsi, les factoïdes peuvent être contradictoires les uns avec les autres. Enfin, un certain degré de fiabilité peut être associé aux factoïdes. PASE est un exemple de modèle factoïde.

Un autre exemple de modélisation est le projet de la République Romaine [Figueira & Vieira, 2017]. Fondé aussi sur les principes du modèle factoïde, le modèle de données est organisé autour de quatre entités principales : Person, PostAssertion, StatusAssertion et RelationshipAssertion. Person stocke des informations sur chaque individu, telles que les noms, le sexe, les dates de vie, l'indicateur patricien, etc. L'entité PostAssertion stocke les données relatives aux postes détenus par des individus (tels que le préteur, le consul, le légat, etc.), tandis que StatusAssertion stocke les informations sur les classes sénatoriales. Le RelationshipAssertion représente les relations personnelles (frère de, fils de, etc.), reliant deux individus différents dans la base de données. Tous les types d'assertion sont liés à une entité SecondarySource avec des détails sur les références bibliographiques qui supportent les données (Figure 2).



**Figure 2.** Modèle de base de données République Romaine (adapté de [Figueira & Vieira, 2017])

<sup>15</sup> <https://factoid-digum.kcl.ac.uk/>

D'autres bases prosopographiques utilisent des modèles factoïdes, par exemple PoMS<sup>16</sup>, PBE<sup>17</sup> ou PONE<sup>18</sup>. Le modèle de la base de l'Europe de Charlemagne<sup>19</sup> s'appuie sur quatre types de factoïdes : 1) les attributs/relations qui représentent les caractéristiques d'un agent ou les liens entre deux agents ; 2) les factoïdes liés aux lieux qui enregistrent la décomposition hiérarchique médiévale des lieux ; 3) les factoïdes de transactions qui lient entre eux agents, possessions et lieux et 4) les autres factoïdes qui enregistrent les autres événements.

[Akoka *et al.*, 2019] présentent un modèle conceptuel qui rassemble et rend plus génériques les informations contenues dans différentes bases de données prosopographiques, à savoir les concepts de personnes, factoïdes, lieux et sources. Il intègre également une large représentation de l'incertitude. Dans ce modèle, la notion de factoïde est prise au sens large. Il comprend les factoïdes de certaines représentations prosopographiques, mais aussi tous les faits qui caractérisent les individus. Les dimensions de tous les concepts prosopographiques, y compris les factoïdes, sont associées à des référentiels hiérarchiques. Le mécanisme d'agrégation incorpore le temps comme dimension, permettant de faire varier la catégorisation dans le temps. Dans ce modèle, les noms des individus peuvent être connus de manière imprécise. Certaines relations entre les concepts sont typées. Par exemple, la nature de l'attribut entre le factoïde et l'objet indique clairement que, lors d'un événement de troc, un objet est attribué et un objet est accordé en échange. La représentation du temps intègre le temps discret (une date), le temps continu (un intervalle) et leur composition (plusieurs dates potentielles, ou plusieurs intervalles possibles, ou plusieurs intervalles cumulés, par exemple "il était présent de 1492 à 1500 puis de 1503 à 1508"). Enfin, il intègre la gestion des informations incertaines sous trois formes : un degré de certitude, de confiance et de précision. La certitude est une représentation du degré de fiabilité des informations auxquelles elle est attachée. La confiance est une caractéristique d'une source d'information qui agrège une opinion partagée par les historiens quant à sa crédibilité. La précision permet la représentation d'informations approximatives. Par exemple, la précision peut être relative à l'emplacement d'un événement. Les valeurs qu'elle peut prendre dans ce cas sont : près, autour, non loin de, à quelques kilomètres de, etc. Lorsqu'elle caractérise le moment où un événement a lieu, elle peut prendre les valeurs de : autour, avant, bien avant, peu de temps après, etc.

La modélisation des projets prosopographiques fait aussi appel à des ontologies. Nous présentons ci-dessous les trois ontologies prosopographiques les plus représentatives.

L'ontologie FPO (Factoïd Prosopography Ontology)<sup>20</sup> du projet PASE est une ontologie fondée sur OWL / RDFS (RDF Schema). Elle définit de manière formelle les concepts de l'approche factoïde telle qu'utilisée par le département des humanités numériques du King's College de Londres. FPO peut être utilisée comme base de tout projet prosopographique qui souhaite utiliser l'approche factoïde.

La seconde ontologie est celle du CIDOC-CRM [Doerr, 2003]. C'est une ontologie qui représente l'information relative au patrimoine culturel. Elle a été développée par le Comité International pour la DOCumentation (CIDOC) de l'International Council of Museum (ICOM). Issue du monde des musées, elle possède néanmoins une ambition multi-domaine, puisqu'elle intègre les données du

<sup>16</sup> <https://www.poms.ac.uk/>

<sup>17</sup> <http://www.pbe.kcl.ac.uk/>

<sup>18</sup> <http://www.pone.ac.uk/>

<sup>19</sup> <http://www.charlemagneseurope.ac.uk/>

<sup>20</sup> <https://factoid-digham.kcl.ac.uk/fpo-factoid-prosopography-ontology/>

monde des bibliothèques, des archives et des institutions de recherche. Elle est un standard aidant à intégrer l'information culturelle issue de différentes sources dans un seul environnement et est utilisée pour représenter les collections afin d'en améliorer le partage d'information. Elle a fait l'objet, en 2006, d'une normalisation auprès de l'Organisation internationale de normalisation (ISO) sous la référence ISO 21127 [Crofts, 2007]. En décembre 2014, est parue une nouvelle version (ISO 21127:2014)<sup>21</sup>. Le consortium Data for History (<http://dataforhistory.org/>) a entrepris des travaux relatifs à l'utilisation du CIDOC-CRM pour les données historiques. Mentionnons enfin que la notion du temps est aussi discutée au sein du SIG du CIDOC-CRM. Ainsi, l'effort de conceptualisation n'est pas l'apanage des bases de données mais il est aussi au cœur des projets de représentation des connaissances sous forme d'ontologies. Le modèle conceptuel de CIDOC-CRM s'appuie non pas sur des factoïdes mais sur des entités temporelles qui généralisent le concept d'événement.

Le projet SyMoGIH [Gedzelman et al., 2014 ; Beretta et al., 2016] a créé une plateforme modulaire ouverte pour stocker des informations géo-historiques. En offrant l'accès à une ontologie conçue en collaboration entre différents académiques de différents pays, et à une solution technologique robuste le projet SyMoGIH vise à l'évolution des pratiques dans le domaine de la production et de la conservation des données dans la recherche historique. La plateforme héberge 16 projets en cours, dont plusieurs à l'échelle européenne (France, Allemagne, Suisse, Belgique). Il y a environ 50 utilisateurs actifs et la plateforme contient près de 1.700.000 lignes de données. La nature générique du modèle de données SyMoGIH a permis de le transposer dans une ontologie OWL DL afin de publier les données. Elle est basée sur quatre classes principales: Object, KnowledgeUnit, Role et Sourcing. Bien que conçue de manière indépendante, l'ontologie de SyMoGIH présente également des similitudes étroites avec l'ontologie Simple Event Model [van Hage et al. 2011] et TemporalEntity dans CIDOC-CRM [Doerr, 2003] permettant une grande interopérabilité avec ces ontologies de référence.

## 6. Conclusion

La prosopographie est une branche de la recherche historique qui se concentre sur l'identification des personnes qui apparaissent dans les sources historiques. Il s'agit de représenter, de traiter et d'interpréter des données historiques, généralement issues de textes et de champs de base de données, de manière significative. La prosopographie s'appuie sur une base de données contenant des informations relatives à des personnes d'un milieu spécifique définies chronologiquement et géographiquement. Dans cet article, nous présentons un état de l'art des bases de données prosopographiques mettant l'accent sur les concepts spécifiques de la prosopographie et la recherche afférente à la modélisation de ces concepts. Nous décrivons la méthode prosopographique et présentons un panorama des bases de données prosopographiques en mettant l'accent mis sur les avantages et les limites. Nous présentons et discutons les principaux concepts de ces systèmes notamment les *personnes*, les *événements* auxquels les personnes sont liées, le *temps*, *l'incertitude* et les *sources*. Nous analysons les efforts de modélisation qui intègrent tout ou partie de ces concepts. Enfin, nous discutons l'apport des approches ontologiques utilisées par ces bases de données prosopographiques.

Cet article, bien qu'il cherche à intégrer des points de vue apparemment disparates tels qu'ils sont exprimés dans la communauté de la prosopographie, ne prétend pas à l'exhaustivité. Toutefois, il ouvre la voie d'une modélisation qui pourrait tirer bénéfice des efforts de représentation de chacun des concepts de base (nom, événement, temps, incertain, source).

<sup>21</sup> <https://www.iso.org/fr/standard/57832.html>

Cet article souffre de plusieurs limites à mentionner. Bien qu'il existe des bases de données portant sur la période contemporaine (<https://www.unil.ch/elitessuisses/index.php?page=accueil> et <https://www.odos.be/hercules/>), les bases de données prosopographiques citées sont, pour la plupart, liées à la période médiévale. La recherche bibliographique n'a pas fait l'objet d'une revue systématique de la littérature mais s'est appuyée sur des techniques de « boule de neige » (snowballing technique).

En termes de recherche future, il serait utile de proposer un modèle conceptuel générique plus complet intégrant la variété des recherches prosopographiques sur toutes les périodes de l'histoire, voire dans les autres domaines des sciences sociales ayant recours à la prosopographie. Enfin, une recherche utile consisterait à proposer des outils de migration des bases existantes vers une base prosopographique s'appuyant sur ce modèle conceptuel générique.

## References

- [Akoka *et al.*, 2019] Akoka, J., Comyn-Wattiau, I., Lamassé, S., & Du Mouza, C. Modeling historical social networks databases. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
- [Akoka *et al.*, 2020] Akoka, J., Comyn-Wattiau, I., & Du Mouza, C. Evaluer la crédibilité des sources historiques. In *Atelier EGC DAHLIA*, 2020.
- [Allen, 1983] Allen J.F. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843 (1983).
- [Allen & Ferguson, 1994] Allen, J. F., & Ferguson, G. (1994). Actions and events in interval temporal logic. *Journal of logic and computation*, 4(5), 531-579.
- [Bédard *et al.*, 2004] Bédard, Y., Larrivée, S., Proulx, M. J., & Nadeau, M. Modeling geospatial databases with plugins for visual languages: A pragmatic approach and the impacts of 16 years of research and experimentations on perceptory. In *International Conference on Conceptual Modeling* (pp. 17-30). Springer, Berlin, Heidelberg, 2004.
- [Bedford & Cooke, 2001] Bedford, T. and Cooke, R., “What is Uncertainty?,” *Probabilistic Risk Analysis: Foundations and Methods*, Cambridge University Press, Cambridge, United Kingdom, 2001, pp. 17-38.
- [Beretta *et al.*, 2016] Beretta, F., [Alamercery](#), V., [Ferhod](#), D.: The symogih.org Project: Towards an International Consortium. *Intl. Conf. On Digital Humanities (DH) 2016*: 743-744
- [Bernaudeau 2008] Bernaudeau, V. (Ed.). (2008). *Les praticiens du droit du Moyen Âge à l'époque contemporaine: Approches prosopographiques (Belgique, Canada, France, Italie, Prusse)*. PU Rennes. [Bradley & Short, 2002] Bradley, J., Short, H. (2002), Using formal structures to create complex relationships: the prosopography of the Byzantine Empire - a case study. *Resourcing Sources Prosopographica et Genealogica* 7.
- [Bry *et al.*, 2005] Bry F. and Marchiori M., Reasoning on the semantic web: beyond ontology languages and reasoners, 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT 2005), 2005, p. 317 – 321.
- [Bulst, 1991] Bulst N., « Objet et méthode de la prosopographie », in L'État moderne et les élites. 13e -18e siècles. Apports et limites de la méthode prosopographique. Actes du colloque international CNRS-Paris I, 16-19 octobre 1991, ed. J.-P. Genet et G. Lottes, Paris, pp. 467-482.
- [Charle, 2001] Charle, C. (2001). « Prosopography (collective biography) », in *International Encyclopedia of the Social and Behavioral Sciences*, Oxford,,Elsevier Science Ltd, 2001, vol.18.
- [Carrié 2003] Carrié, J. M. (2003). The contribution of papyri to the prosopography of the ancient world: Evaluation and prospects. In *Proceedings of the British Academy* (Vol. 118, pp. 73-93). Oxford University Press.
- [Costa *et al.*, 2012] Costa, P. C. G., Laskey, K.B., Blasch, E., & Jousselme, A-L. “Towards Unbiased Evaluation of Uncertainty Reasoning: The URREF Ontology,” *International Conf. on Info Fusion*, 2012.
- [Crofts *et al.*, 2003] Crofts, N., Doerr, M., Gill, T., Stead, S. and Stiff. M. Definition of the CIDOC, object-oriented conceptual reference model and crossreference manual, 2003. <http://cidoc.ics.forth.gr/docs/> cidoc crmv 3.4.9 cross referenced.zip.

- [Crofts, 2007] Crofts, N. (2007). La norme récente ISO 21127: une ontologie de référence pour l'échange d'informations de patrimoine culturel. *Systèmes d'informations et synergies entre musées, archives, bibliothèques universités, radios et télévisions, Lausanne–2007*.
- [Dedieu, 2004] Dedieu, J.-P., « Les grandes bases de données : une nouvelle approche de l'histoire sociale. Le système Fichoz », *Història*, III/5 (2004), p. 101-114.
- [Dedieu & Chaparro, 2013] Dedieu, J.-P., Chaparro Sainz Á., « De la base au réseau. L'apport des bases de données à l'étude des réseaux », *Cahiers de Framespa*, 12 (2013) [en ligne : <http://framespa.revues.org/2173>, consulté le 3 mars 2020].
- [Dequech, 2000] Dequech, D., 2000. Fundamental uncertainty and ambiguity. *Eastern Economic Journal* 26, 41–60.
- [De Ridder-Symoens, 1991] De Ridder-Symoens, H., « Prosopografie en middeleeuwse geschiedenis : een onmogelijke mogelijkheid? », in *Handelingen der Maatschappij voor Geschiedenis of Oudheidkunde te Gent*, Nieuwe reeks, deel XLV, Gent, 1991, pp. 95-117.
- [Doerr, 2003] Doerr, M. “The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata”. In *AI Magazine* Vol. 24 No. 3 (2003).
- [Figueira & Vieira, 2017] Figueira, L., Vieira, M., Modelling a Prosopography for the Roman Republic. The Digital Prosopography of the Roman Republic Project, Digital Humanities Conference, Montreal, <https://dh2017.adho.org/abstracts/091/091.pdf> (2017)
- [Fossier & Genet, 1983] Fossier, L. et Genet, J.-P., Dossier spécial « La prosopographie. Les “fantassins de l'histoire” à l'honneur », *Le médiéviste et l'ordinateur*, 10 (1983), [en ligne : <http://www.menestrel.fr/spip.php?rubrique667&lang=fr#2108>, consulté le 3 mars 2020].
- [François & Muller, 2015] François, A., Muller, F., La prosopographie à l'heure du numérique : bilan, prospectives et perspectives à l'aune de l'application « Belgian Magistrates » DOI: [10.25518/1370-2262.367](https://doi.org/10.25518/1370-2262.367)
- [Gangemi & Mika, 2003] Gangemi, A., & Mika, P. Understanding the semantic web through descriptions and situations. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 689-706). Springer, Berlin, Heidelberg (2003).
- [Gedzelman et al., 2014] *Gedzelman*, S. S., *Beretta*, F., *Ferhod*, D., *Boschetto*, S., *Butez*, C.-C., *Vernus*, P., *Hours*, B.: *The SyMoGIH project: Sharing and publishing historical and geographical data in a standard, open and interoperable way*. Int. Conf. On Digital Humanities ([DH 2014](https://dh2014.adho.org/abstracts/091/091.pdf))
- [Gregersen & Jensen, 1999] Gregersen, H., & Jensen, C.S. [Temporal Entity-Relationship models-a survey](https://doi.org/10.1109/19.770511). In: *IEEE Transactions on knowledge and data engineering*, 11 (3), 464-497 (1999).
- [Hengchen et al., 2015] Hengchen, S., Van Hooland, S., Verborgh, R., & De Wilde, M. (2015). L'extraction d'entités nommées: une opportunité pour le secteur culturel? *I2D Information, données documents*, 52(2), 70-79.
- [Hobbs & Pan, 2004] Hobbs, J. R., & Pan, F. (2004). An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing (TALIP)*, 3(1), 66-85.
- [Hyvönen, 2020] Hyvönen, E. Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web* 11(1): 187-193 (2020)
- [Keats-Rohan, 2007] Keats-Rohan, K.S.B. (ed.), *Prosopography Approaches and Applications A Handbook*, Oxford.
- [Lagoze & Hunter, 2001] Lagoze, C., Hunter, J. The ABC Ontology and Model. International Conference on Dublin Core and Metadata Applications, [S.1.], pp. 160-176, oct. 2001. ISSN 1939-1366. Available at <https://dcpapers.dublincore.org/pubs/article/view/655>. Date accessed: 03 Mar. 2020.
- [Lemercier & Picard, 2012] Lemercier, C., Picard, E. Quelle approche prosopographique ? Laurent Rollet; Philippe Nabonnaud. Les uns et les autres. *Biographies et prosopographies en histoire des sciences*, Presses Universitaires de Nancy; Editions Universitaires de Lorraine, pp.605-630, 2012, 9782814300965. fffalshs00521512v2f
- [Li et al., 2012] Li, Y., Chen, J., & Feng, L. (2012). Dealing with uncertainty: A survey of theories and practices. *IEEE Transactions on Knowledge and Data Engineering*, 25(11), 2463-2482.
- [Liu et al., 2010] Liu, W., Liu, Fu, Z. J., Hu, R. and Zhong, Z. "Extending OWL for Modeling Event-oriented Ontology," 2010 International Conference on Complex, Intelligent and Software Intensive Systems, Krakow, 2010, pp. 581-586.
- [Loriga, 2010] Loriga, S. (2010). *Le Petit x. De la biographie à l'histoire: De la biographie à l'histoire*. Le Seuil.

- [Manola *et al.*, 2004] Manola, F., Miller, E., McBride, B. : [RDF primer](#), W3C recommendation, uazuay.edu.ec (2004)
- [Mathisen, 1988] Mathisen, Ralph W., Medieval Prosopography and Computers: Theoretical and Methodological Considerations, Medieval Prosopography, 9 (1988), pp. 73-128.
- [Matousek *et al.*, 2007] Matousek, K., Falc, M., Kouba, Z. Extending Temporal Ontology with Uncertain Historical Time, Computing and Informatics, Vol. 26, 2007, 239–254.
- [Maurin, 1982] Maurin J., « La prosopographie romaine : pertes et profits », Annales. Économies, Sociétés, Civilisations, 1982, n° 37 (5-6), pp. 824-836.
- [McGuinness & Van Harmelen, 2004] McGuinness, D.L., Van Harmelen, F.: [OWL web ontology language overview](#), W3C recommendation, academia.edu (2004)
- [Moisuc *et al.*, 2005] Moisuc, B., Davoine, P. A., Gensel, J., & Martin, H. (2005). Design of spatio-temporal information systems for natural risk management with an object-based knowledge representation approach. *Geomatica*, 59(4), 401-413.
- [Parent *et al.*, 1998] Parent, C., Spaccapietra, S., Zimanyi, E., Donini, P., Plazanet, C., & Vangenot, C. (1998). Modeling spatial data in the MADS conceptual model. In *Proc. of the 8th Int. Symp. on Spatial Data Handling, SDH'98* (No. CONF).
- [Pasin & Bradley, 2013] Pasin, M., Bradley, J. (2013) Factoid-based prosopography and computer ontologies: towards an integrated approach, *Digital Scholarship in the Humanities* 30.1 (2013): 86-97.
- [Prost, 1996] Prost A. (1996) Douze leçons sur l'histoire, Paris, le Seuil, 1996.
- [Raimond & Abdallah, 2007] Raimond, Y., Abdallah, S. 25th October 2007, “The Event Ontology”, <http://motools.sf.net/event/event.html> (rdf).
- [Shaw & Larson, 2008] Shaw, R. and Larson, R.R. Event Representation in Temporal and Geographic Context International Conference on Theory and Practice of Digital Libraries. Springer, Berlin, Heidelberg, 2008.
- [Stone, 1971] Stone, L.: Prosopography, *Daedalus*, 100, 1, (1971), pp. 46-79.
- [Smithson, 1989] Smithson, M., 1989. Ignorance and Uncertainty: Emerging Paradigms. Springer Press, New York.
- [Smithson, 2009] Smithson, M., 2009. Psychology's ambivalent view of uncertainty. In: Bammer, G., Smithson, M. (Eds.), *Uncertainty and Risk: Multidisciplinary Perspectives*. Earthscan, Sterling, VA.
- [Snodgrass, 2012] Snodgrass, R.T. The TSQL2 Temporal Query Language, Computers 2012 - <https://books.google.fr/books?isbn=1461522897>.
- [Strong *et al.*, 1997] Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110.
- [Taleb, 2010] Taleb, N.N., 2010. *The Black Swan: The Impact of the Highly Improbable*. Random House, New York.
- [Tannert *et al.*, 2007] Tannert, C., Elvers, H., Jandrig, B., 2007. The ethics of uncertainty. In the light of possible dangers, research becomes a moral duty. *EMBO Reports* 8 (10), 892–896.
- [Thunnissen, 2003] Thunnissen, D., 2003. Uncertainty classification for the design and development of complex systems. Proceedings of the 3rd Annual Predictive Methods Conference, Veros Software, Santa Ana, CA.
- [Tran, 2006] Tran, M. Prolexbase. Un dictionnaire relationnel multilingue de noms propres : conception, implémentation et mise en ligne. *Traitement du texte et du document*. Université de Tours, 2006. Français. fftel-01726999f.
- [UKY, 2019] University of Kentucky, Evaluating Historical Sources, [www.uky.edu/~dolph/HIS316/handouts/sources.html](http://www.uky.edu/~dolph/HIS316/handouts/sources.html)
- [van Asselt & Rotmans, 2002] Van Asselt, M., Rotmans, J., 2002. Uncertainty in integrated assessment modeling: from positivism to pluralism. *Climatic Change* 54, 75–105.
- [van Hage *et al.*, 2009] Van Hage, W. R., Malaisé, V., Segers, R., Hollink, L., & Schreiber, G. (2011). Design and use of the Simple Event Model (SEM). *Journal of Web Semantics*, 9(2), 128-136.
- [Verboven *et al.*, 2007] Verboven, K., Carlier, M., & Dumolyn, J., « A Short Manual to the Art of Prosopography », in Katharine S. B. Keats-Rohan (ed.), *Prosopography Approaches and Applications A Handbook*, Oxford : Occasional Publications UPR, 2007, disponible en ligne (<http://prosopography.modhist.ox.ac.uk/images/01%20Verboven%20pdf.pdf>)

- [West, 2007] Vest, K., *Using Primary Sources in the Classroom*. Upper Saddle River: Shell Education 2007.
- [Wakeham , 2015] Wakeham, J. *Uncertainty: History of the Concept*, International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, Volume 24, 2015 Elsevier.
- [Westermann & Jain, 2007] Westermann, U., and Jain, R. Toward a common event model for multimedia applications. *IEEE Multimedia*, 14(1):19–29, 2007.
- [Zaveri *et al.*, 2016] Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2016). Quality assessment for linked data: A survey. *Semantic Web*, 7(1), 63-93.
- [Zio & Pedroni, 2013] Zio, E., & Pedroni, N. (2013). *Literature review of methods for representing uncertainty*. FonCSI.