

Graphes de connaissances pour représenter et analyser l'évolution des territoires en Histoire

Knowledge graphs to represent and to analyse the evolution of territories in History

Lucas Bourel¹, William Charles¹, Nathalie Hernandez^{1,2}, Nathalie Aussenac-Gilles², Victor Gay¹ and Sébastien Poublanc³

¹IRIT- CNRS et Université de Toulouse <http://www.irit.fr>

²Université Toulouse2 Jean Jaurès

³Toulouse School of Economics (TSE) Institute for Advanced Study in Toulouse (IAST) University of Toulouse Capitole

Abstract

The notion of territory amounts for an important part of human and social sciences. As a spatio-temporal object considered from the digital humanities viewpoint, the question of its digital representation brings forward the need to represent its various aspects. Territory is considered here as a geographical area defined by stakeholders exerting a defined kind of power (religious, legal, ...) over said area, or at least trying to exert it. Inspired by the territorial ontologies TSN and TSN-Change, we propose the HHT ontology (Hierarchical Historical Territory) which is designed to represent the state of affairs of various hierarchical territorial divisions identified by historians' research. Such territory representation implies an overcoming of the main difficulty of such work, which is the lack of exhaustive historical sources regarding the whole territory. Additionally, the HHT ontology endeavors to represent the former states of historical knowledge of territories, now considered deprecated. Defined as part of the ANR ObARDI project, which aims to study territories under the historical period known as Ancien Régime (1661-1789), HHT fills the specific need of representing hierarchical historical territorial divisions.

Keywords

ontology, Semantic Web, digital humanities, spatio-temporal objects

1. Introduction

Dans le domaine des sciences humaines et sociales, la notion de territoire constitue un objet d'étude central à cheval entre l'histoire, la géographie et l'économétrie. Pour ces disciplines, il se caractérise par les éléments suivants :

- un pouvoir exercé par un acteur sur le territoire (**la domination**) ;
- l'espace dominé par ce contrôle territorial (**l'aire**) ;
- la connaissance des **limites** qui enserrent le territoire.

En conséquence, un territoire peut être un découpage administratif, un espace étatique, ou tout espace revendiqué par ses habitants¹. Il s'agit toujours de l'appropriation d'un espace par un acteur.

Mais un territoire n'est pas réductible à sa seule délimitation spatiale; la notion intègre également une dimension temporelle car le territoire évolue au cours du temps. Les paroisses de

. *Workshop on Digital Humanities and Semantic Web*

. ✉ lucas.bourel@irit.fr (L. Bourel); william.charles@irit.fr (W. Charles); nathalie.hernandez@irit.fr (N. Hernandez); aussenac@irit.fr (N. Aussenac-Gilles); victor.gay@tse-fr.eu (V. Gay)

1. cf <http://geoconfluences.ens-lyon.fr/glossaire/territoire>

Paris en 1789 sont très différentes des communes - leurs équivalents actuels - tant par leur surface, leurs caractéristiques d'urbanisation, ou leur place dans les nomenclatures administratives. Pourtant, on peut établir un lien de filiation directe entre les paroisses et les communes. Par conséquent, il s'agit de parler d'un objet spatio-temporel dont l'évolution est déterminée par des activités humaines. Enfin, les territoires sont imbriqués dans des rapports de force les uns par rapport aux autres, s'inscrivant dans des hiérarchies territoriales complexes, faisant l'objet de revendications ou de conflits.

Dans une perspective d'humanités numérique, la notion de territoire implique de prendre en compte chacune de ces dimensions afin de pouvoir correctement visualiser et analyser les évolutions territoriales étudiées.

Dans le domaine du web sémantique, les ontologies ont montré leur intérêt pour définir des vocabulaires partagés servant à décrire des entités d'un domaine en vue de les manipuler. Ces descriptions forment des graphes de connaissances. Représenter la notion de territoire à partir d'ontologies présente plusieurs avantages. Tout d'abord, les représentations des territoires réalisées à partir d'ontologies peuvent être facilement interrogées et visualisées. Décrites à partir de vocabulaires partagés, ces représentations de territoires peuvent alors être publiées sur le Web de données liées, partagées et exploitées par les chercheurs de différentes disciplines. Les graphes de connaissances permettent également de représenter plusieurs points de vue d'intérêt différents sur ces territoires, mais aussi de compléter la représentation intrinsèque d'un territoire à l'aide de données connexes (données démographiques, sociales, institutionnelles recensement de la population, présence de rébellions, etc.). Cette démarche s'inscrit dans celle des science ouvertes et collaboratives.

Plusieurs travaux proposent déjà une modélisation des territoires à partir de graphes de connaissances [1, 2, 3, 4]. Dans la plupart de ces approches, on suppose qu'une institution normative (comme l'Union européenne ou l'État français) détermine les territoires en jeu à partir d'une date donnée. Ces territoires sont décrits dans des sources primaires. Par exemple, l'INSEE fournit, sous forme de données ouvertes, la liste des différents types de territoires administratifs français, leur code et leur géographie pour une date donnée^{2 3}. Les territoires sont organisés hiérarchiquement dans la perspective administrative établie par l'institution normative.

Cette approche moderne des modélisations hiérarchiques des territoires ne convient toutefois pas à la représentation d'une organisation hiérarchique territoriale du point de vue historique. En effet, sous l'Ancien Régime par exemple, il n'existe pas de source unique décrivant l'ensemble des territoires et leur organisation hiérarchique pour une date précise et à l'échelle de tout le royaume de France. Les archives utilisées pour reproduire les nomenclatures en usage à cette époque sont forcément partielles, multiples et souvent contradictoires. C'est aux historiens de les interpréter pour arriver à un état des connaissances que l'ontologie se devra de restituer.

A la croisée de l'histoire, de la géographie, de l'économie et des sciences informatiques, l'un des enjeux du projet ANR transdisciplinaire ObARDI est de réussir à modéliser les unités territoriales et leurs hiérarchies entre 1661, début du règne personnel de Louis XIV qui correspond à la mise place de la monarchie "administrative", et 1789, fin de l'Ancien Régime.

2. <https://www.insee.fr/fr/information/2016807>

3. <http://rdf.insee.fr/>

Dans le cadre d'ObARDI, l'objectif pour les historiens est d'améliorer la compréhension des dynamiques de pouvoir qui sous-tendent la construction de l'État moderne en France. En étudiant ses mécanismes de développement et sa manière de représenter un territoire politique, il s'agit de dépasser le méta-récit de la construction de l'État qui en freine encore sa compréhension⁴ [5] [6]. Pour cela, la représentation à travers une ontologie de ce concept-clé qu'est le territoire est centrale non seulement dans la démarche des historiens mais également dans l'analyse des données territoriales de l'Ancien Régime.

Dans cet article, nous proposons l'ontologie HHT (*Historical Hierarchical Territory*) pour représenter les caractéristiques intrinsèques d'un territoire historique. Cette ontologie permet de représenter plusieurs découpages hiérarchiques simultanés du territoire et l'évolution de ces découpages territoriaux au cours du temps. Elle s'inspire fortement des ontologies TSN et TSN-Change [1] consacrées à la représentation de nomenclatures territoriales organisées de manière hiérarchique, et de leurs changements dans le temps. Nous proposons également une approche itérative pour construire un graphe de connaissances à l'aide de cette ontologie à partir de descriptions de différentes unités territoriales, de leurs relations hiérarchiques et de connaissances sur leur évolution selon plusieurs points de vue (administratif, juridique ou religieux par exemple) durant une période historique.

La suite de l'article s'organise en 3 parties. La section 2 présentera la définition choisie des découpages en unités territoriales à travers le temps. La section 3 exposera les fondements de l'ontologie HHT construite à partir de cette réflexion. Enfin, la section 4 décrira la méthode de construction d'un graphe de connaissances à l'aide de cette ontologie.

2. Unité territoriale et hiérarchie

Cette section détaille la définition des unités territoriales et leurs multiples hiérarchies du point de vue de la recherche en histoire.

2.1. Identité d'une unité territoriale à travers le temps

Les historiens attribuent différentes caractéristiques aux territoires, en particulier des liens de filiation, disruptive ou non, que l'on peut définir à travers le temps entre des territoires.

Dans le cadre d'ObARDI, les historiens appréhendent le territoire comme un rapport de force d'un acteur sur un espace géographique. Comme ce rapport de force est amené à évoluer dans le temps, la question se pose alors de l'identité du territoire à travers le temps [7]. Représenter une entité spatio-temporelle inter-connectée à d'autres de ses semblables se modifiant au cours du temps pose de nombreuses questions de modélisation. Plusieurs approches fondamentales existent pour y répondre. Nous reprenons la dualité entre perdurance (propriété des entités qui ne changent pas dans le temps) et endurance (propriété des entités qui ont une durée déterminée, y compris instantanée) proposée par N. Guarino pour structurer les ontologies formelles [8]. La manière d'organiser une ontologie selon ces notions conduit Grenon et Welty à définir deux types d'ontologies [9]. Pour rendre compte de la perdurance, les ontologies SPAN offrent une vision 4D des objets qui sont des "vers d'espace-temps" perdurant dans le temps. Construite pour

4. <https://obardi.hypotheses.org/270/>

représenter des entités endurantes, les ontologies SNAP donnent une vision tridimensionnelle d'un objet qui dure dans le temps. Il va falloir choisir un camp pour notre objet territorial.

En reprenant les choix retenus dans les travaux sur TSN & TSN-Change[1], l'approche 4D-Fluent (Une approche perdurantiste qui parle de *TimeSlice* pour représenter les versions du perdurant) semble la plus appropriée et correspond à notre définition du territoire établie avec les historiens du projet. Le territoire est vu alors comme un *processus comprenant l'ensemble des tranches de son existence*. La vie du territoire existe à travers toutes les versions de lui-même. Mais chacune de ces versions possède des attributs et des caractéristiques qui lui sont propres. Les historiens s'intéressant particulièrement à la question de l'évolution, le type de changement qui se produit entre deux versions d'un territoire doit être également modélisé dans notre ontologie⁵. Cependant, cette approche soulève deux questions :

- Quelle est le critère d'identité diachronique qui permet d'établir l'identité du territoire ? [7]
- Quelles sont les caractéristiques essentielles d'une unité territoriale qui définissent la singularité de chacune des versions du territoire ?

Nous avons donc cherché à définir les caractéristiques intrinsèques d'un territoire. Les données démographiques, sociales ou les caractéristiques d'urbanisation d'un territoire sont des observations statistiques faites sur le territoire mais ne le caractérisant pas en lui-même. Un échange approfondi avec des historiens ainsi que l'étude des données usuelles associées aux territoires en histoire ont permis d'arriver à d'autres propriétés caractérisant le territoire lui-même :

- Un nom
- Une géométrie (représentant sa délimitation spatiale)
- Son type ou sa catégorie hiérarchique (indiquant son rôle dans une hiérarchie donnée, tel que l'intendance, un diocèse, une élection, etc.)
- Ses relations hiérarchiques.

Le dernier point est particulièrement intéressant car il définit le territoire par rapport à d'autres territoires. Un territoire peut se placer sous la juridiction d'un autre (l'unité territoriale supérieure) et plusieurs autres territoires peuvent tomber sous sa propre juridiction (unités territoriales inférieures). Ces relations hiérarchiques reflètent encore une fois des rapports de force entre acteurs. Elles sont cruciales car elles vont former le squelette hiérarchique des institutions de l'Ancien Régime, un des principaux objets d'étude dans le cadre d'ObARDI.

Aucune de ces quatre caractéristiques n'est cependant essentielle pour l'identité d'un territoire [10]. Mais, chacune de ces caractéristiques peut engendrer un changement accidentel [7] qui fait passer d'une version du territoire à l'autre. Chacune de ses modifications peut se modéliser sous la forme d'un événement *ChangeBrige* [11] Cela répond à notre deuxième question sur les changements non-disruptifs [7], c'est-à-dire ceux qui ne modifient pas l'identité du perdurant représentant la vie d'un territoire. Reste à déterminer le critère permettant de qualifier un changement disruptif. Toutefois, ce critère ne doit rester qu'un guide. En effet, dans une démarche de recherche historique, seul l'historien peut, grâce à l'analyse des sources, trancher sur la véracité ou non d'une connaissance historique. Néanmoins, il est utile d'explicitier le critère diachronique d'identité que l'on veut justement expérimenter à l'épreuve de l'analyse des

5. L'approche est identique encore une fois à celle de TSN&TSN-Change.

historiens. En reprenant celui développé par Garbaza [7], on énonce le critère d'identité local suivant :

Changement disruptif il y a si et seulement si le nom s'en retrouve modifié en même temps qu'une autre caractéristique du territoire (catégorie, géométrie, relation hiérarchique). Tout autre changement dans lequel serait impliqué un territoire sera non disruptif.

2.2. Relations hiérarchiques

Une hiérarchie territoriale est considérée ici comme une classification des territoires. Elle repose sur un critère hiérarchique identifié par l'historien qui va servir de caractère discriminant permettant d'établir cette classification. Chacune de ces classes possède un niveau hiérarchique, et peut être rattachée à un ou plusieurs niveaux supérieurs et inférieurs. La plupart des découpages administratifs modernes se construisent à l'aide de tel système hiérarchique de découpage du territoire qui sert de nomenclature.

Ces hiérarchies sont le reflet des relations hiérarchiques des territoires qui les constituent. Autrement dit, si une catégorie en domine une autre, une unité de la première catégorie pourra alors dominer une unité de cette deuxième catégorie.

Une différence est à noter entre territoire et unité territoriale. Cette dernière est une catégorie plus générale de la première, définie par les mêmes caractéristiques, mais ne représentant pas historiquement l'appropriation par un acteur de l'espace. Cette distinction est utile lorsque des catégories étiques⁶ sont à modéliser.

Par exemple, dans le cadre de la norme européenne NUTS, des seuils démographiques sont fixés par niveau hiérarchique. Pour respecter des réalités socio-économiques, la nomenclature NUTS tente également de respecter les divisions administratives locales déjà mises en place⁷. Cette hiérarchie représente donc à la fois des unités territoriales simples et des territoires.

Le système hiérarchique obtenu découpe alors le territoire européen en plusieurs niveaux :

- Pays membre ;
- NUTS 1 : grandes régions socio-économique. (entre 3 millions et 7 millions d'habitants) ;
- NUTS 2 : régions de bases pour l'application de politiques régionales (entre 800 000 et 3 millions d'habitants) ;
- NUTS 3 : petites régions (entre 150 000 et 800 000 habitants).

L'organisation des unités territoriales sous l'Ancien Régime est plus complexe que les normes européennes actuelles. Après discussion avec les historiens, pour étudier les territoires de l'Ancien Régime, quatre dimensions de classification ont émergé : administrative, religieuse, judiciaire, fiscale. Ces quatre dimensions permettent d'établir quatre découpages hiérarchiques territoriaux différents du royaume de France, quatre filtres d'étude possibles qui se superposent les uns aux autres. Ils ne sont pas étanches entre eux puisque les unités territoriales peuvent posséder chacune une ou plusieurs dimensions.

Chacune de ces dimensions possède ses propres niveaux hiérarchiques. Dans ObARDI, le *découpage administratif* est constitué de pays d'États, de généralités et d'intendances, d'élections et de subdélégations, mais aussi de diocèses civils. Le *découpage judiciaire* est constitué des justices royales que sont les parlements, présidiaux, bailliages et sénéchaussées. Le *découpage*

6. correspondant à un filtre du point de vue de l'observateur et non pas de la réalité historique de l'époque

7. <https://ec.europa.eu/eurostat/fr/web/nuts/principles-and-characteristics>

religieux contient des archevêchés, des évêché, des archidiaconés, des doyennés et des paroisses ecclésiastiques. Enfin, le *découpage fiscal* contient les greniers à sels des gabelles, les assiettes des tailles, ou encore les départements des tabacs.

2.3. Un état de fait évolutif

Les découpages territoriaux de l’Ancien Régime relèvent d’un état de fait qui atteste de rapports de force entre différents acteurs. Ils ont laissé comme trace de leurs conflits juridiques ces hiérarchies territoriales qui structurent l’espace du royaume de France. Analyser les changements que subissent ces unités permet de comprendre la mise en concurrence des pouvoirs sous l’Ancien Régime, les oppositions développées entre acteurs, agents de la puissance publique et les institutions.

Cependant, aucune source ne décrit de manière exhaustive l’état de l’ensemble des territoires étudiés pour une date donnée. Aucune source ne décrit non plus à elle seule l’ensemble des changements subis par ces territoires.

Parallèlement, les historiens souhaitent également documenter l’évolution des connaissances qui a mené à la représentation d’un état de fait à un moment donné. À la lecture de nouvelles sources ou à la réalisation d’enquêtes nouvelles, les représentations changent et se modifient. Il convient donc d’en garder trace dans une perspective méthodologique et épistémologique.

2.4. Etat de l’art

Cette section décrit les ontologies utilisées dans le domaine du web sémantique en lien avec la notion de territoire.

2.4.1. Temps, espace et unités territoriales

L’ontologie OWL-time propose un vocabulaire standardisé par le W3C pour représenter le temps [12]. Elle permet de représenter des instants ou des intervalles temporels dans divers calendriers et d’exprimer des relations topologiques entre eux.

Concernant l’espace, GeoSPARQL propose non seulement un vocabulaire mais aussi un mécanisme de raisonnement spatial [13]. Ce vocabulaire permet de définir à l’aide de coordonnées et d’une forme (un *shapefile*) une zone définie dans l’espace. Néanmoins, dans notre cas, plus qu’une simple zone géographique, nous cherchons à définir un territoire.

Des ontologies d’applications sont spécifiquement dédiées à des découpages territoriaux administratifs. On peut citer ainsi l’INSEE qui propose une ontologie capturant le découpage administratif français actuel pour décrire un jeu de données sur le web des données liées⁸.

Néanmoins, ce que nous cherchons est une ontologie de domaine permettant de décrire des découpages territoriaux hiérarchiques génériques. TSN et TSN-Change sont justement deux ontologies construites à cette fin [1]. TSN décrit un découpage pour l’ensemble d’un territoire à une période donnée tandis que TSN-change décrit les changements permettant de passer d’une version de ce découpage territorial à une autre.

8. <http://rdf.insee.fr/>

Dans TSN, les notions d'unité territoriale, de niveau hiérarchique, et de nomenclature représentant une hiérarchie stable des territoires sont définies.

Contrairement aux agences statistiques auxquelles sont principalement dédiées TSN et TSN-Change, des découpages stables de la France, même pour une courte durée, n'existent pas durant l'Ancien Régime. Chacune des unités territoriales de l'Ancien Régime est documentée par des sources différentes qui ont leur propre temporalité.

2.4.2. Trajectoire de vie

Nous avons évoqué dans la section précédente, la question d'identité d'un territoire à travers ses différentes versions. La *trajectoire de vie* d'une unité territoriale correspond à l'ensemble des différentes versions de cette unité au cours du temps [1]. Dans TSN et TSN-Change, la trajectoire de vie des territoires est rythmée par des versions régulières de la nomenclature toute entière. Dans notre cas, chaque trajectoire de vie peut posséder ses propres références, ses propres sources. C'est sur cette notion différente de la temporalité de ces trajectoires de vie que nous allons devoir diverger de TSN et de TSN-Change.

CIDOC-CRM est une des ontologies de référence pour le patrimoine [14]. Elle a été créée en premier lieu pour répondre à la question d'archivage d'objets culturels, mais elle peut s'étendre hors de ce cadre. Elle est organisée autour de la notion d'évènement. Dans CIDOC-CRM, la notion de zone géographique est définie ainsi que des relations d'inclusion spatiale. Mais, ni les niveaux hiérarchiques ni la notion de territoire ne sont considérés. Cependant les changements territoriaux que nous cherchons à décrire (fusion de deux territoires, apparition d'un nouveau territoire, etc...) peuvent être perçus comme des événements historiques du point de vue de CIDOC-CRM. Certaines notions que nous cherchons à représenter s'alignent donc avec des concepts de CIDOC-CRM.

Pour représenter les hiérarchies territoriales historiques, nous allons donc devoir créer notre propre ontologie en réutilisant des concepts ou principes de ces vocabulaires.

2.5. Spécification de l'ontologie

Nous présentons ici la démarche suivie pour représenter au sein d'une ontologie les caractéristiques et les spécificités des hiérarchies territoriales historiques que nous venons d'énoncer.

La première étape dans une telle démarche, selon la méthodologie NEON [15] est d'identifier les exigences de l'ontologie. En premier lieu, une discussion permanente avec les experts du domaine, ici les historiens, a permis de cerner plus finement la problématique de cette modélisation par rapport à d'autres. L'objet d'étude et ses caractéristiques ont pu être détaillés dans les sections précédentes suite à ces échanges. Il a fallu également préciser les spécificités des sources disponibles pour les données historiques prévues dans le cadre d'ObARDI [16].

Le *but* de l'ontologie est donc de représenter des hiérarchies territoriales multiples dans le cadre de la recherche historique afin de pouvoir analyser leur évolution spatio-temporelle dans le cas de l'Ancien Régime.

Mais la *portée* de cette ontologie dépasse le seul cas d'étude d'ObARDI pour traiter la question de la représentation d'un territoire en histoire. Le développement d'un portail sémantique

est également prévu afin d'interagir avec cette ontologie, les données du projet et de pouvoir proposer des règles de raisonnements supplémentaires pour ces données.

En terme d'*implémentation*, l'ontologie sera implémentée avec la technologie RDF et le langage OWL.

Les *utilisateurs finaux* de cette ontologie se situeront à la croisée des différentes disciplines impliquées dans le projet ANR ObARDI. Des historiens, des économistes ou des géographes pourront utiliser ce portail afin de manipuler l'ontologie et les connaissances ainsi modélisées.

Les discussions avec les membres du projet ObARDI ont permis d'établir des cas d'usages de cette ontologie ainsi que des *Competency Question (CQ)*.

Néanmoins, l'ontologie HHT proposée dans cet article, n'est pas encore à complète maturité. En suivant les méthodes agiles tel que SAMOD [17], nous travaillons par cycles itératifs pour affiner cette ontologie. Voici quelques exemples de CQ relevées pour le moment :

- Comment évoluent les institutions en concurrence au cours du temps ?
- Quelles sont les superpositions de juridictions à un moment donné et à un endroit donné ?
- Quelles sont toutes les paroisses en conflit à une date donnée ?
- Quelle est l'évolution de l'institution X dans la durée ?

L'ontologie HHT nécessite encore plusieurs itérations supplémentaires afin de pouvoir répondre à tous ces cas d'usage. Néanmoins, dans l'état actuel des choses, elle répond à un sous-ensemble des questions soulevées : celles portant sur l'état de fait et sur son évolution.

Dans le suite de la méthodologie NEON, l'ontologie doit s'inscrire dans le réseau grandissant des ontologies déjà existantes à travers différents scénarios. Le scénario 1⁹ a été suivi pour partir des spécifications jusqu'à l'implémentation de l'ontologie HHT. En cherchant des ressources ontologiques supplémentaires, des ontologies décrivant l'espace (GeoSPARQL) et le temps (OWL-Time) ont été intégrées au projet (scénario 3¹⁰). HHT s'est concentré alors seulement sur l'aspect territorial et hiérarchique. TSN&TSN-Change a semblé la meilleure piste pour l'aspect hiérarchique, mais des différences fondamentales existant avec le cadre de la recherche historique nous ont poussé à implémenter sous une nouvelle forme une partie de cette ontologie, avec des bases plus générales (Scénario 4¹¹). Concernant les événements qui retraduisent les changements territoriaux, ces concepts de l'ontologie HHT peuvent étendre la notion d'évènement déjà présent dans l'ontologie de référence en histoire CIDOC-CRM (Scénario 8¹²).

L'ontologie HHT ainsi produite s'inscrit donc dans le réseau des ontologies déjà existantes sur ce domaine.

Les objectifs auxquels HHT tente de répondre sont alors de modéliser les éléments suivants :

- les unités territoriales ;
- leurs niveau hiérarchiques ;
- différents critères de classification hiérarchique ;
- les changements subis par ces unités et ces niveaux ;
- l'évolution des connaissances des historiens.

9. "From specification to implementation"

10. "Reusing ontological resources"

11. "Reusing and re-engineering ontological resource"

12. "Restructuring ontological resources"

3. Ontologie HHT

Dans cette section, nous détaillons l'ontologie créée pour le projet ObARDI et destinée à représenter des unités territoriales (UT) organisées hiérarchiquement en histoire. Nommée HHT pour *Hierarchical Historical Territory*, elle s'inspire très fortement des ontologies TSN et TSN-Change mais s'en éloigne sur des points fondamentaux.

Cette ontologie repose sur trois modules :

- **Module 1** : La modélisation des territoires et de leurs hiérarchies.
- **Module 2** : Les changements et les modifications appliquées à ces territoires.
- **Module 3** : Les relations de revendication s'opposant à l'état de fait.

Dans la suite de l'article, nous choisirons *hht* comme préfixe de cette ontologie.

3.1. Module 1 : Unités territoriales et hiérarchies

Cette partie de l'ontologie cherche à représenter n'importe quelles hiérarchies territoriales historiques. Elle repose sur différents concepts :

hht:Area Simple zone géographique (héritant de `geosparql:Feature`¹³ provenant de l'ontologie GeoSPARQL [13]. Ce concept n'est défini que par sa dimension spatiale.

hht:Unit sous-classe de `hht:Area`, représentant une zone géographique appartenant à une hiérarchie. Ce concept est défini par sa dimension spatiale et par sa dimension hiérarchique.

Les paroisses : Notons que dans le cadre du projet ObARDI nous ne disposons pas de la géométrie des unités territoriales. Et bien souvent d'ailleurs, en histoire, il est difficile de reconstruire une représentation spatiale précise des lieux décrits dans les sources.

Pour combler ce manque, nous avons identifié un niveau de référence à notre découpage. La géométrie des unités de ce niveau de référence sera considérée comme fixe sur toute la période d'étude. Pour le cas d'ObARDI, ce sont les paroisses qui ont rempli ce rôle¹⁴. Niveau hiérarchique le plus bas suivant nos quatre critères hiérarchiques, les frontières des paroisses ont très peu changé durant l'Ancien Régime. En considérant la géométrie des unités du niveau hiérarchique le plus bas possible comme fixe, on peut alors reconstruire une notion d'espace discret pour tous ceux qui lui sont supérieurs.

hht:historicalTerritory sous-classe d'`hht:Unit`, représentant une portion de l'espace géographique réclamé ou occupé par une personne, un groupe de personnes ou une institution qui en définit elle-même les frontières.

hht:Level Niveau hiérarchique permettant de classer des `hht:Unit`

13. <http://www.opengis.net/ont/geosparql#{#}Feature>

14. <https://obardi.hypotheses.org/526>

hht:HierarchicalCriterion Critère hiérarchique, caractère discriminant qui permet de définir un découpage hiérarchique de l'espace en différents `hht:Level`. Dans ObARDi, on dispose de quatre critères : administratif, religieux, judiciaire, fiscal.

Chacun de ces concepts permet de décrire l'espace. Pour prendre en compte leur évolution dans le temps, nous définissons 3 nouveaux concepts, versions des concepts précédents, possédant chacun leur propre période de validité (à travers la propriété `hht:validityPeriod`).

hht:UnitVersion Version d'une `hht:Unit` sur une période de validité donnée. Elle possède des unités supérieures auxquelles elle est liée par la propriété `hht:hasSuperUnit` et des unités inférieures auxquelles elle est liée par la propriété `hht:hasSubUnit`.

hht:HistoricalTerritoryVersion Version d'un `hht:HistoricalTerritory` sur une période de validité donnée.

hht:LevelVersion Version d'un niveau hiérarchique sur une période de validité donnée. Un niveau possède un niveau supérieur par la propriété `hht:hasSuperLevel` et un niveau inférieur par la propriété `hht:hasSubLevel`. Un niveau hiérarchique possède des `hht:UnitVersion` à travers la propriété `hht:hasMember`.

La figure 1 représente les concepts ainsi que les relations définies dans le module. Ce schéma est très fortement inspiré de la structure de TSN. Mais, la principale différence réside dans les trajectoires de vie de chaque unité. Dans notre cas, les différentes versions d'une même unité ne dépendent que d'elle-même.

3.2. Module 2 : Changements

Le module 2 vise à représenter les changements ayant mené à la création d'une nouvelle version d'unité territoriale. Trois grands types de changements sont considérés, dont les deux premiers sont inspirés de TSN-Change :

- `hht:FeatureChange` : représentant la modification d'une simple caractéristique entre deux versions d'une unité territoriale : `hht:Expansion`, `hht:Contraction`, `hht:Deformation`, `hht:Disappearance`, `hht:Appereance`, `hht:NameChange`, `hht:UpperUnitChange`, `hht:SubUnitChange`,...
- `hht:CompositeChange` : changement composite, c'est-à-dire un évènement regroupant plusieurs changements simples (`hht:FeatureChange`) : `hht:Merge`, `hht:Split`, `hht:Redistribution`.

Par exemple, une fusion consiste en la disparition de deux territoires et l'apparition d'un nouveau territoire sur l'espace des deux derniers.

- `hht:UpdateKnowledge` : Cet évènement est différent des deux autres dans le sens où il ne décrit pas un évènement historique mais la mise à jour du graphe de connaissances. Il permet de retracer ainsi l'ensemble des versions par lesquelles est passée l'information sur une unité territoriale pour arriver à l'état actuel des connaissances. Seule la dernière



FIGURE 1 : Schéma de l'ontologie HHT

est considérée valide tandis que les autres sont considérées comme des connaissances obsolètes.

Chacun de ces concepts est une sous-classe du concept E5 Évènement¹⁵ de CIDOC-CRM .

3.3. Module 3 : Conflits et Revendications

Une troisième partie de cette ontologie concerne la représentation des conflits territoriaux. Analyser l'évolution de ces conflits et les visualiser sont des cas d'usage qui sont revenus souvent de nos discussions avec les historiens du projet.

Dans un premier temps, pour parler de conflits, il faut parler d'acteurs. Comme énoncé précédemment, un territoire est vu comme l'appropriation d'un espace par un acteur. On définit `hht:Institution` comme un sous-type du Concept E39 d'Acteurs¹⁶. Une institution est l'acteur qui dirige un `hht:historicalTerritory` à travers la relation `hht:attachedTo`.

Chacun des territoires présents dans une hiérarchie peut donc se doter de l'acteur qui représente son institution. Les liens de hiérarchies entre territoires sont alors compris comme des rapports de force entre les acteurs associés. Un territoire placé hiérarchiquement sous un autre tombe sous sa juridiction. Cette relation est à dissocier de celle de propriété. Le territoire n'appartient pas à son territoire supérieur. Ces relations qui sont matérialisées par la hiérarchie

15. <https://cidoc-crm.org/Entity/e5-event/version-6.2>

16. <https://www.cidoc-crm.org/Entity/e39-actor/version-6.2> provenant de CIDOC-CRM

relèvent de ce que les historiens nomment *l'état de fait*. Cet état de fait n'est pas toujours l'image la plus précise de la réalité historique mais traduit l'interprétation par les instances les plus supérieurs de la hiérarchie à une date donnée.

A cet état de fait, on doit donc ajouter la notion de *revendication*. Une revendication provient d'un acteur qui souhaiterait l'établissement d'une nouvelle relation hiérarchique dans l'état de fait. Cet acteur peut être une institution dirigeant déjà un territoire, un groupe de personnes membres d'un territoire ou une personne seule. Dans HHT, cette relation de revendication est réifiée à travers la notion `hht:Claim`. Cette revendication ne dure que durant une période donnée (à travers la relation `hht:validityPeriodOfClaim`) et elle est réalisée par un Acteur à travers la propriété `hht:makeAClaim`. Une revendication cherche à établir une nouvelle relation `hht:hasSubUnit` entre deux territoires. Elle se définit donc par un territoire inférieur (`hht:subTerritory`) et un territoire supérieur (`hht:upperTerritory`) par rapport à la nouvelle relation hiérarchique souhaitée.

Enfin, plusieurs types de `hht:Claim` existent, chacun dépendant de la position de l'acteur source de la revendication :

- `DeclarationUnder` : Une revendication provenant de l'acteur du territoire inférieur qui souhaiterait se placer sous une juridiction plus avantageuse pour lui.
- `ClaimTo` : Une revendication provenant de l'acteur du territoire supérieur qui cherche à placer le territoire inférieur sous sa juridiction.
- `AutonomyRequest` : Une revendication sans territoire supérieur, car cherchant justement la création d'un territoire supérieur pour gouverner le territoire inférieur.

Ces différents types de revendications retenus pour l'instant proviennent d'échanges avec les historiens sur cette question.

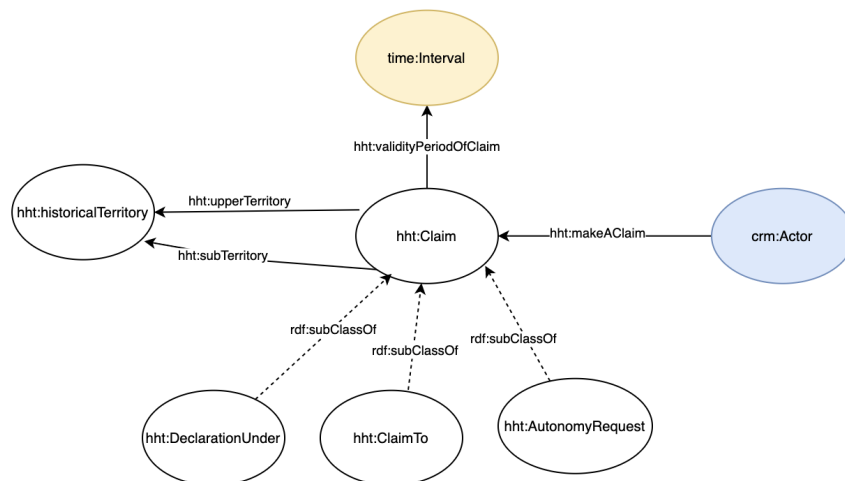


FIGURE 2 : Schéma du concept de revendication dans HHT

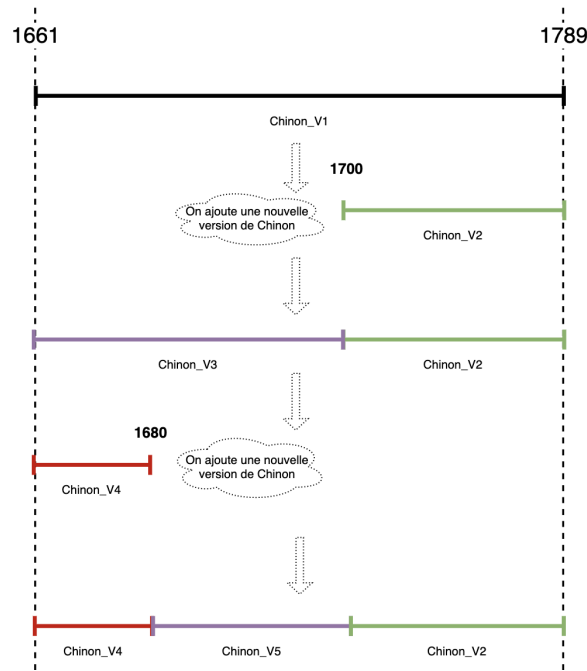


FIGURE 3 : Différents états pris par la trajectoire de vie d'une unité au fur et à mesure que le graphe de connaissances se peuple.

4. Une construction incrémentale du graphe de connaissances

Le graphe de connaissance d'ObARDI représente l'état de connaissances à propos de l'ensemble des différentes versions des territoires de la France d'Ancien Régime, ainsi que les états de connaissances, désormais obsolètes, par lesquels le graphe est passé. Ce graphe se construit donc de manière continue et incrémentale en lui ajoutant régulièrement de nouvelles ressources.

Au début d'un processus de recherche, on ne fixe que le cadre, et les différents critères hiérarchiques qui seront à l'étude. Ensuite, en consultant les sources, on fixe les différents niveaux hiérarchiques qui correspondent à ces critères. Puis, vient le minutieux et progressif travail de l'analyse des sources pour remplir ce graphe d'unités territoriales.

On considère qu'au premier ajout d'une version d'une de ces unités, la période de référence est par défaut définie sur toute la période d'étude (dans notre cas, 1661-1789) et ceci jusqu'à preuve du contraire.

Lorsqu'une nouvelle version d'unité territoriale vient s'ajouter dans le graphe, les précédentes informations du graphe sont considérées obsolètes. Les différentes versions d'une UT vont donc se modifier en conséquence. Le schéma 3 montre les différents états successifs de la trajectoire de vie d'une unité (ici Chinon) lors de la mise à jour du graphe.

5. Conclusion

L'ontologie HHT permet de représenter des hiérarchies territoriales flexibles et adaptées à la recherche en histoire. Elle ne se concentre que sur l'aspect géométrique et hiérarchique de ces territoires, mais elle permet de les représenter, indépendamment de l'existence d'une source de données décrivant l'entièreté du territoire. Elle facilite une gestion locale des périodes de validité de chacune des unités.

HHT se place dans une perspective historique, s'inscrivant dans les humanités numériques, tandis que TSN est à portée statistique pour proposer une norme commune à tous les découpages territoriaux actuels. Dans cet objectif, HHT cherche également à modéliser les états passés des connaissances historiques. Elle s'aligne aussi avec CIDOC-CRM concernant la notion d'évènement.

Cette ontologie devra être étendue à d'autres concepts pour aider la recherche en histoire. Un aspect non encore modélisé est la notion d'*identité d'un territoire* au-delà des différents critères hiérarchiques. La question se pose en effet de l'existence d'un concept à l'intersection des différentes hiérarchies représentant la même zone.

La notion de revendication *hht:Claim* va également se retrouver centrale lorsqu'il va falloir représenter la notion de front juridictionnel en histoire, ce qui consiste à modéliser des zones géographiques de conflit regroupant plusieurs revendications provenant d'une même cause historique. Modéliser et visualiser ces fronts juridictionnels évoluant au cours du temps est une des finalités du projet ObARDI.

Enfin, un des derniers aspects de la recherche historique dans lequel nous aimerions ensuite poursuivre est la gestion des sources. Le concept d'*HistoricalResource* dont dépendra la plupart de nos concepts dans HHT, permettra d'indiquer la source de chaque affirmation d'existence d'un territoire, d'un acteur, d'un changement, ou d'un niveau hiérarchique. Dans cette perspective, l'ontologie symogih sera considérée [18].

De nombreuses pistes restent donc encore à explorer pour affiner et étendre ce modèle. L'ontologie sera publiée sur le web ainsi qu'un site dont l'objectif sera de faciliter la construction et la visualisation du graphe de connaissance qui sera partagé par la communauté.

Références

- [1] C. Bernard, Immersing evolving geographic divisions in the semantic Web, Ph.D. thesis, Université Grenoble Alpes, 2019.
- [2] M. Villanova-Oliver, Représentations de connaissances spatiales évolutives : des ontologies aux géovisualisations, Ph.D. thesis, Communauté Université Grenoble Alpes, 2018.
- [3] A. Ezoji, DOTK : territorial ontology as a tool to help the industries for sustainable development JURY, Ph.D. thesis, Université de Technologie de Troyes, 2019.
- [4] G. Hiebel, M. Doerr, Ø. Eide, Crmgeo : A spatiotemporal extension of cidoc-crm, International Journal on Digital Libraries 18 (2017) 271–279.
- [5] W. Blockmans, The origins of the modern state in Europe : 13th to 18th centuries, Clarendon Press, 1995.

- [6] W. P. Blockmans, A. Holenstein, J. Mathieu, Empowering interactions : political cultures and the emergence of the state in Europe, 1300-1900, Ashgate Publishing, Ltd., 2009.
- [7] P. Garbacz, B. Szady, A. Ławrynowicz, Identity of historical localities in information systems, *Applied Ontology* 16 (2021) 55–86.
- [8] N. Guarino, Some ontological principles for designing upper level lexical resources, in : First International Conference on Language Resources and Evaluation, Granada, Spain, volume 1, 1998, pp. 527–534.
- [9] P. Grenon, B. Smith, Snap and span : Towards dynamic spatial ontology, *Spatial Cognition & Computation* 4 (2004) 104 – 69.
- [10] N. Guarino, C. Welty, Identity and subsumption, in : *The Semantics of Relationships*, Springer, 2002, pp. 111–126.
- [11] T. Kauppinen, E. Hyvönen, Modeling and reasoning about changes in ontology time series, in : *Ontologies*, Springer, 2007, pp. 319–338.
- [12] J. R. Hobbs, F. Pan, Time ontology in owl, W3C working draft 27 (2006) 133.
- [13] R. Battle, D. Kolas, Geosparql : enabling a geospatial semantic web, *Semantic Web Journal* 3 (2011) 355–370.
- [14] G. Bruseker, N. Carboni, A. Guillem, Cultural heritage data management : the role of formal ontology and cidoc crm, *Heritage and Archaeology in the Digital Age* (2017) 93–131.
- [15] M. C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The neon methodology for ontology engineering, in : *Ontology engineering in a networked world*, Springer, 2012, pp. 9–34.
- [16] V. Gay, S. Poublanc, J.-L. Demonsant, Plan de gestion de données du projet ANR ObARDI, Ph.D. thesis, MSHS Toulouse ; FRAMESPA ; TSE-Toulouse School of Economics ; IAST-Institute ..., 2021.
- [17] S. Peroni, A simplified agile methodology for ontology development, in : *OWL : Experiences and Directions–Reasoner Evaluation*, Springer, 2016, pp. 55–69.
- [18] F. Beretta, L’interopérabilité des données historiques et la question du modèle : l’ontologie du projet SyMoGIH, Presses universitaires de Paris Nanterre, 2017. URL : <https://halshs.archives-ouvertes.fr/halshs-01559816>.