



HAL
open science

TSN et TSN-change : ontologies pour représenter l'évolution des découpages territoriaux statistiques

Camille Bernard, Marlène Villanova-Oliver, Jérôme Gensel, Hy Dao

► **To cite this version:**

Camille Bernard, Marlène Villanova-Oliver, Jérôme Gensel, Hy Dao. TSN et TSN-change : ontologies pour représenter l'évolution des découpages territoriaux statistiques. SAGEO 2017 - Spatial Analysis and GEomatics, GdR Magis CNRS, Nov 2017, Rouen, France. hal-01634424v2

HAL Id: hal-01634424

<https://hal.science/hal-01634424v2>

Submitted on 26 Jun 2018

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

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

TSN et TSN-change : ontologies pour représenter l'évolution des découpages territoriaux statistiques

Camille Bernard¹, Marlène Villanova-Oliver¹,
Jérôme Gensel¹, Hy Dao²

1. Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, F-38000 Grenoble,
France

camille.bernard@imag.fr, marlene.villanova-oliver@imag.fr, jerome.gensel@
imag.fr

2. Institute for Environmental Governance and Territorial Development,
University of Geneva, Geneva, Switzerland

hy.dao@unige.ch

RÉSUMÉ. Deux ontologies spatio-temporelles sont présentées ici : TSN, pour Territorial Statistical Nomenclature, et TSN-Change. Ces ontologies permettent la description de découpages territoriaux (en polygones irréguliers), utilisés comme support à la collecte de statistiques. Elles offrent un vocabulaire pour lier les versions successives de ces découpages dans le temps, via des graphes RDF dessinant des changements territoriaux, affectant parfois plusieurs niveaux du découpage. Ces ontologies fournissent aux acteurs des territoires un modèle harmonisé, au plus proche du vocabulaire du domaine, pour la publication de toute Nomenclature Territoriale Statistique dans le Web des Données Liées.

ABSTRACT. We present in this paper two innovative spatio-temporal ontologies: TSN (i.e., Territorial Statistical Nomenclature) and TSN-Change. These ontologies allow for description of territorial partitions composed of territorial units (irregular polygons), used as a support to the collection of statistics. They enable linking the successive versions of these divisions through RDF graphs that describe territorial changes, potentially affecting several nested levels of the nomenclature. These ontologies provide stakeholders with a harmonized model, as close as possible to the vocabulary of the domain, for the publication of any Territorial Statistical Nomenclature, in the Web of Linked Data.

MOTS-CLÉS : Nomenclature Territoriale Statistique, NUTS, Ontologie du changement, Spatio-temporelle ontologie, Web des Données Liées

KEYWORDS: Territorial Statistical Nomenclature, NUTS, Change Ontology, Spatio-temporal Ontology, Web of Linked Data

1. Introduction

Une *Nomenclature Territoriale Statistique* (TSN pour Territorial Statistical Nomenclature en anglais) est un découpage du territoire, utilisé en statistique à des fins de collecte de données socio-économiques. Ces nomenclatures (e.g., *Functional Urban Areas* (FUA)¹ ou *Nomenclature of Territorial Units for Statistics* (NUTS)²) varient par leur forme (e.g., *zonage* couvrant ou non) ou par les territoires qu'elles décrivent (pays, métropoles, etc.) mais ont en commun de diviser les territoires en Unités Territoriales (TU), parfois sur plusieurs niveaux (e.g., pays – *niveau 0*, grandes régions – *niveau 1*, régions – *niveau 2* dans la NUTS) (Bernard *et al.*, 2017). Cet article aborde en particulier les TSN dont les TU sont des polygones irréguliers vectoriels (nous ne traitons pas les formats raster). Le terme *Territoire* désigne ici tout espace géographique délimité par et sous le contrôle d'un groupe humain (souvent une autorité administrative) (Sack, 1983). Ces découpages territoriaux dépendants de politiques varient au cours du temps : le nom d'une communauté de communes peut changer ou les régions administratives d'un pays peuvent être restructurées, comme en France en 2016. Or, les statisticiens doivent disposer de découpages stables lors de la collecte de données afin d'assurer la comparabilité des observations relevées. Ils choisissent alors de figer le découpage du territoire le temps de la collecte de données puis d'en restituer les changements subis dans une nouvelle version à la fin de la période de stabilité arbitrairement déterminée (e.g. trois ans au minimum pour la NUTS³). Si cette pratique résout des problèmes de collecte, il n'en reste pas moins que les séries temporelles statistiques sont brisées, du fait de relevés effectués en ayant recours à des versions différentes d'une même TSN. Ce problème, bien connu sous le nom de *Change of Support Problem* (COSP) ou *Modifiable Areal Unit Problem* (MAUP) (Openshaw, Taylor, 1979 ; Gotway, Young, 2002 ; Howenstine, 1993), décrit le phénomène selon lequel des données, relevées pour des TSN ou versions de TSN différentes, ne sont pas comparables du fait des différences potentielles entre objets géographiques supports. Cette situation peut introduire des biais dans l'interprétation de l'information statistique (Grasland, 1997).

Dans cet article, nous proposons un modèle de description des différences entre deux versions d'une même TSN, afin de répondre, à terme, à plusieurs besoins identifiés par les acteurs des territoires : 1) transférer semi-automatiquement des données statistiques d'une version d'une TSN à une autre pour construire des séries temporelles et ainsi analyser l'impact de politiques passées ; 2) simuler l'évolution d'un territoire (e.g., la fusion avec un territoire adjacent) et observer les effets de ce nouveau découpage à des fins de prospection ; 3) comparer un territoire à un autre, notamment en termes de trajectoire d'évolution

1. http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Metro_regions

2. <http://ec.europa.eu/eurostat/web/nuts/overview>

3. <http://ec.europa.eu/eurostat/web/nuts/history>

territoriale, afin d'étudier la pertinence de suivre ou non une voie similaire (e.g. engager ou non la scission d'une commune). Une des conditions nécessaires au transfert de données statistiques d'une version à une autre d'un découpage est de disposer d'une description précise des changements subis par le territoire impacté à chaque niveau de découpage de celui-ci, i.e. pour chacune des TU concernées (e.g., le changement de surface d'une TU doit être connu ainsi que son impact sur ses TU de niveau inférieur). Ces descriptions territoriales sont nécessaires mais ne sont pour autant pas suffisantes pour répondre complètement aux besoins précédemment évoqués. En effet, assurer le transfert de données nécessite de disposer d'informations relatives à la nature des indicateurs statistiques observés (par exemple, des stocks, des ratios correspondant à des quantités mesurées) et de méthodes de calcul *ad hoc*. En ce sens, notre proposition doit être vue comme un premier pas vers une réponse aux besoins identifiés précédemment. En fournissant des descriptions détaillées des changements subis par le territoire, nous apportons une connaissance utile pour le choix des méthodes statistiques à appliquer pour estimer de nouvelles valeurs selon la nature du changement territorial observé (Goodchild *et al.*, 1980; Flowerdew, 1991). Également, ces descriptions du changement permettent d'identifier des *patterns* d'évolution des territoires, de cibler des territoires dont les trajectoires sont similaires ou opposées pour conduire des comparaisons ou des analyses d'impact de politiques menées.

Notre travail s'inscrit dans le Web des Données Liées (Linked Data - LD) dont les technologies (ontologies et graphes *Resource Description Framework* (RDF) (Cyganiak *et al.*, 2014) notamment) permettent la description de la connaissance liée aux TSN sous une forme compréhensible à la fois par des humains et des machines (Maedche, Staab, 2001). Cela constitue également un support pertinent pour aller vers la semi-automatisation du transfert de données d'une version à une autre. De plus, l'usage de graphes RDF est adapté à la recherche automatique de *patterns* d'évolution de territoires. Également, le recours à ces technologies permet, du fait de la richesse de l'information disponible dans les LD et de son accessibilité, de contextualiser les territoires observés (information relative à la population, la superficie, etc.), mais aussi de contextualiser un changement territorial (réforme territoriale, etc.). Un dernier aspect non négligeable pour les acteurs des territoires est que ces technologies sont un pas vers l'interopérabilité des systèmes et répondent à des directives telles que la *Directive Open Data*⁴ en Europe.

Dans le Web des LD, les ontologies sont des documents qui définissent formellement des concepts et leurs relations (Berners-Lee *et al.*, 2001). Nous présentons ici deux ontologies nommées *TSN Ontology* et *TSN-Change Ontology* pour la description de toute TSN et des changements qu'elle subit au cours du temps. Ces ontologies prolongent les travaux présentés dans (Plumejeaud *et al.*, 2011). Elles vont plus loin dans la description des changements subis par les terri-

4. <https://ec.europa.eu/digital-single-market/en/open-data>

toires au cours du temps, en proposant la modélisation d'événements de fusion ou scission de territoires, mais également en regroupant sous un même nœud des changements ayant lieu à différents niveaux de la TSN (lorsqu'il fait sens de regrouper ces changements pour leur compréhension). Nous proposons ainsi, immergée dans le Web des LD, une représentation détaillée et multi-niveaux des changements territoriaux ayant lieu au sein d'une TSN. Cet article présente les ontologies TSN et TSN-Change qui constituent une première étape nécessaire pour répondre aux trois besoins évoqués plus haut et montre leur capacité à associer des étiquettes sémantiques précises à des changements territoriaux complexes, affectant simultanément plusieurs niveaux d'une TSN.

La section 2 donne un aperçu des travaux connexes aux nôtres ; la section 3 présente les ontologies *TSN Ontology* et *TSN-Change Ontology* ; la section 4 conclut l'article et introduit nos propositions pour la suite des travaux.

2. Travaux connexes et positionnement

2.1. Gestion de versions

Une TSN est un ensemble de connaissances décrivant une structure territoriale que l'on peut représenter sous la forme d'une ontologie. La structure territoriale pouvant évoluer au cours du temps, une TSN existe généralement en plusieurs versions, chacune visant à rapporter des changements territoriaux survenus au cours du temps. (Noy, Klein, 2004) note que, dans le contexte des ontologies, il faut différencier les *changements du domaine ou du monde réel*, des *changements de conceptualisation* (nommés aussi dans la littérature *gestion des versions ou évolution d'une ontologie*). Les modifications du découpage d'un territoire relève du premier type (Kauppinen, Hyvönen, 2007). En effet, ce n'est pas la structure de description qu'est la TSN (i.e. le contenant) qui change avec une nouvelle version, mais bien son contenu. Les concepts de description de la connaissance territoriale, leurs propriétés et relations restent les mêmes d'une version à l'autre, alors que les instances de ces concepts, les valeurs de propriétés et de relations peuvent eux évoluer.

En référence au domaine de l'évolution des systèmes logiciels (Tichy, 1988), notre approche s'inscrit dans le courant de la gestion des versions vues comme des *variantes* (versions d'un même item destinées à coexister) plutôt que comme des *révisions* (versions destinées à remplacer la précédente) (Conradi, Westfechtel, 1998). Les versions d'une TSN doivent en effet coexister afin d'observer l'influence du découpage sur les données statistiques (Grasland, 1997). Du côté des LD, les spécifications RDF indiquent que le modèle de données est atemporel, mais que les graphes RDF peuvent décrire des événements et leur temporalité, *via la définition de concepts ontologiques appropriés*. Il est ainsi recommandé d'utiliser les vocabulaires existants (tels que OWL-Time⁵)

5. <https://www.w3.org/TR/owl-time/>

pour l'interopérabilité sémantique entre systèmes⁶. Exprimer différents états d'une ressource au fil du temps, grâce à différents graphes RDF inclus dans une ressource "contenant", est également soutenu par le W3C⁷ (Cyganiak *et al.*, 2014). Dès lors, nous proposons de représenter une TSN via une ressource RDF, avec ses versions successives, représentées par différents graphes RDF, ayant chacun un URI pour l'identification unique des états de la TSN au fil du temps. (Stefanidis *et al.*, 2014) critique cette approche qui peut s'avérer lourde en termes d'espace de stockage des données. Une alternative consiste à ne stocker qu'une seule version (la première, le *version model* selon (Conradi, Westfechtel, 1998)) et les *deltas* (les différences entre deux versions consécutives). Néanmoins, cette solution requiert la génération des versions à chaque requête adressée aux systèmes, ce qui peut entraîner une surcharge, au moment du calcul cette fois-ci. En termes d'usages, les différentes versions de TSN sont des variantes destinées à coexister dans le Web des LD sous forme de documents directement consultables (e.g. accéder à une version de NUTS en particulier). Ces versions ne constituent par ailleurs pas de grands ensembles de données qui évoluent rapidement (e.g., NUTS change tous les trois ans). L'approche que nous adoptons, appelée *Materialization Using Versions and Deltas* (Stefanidis *et al.*, 2014), consiste donc à conserver toutes les versions d'une même TSN et à décrire les deltas entre deux versions consécutives d'une TSN, car ces deltas, dans le contexte de la statistique, sont essentiels (voir section 1).

2.2. Ontologies pour les TSN et les changements territoriaux

De nombreuses TSN définissent différents niveaux d'observation d'un territoire. Parmi elles, la nomenclature *Australian Standard Geographical Classification (ASGC)*⁸ divise en sept niveaux imbriqués le territoire australien. Chaque découpage (e.g., Section of State, Collection District, Statistical Local Area) est conçu pour répondre à différents usages statistiques. Différentes versions de cette TSN sont disponibles, révisées chaque année depuis 1984⁹. En Europe, la NUTS offre un découpage de l'ensemble du territoire, à plusieurs niveaux. De nombreux instituts statistiques ont également créé leur propre ontologie pour la description et la publication de leur TSN dans le Web des LD. Ainsi, l'INSEE publie l'*Ontologie géographique de l'INSEE*¹⁰, dont les concepts décrits en français sont extrêmement dépendants du contexte territorial national (e.g., concept *ArrondissementMunicipal*). L'Ordnance Survey (agence gouvernementale chargée de la cartographie du Royaume-Uni) publie l'ontologie *Administra-*

6. <https://www.w3.org/2001/sw/BestPractices/OEP/SemInt/>

7. World Wide Web Consortium <https://www.w3.org/>

8. <http://www.abs.gov.au/websitedbs/D3110124.NSF>

9. <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Standard+Geographical+Classification+%28ASGC%29>

10. <http://rdf.insee.fr/def/geo/insee-geo-onto.ttl>

*tive geography and civil voting area ontology for UK*¹¹. Elle permet de décrire les niveaux de la NUTS, mais la dimension temporelle des objets géographiques composants est oubliée (Correndo *et al.*, 2010). Là aussi, les concepts définis dépendent fortement du contexte britannique. Eurostat propose l'ontologie *Reference And Management Of Nomenclatures Ontology (RAMON)*¹² qui permet la description des TSN NUTS et *Local Administrative Units (LAU)*¹³. Cependant, pour décrire des TU provenant d'une autre TSN que la NUTS ou LAU, il est nécessaire de modifier l'ontologie et d'ajouter un sous-concept au concept "Geographical Region". L'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) publie l'ontologie *FAO Geopolitical Ontology*¹⁴ et un jeu de données contenant : 1) une liste de pays du monde, décrits en RDF grâce aux concepts géopolitiques de l'ontologie FAO (par exemple, *self-governing*, *non-self-governing*, *disputed*) et 2) les versions successives de ces pays (via les prédicats *predecessor*, *successor*, *valid since*, *valid until*). Cependant, les changements subis par un pays, d'une version à l'autre, ne sont pas décrits. Concernant plus particulièrement le support des différentes versions de la NUTS, on peut noter que Eionet publie au format RDF¹⁵ des versions décrites via l'ontologie *RAMON* dont il propose une version simplifiée¹⁶. Il existe deux types de liens (*owl:sameAs*, *rdfs:seeAlso*) pour connecter les TU de la NUTS version 2006 aux TU de la version 2003, mais les changements subis d'une version à l'autre ne sont pas décrits. Le projet EnAKTing¹⁷ publie plusieurs versions de la NUTS au format RDF, sous le nom *Linked NUTS* (Correndo, Shadbolt, 2013). Les *Linked NUTS* décrivent les modifications subies par les TU de la NUTS, au fil des versions, mais la typologie des changements utilisée est restreinte (changement de code d'une TU, fusion, scission de TU), alors qu'une typologie plus fine doit être utilisée pour décrire les changements et éviter par la suite une interprétation erronée des données statistiques.

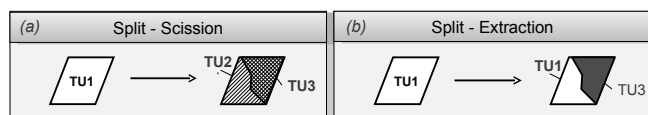


FIGURE 1. *Division d'une TU: distinction entre Scission et Extraction*

Aucune des ontologies présentées ci-dessus n'a un niveau d'abstraction suffisant pour décrire n'importe quelle TSN. De plus, seules l'INSEE et ENACKTing proposent une description des changements qui se produisent d'une version

11. <http://data.ordnancesurvey.co.uk/ontology/admingeo/>

12. <http://ec.europa.eu/eurostat/ramon/ontologies/geographic.rdf>

13. <http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>

14. <http://www.fao.org/countryprofiles/geoinfo/en/>

15. <http://rdfdata.eionet.europa.eu/ramon/nuts.rdf> <http://rdfdata.eionet.europa.eu/ramon/nuts2008.rdf> <http://rdfdata.eionet.europa.eu/ramon/nuts2003.rdf>

16. <http://rdfdata.eionet.europa.eu/ramon/ontology.rdf>

17. <http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/G008493/1>

à une autre. Il est cependant utile et possible d'enrichir ces descriptions sur la base des travaux de (Plumejeaud *et al.*, 2011) traitant particulièrement le cas des événements territoriaux dans les TSN. La *Typologie des changements territoriaux* proposée distingue par exemple deux types de *Scission* de TU (présentés Figure 1). Dans le premier cas (a) *Scission*, toutes les TU impliquées disparaissent alors que, dans le cas (b) d'une *Extraction*, la TU (*TU1*) existant avant le changement persiste (son identifiant dans la nomenclature reste inchangé alors que sa surface diminue, en raison de la création de la nouvelle *TU3*). L'approche que nous adoptons est celle décrite dans les ontologies de type *fluent*, où l'on distingue les attributs porteurs de l'identité d'un objet (qui, s'ils varient, entraînent la création d'un nouvel objet), des autres attributs (qui, s'ils varient, entraînent la création d'un nouveau sous-objet, appelé *TimeSlice*, porteur de tous les attributs de l'objet variables dans le temps et valables donc uniquement pour une période de temps) (Welty *et al.*, 2006). Cette approche est souvent utilisée dans le Web des LD, en complément de l'ontologie *OWL-Time* pour représenter l'évolution des ressources (Batsakis, Petrakis, 2011), notamment géospatiales (Harbelot *et al.*, 2015 ; Tran *et al.*, 2015) (voir Figure 2, approche (1)). Le modèle (Plumejeaud *et al.*, 2011) n'est pas publié dans le Web de LD, contrairement à l'ontologie *Change Ontology*¹⁸ de (Kauppinen *et al.*, 2008 ; Kauppinen, Hyvönen, 2007). Les auteurs introduisent la notion de *Change Bridge* (de type *Merge* par exemple), pour chaîner des territoires anciens (e.g. *Allemagne de l'Est* et *Allemagne de l'Ouest*) aux territoires suivants (e.g. *Allemagne*) (voir Figure 2, approche (2)). Cette approche est critiquée par (Lacasta *et al.*, 2014) car elle peut conduire à une prolifération des instances. En effet, l'état complet de la TU est décrit avant et après le changement. Cette approche, comme l'approche *fluent* (que nous jugeons pourtant nécessaires à la représentation de l'évolution des objets), ne résout pas le problème de prolifération des instances (Batsakis, Petrakis, 2011). Afin de limiter ce problème, nous décidons de ne pas dupliquer les TU dont aucun attribut n'aurait changé d'une version à l'autre : la première version de TU porte l'ensemble des propriétés et un lien est établi entre les deux versions de TU identiques.

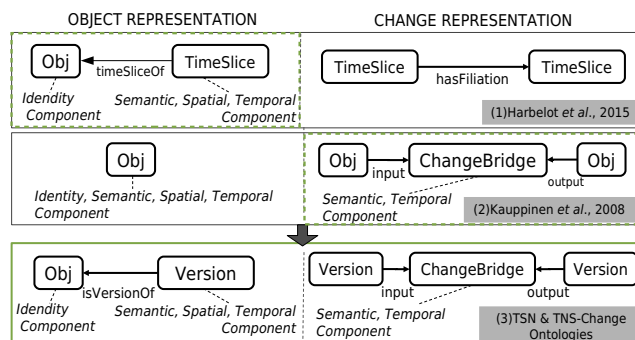


FIGURE 2. Association de deux approches pour le modèle TSN-TSN-Change

18. <http://linkedearth.org/change/ns/>

Notre modèle, décrit ci-après, est trans-temporel : le temps des données statistiques est dissocié du temps de la partition territoriale pour observer des données socio-économiques dans différentes versions d'un découpage. Nous cherchons à constituer des graphes, caractérisant les changements affectant des TU statistiques car c'est cette information qui sera utile pour déterminer les méthodes de calcul à appliquer aux indicateurs socio-économiques pour estimer les observations statistiques dans une version du découpage antérieure ou postérieure. Ces graphes seront composés de nœuds décrivant le changement, auxquels nous pourrions associer une date et des informations contextuelles pouvant expliquer le changement en exploitant la puissance du Web des DL. Nous ne cherchons donc pas uniquement à lier des TU dans le temps, mais à décrire les changements survenus. Nous adoptons l'approche de (Kauppinen *et al.*, 2008) pour l'union de versions successives via des *Change Bridge*, et l'approche *fluent* pour la description des évolutions des objets, à ceci près que nous n'optons pas pour son vocabulaire mais privilégions celui des statisticiens et acteurs du territoire (voir Figure 2, approche (3)). Ainsi, là où les ontologies *fluent* utilisent le terme *TimeSlice*, nous utilisons le terme *Version*.

3. Ontologies TSN et TSN-Change

L'ontologie *TSN-Ontology* permet la description de toute TSN, ce qu'aucune ontologie ne permettait jusqu'à présent à notre connaissance. Notre ontologie peut être spécialisée via des concepts d'ontologies d'application, pour adresser les particularités d'un territoire (les ontologies de l'Ordonance Survey peuvent se relier à notre ontologie pour préciser que le territoire observé est le Royaume-Unis et que les TU sont des *districts* par exemple). *TSN-Ontology* définit un vocabulaire que l'ensemble des acteurs des territoires mais aussi, du fait de son ancrage dans le Web LD, les machines peuvent partager, veillant ainsi à l'interopérabilité entre les systèmes dans le domaine des partitions territoriales utilisées pour la statistique. Cette ontologie introduit le concept de *Territory* (assimilable à l'aire d'étude de la TSN, la TU englobante) et différencie cet objet *Territory* des TU qui le composent. Ainsi, il est désormais possible de décrire des changements du type *Fusion de deux territoires* entre versions de TS qui, par propagation, peuvent affecter chaque niveau du découpage du territoire.

Selon la méthodologie de construction d'une ontologie de (Bachimont *et al.*, 2002), un corpus des TSN et ontologies existantes (basé essentiellement sur le corpus présenté section 2.2) nous a aidé à déterminer le vocabulaire à utiliser pour être proche des désignations employées par les acteurs et ainsi faciliter l'utilisation du modèle. Techniquement parlant, les deux ontologies *TSN* et *TSN-Change* sont écrites en triplets RDF, sous la forme $\langle \textit{subject}, \textit{predicate}, \textit{object} \rangle$. Il s'agit d'ontologies géographiques qui définissent des concepts géographiques (e.g., une *ville*) et non géométriques (Klien, Probst, 2005). Les figures 3 et 5 présentent, par des graphes RDF simplifiés, les principales classes et prédicats des deux ontologies. Les *prédicats* apparaissent le long des flèches. Le départ

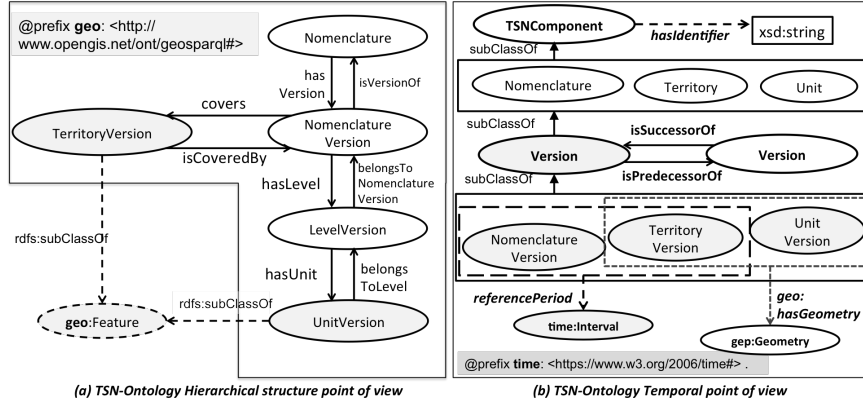


FIGURE 3. TSN Ontology principales classes et prédicats

d'une flèche indique le *domaine* et l'arrivée, le *co-domaine* ou *objet*. Les définitions des concepts sont accessibles depuis les URI <http://purl.org/net/tsn#> et <http://purl.org/net/tsnchange#>. Les éléments de la Figure 3 (a) reproduisent la hiérarchie des éléments composant une TSN, du concept englobant de *Nomenclature* aux plus petits éléments qui peuvent le composer, les *UnitVersion*. Les concepts *TerritoryVersion* et *UnitVersion* sont des sous-classes du concept *geo:Feature* défini par l'OGC dans l'ontologie *Geosparql*¹⁹. La figure 3 (b) montre la mise en œuvre de la modélisation en *fluents* (que nous nommons *Version*) pour la représentation des différentes versions des composants d'une TSN. Le vocabulaire *OWL-Time* est utilisé pour assigner une période de validité à une *NomenclatureVersion* et à un *TerritoryVersion*. Nous choisissons de n'assigner cette période de validité qu'à l'élément père (*NomenclatureVersion*) dans la hiérarchie des composants d'une *Nomenclature*, cette période s'appliquant par propagation aux éléments fils.

Les principaux prédicats qui composent la *TSN Ontology* sont :

- **hasVersion**(*x*, *y*) qui signifie "le TSNComponent *x* a une Version *y*".
- **isCoveredby**(*x*, *y*) qui signifie "la TerritoryVersion *x* est couverte par la NomenclatureVersion *y*".
- **hasLevel**(*x*, *y*) qui signifie "la NomenclatureVersion *x* est composée d'un LevelVersion *y*".
- **hasUnit**(*x*, *y*) qui signifie "le LevelVersion *x* est composé d'une UnitVersion *y*".

Pour chaque prédicat, une propriété inverse est définie, afin de circuler de bas en haut dans la hiérarchie d'une TSN. Les prédicats **hasLowerLevel**, **hasUpperLevel**, **hasSubUnit**, **hasSuperUnit** expriment une relation hiérarchique entre les éléments d'une TSN.

19. <http://www.opengis.net/ont/geosparql>

Comme une Nomenclature ou une TU, un territoire peut changer dans le temps, mais pourtant conserver son identité. Par exemple, l'organisation politique de l'Union Européenne (UE) a connu beaucoup de variations dans le temps, du fait de l'adhésion successive de pays (UE des 15 (1995-2004), des 25 (2004-2006) et 27 (2007-2013) États membres). La typologie des changements que nous proposons dissocie les cas de filiation *continuation*, des cas de *dérivation*, selon que l'attribut d'identité change ou non entre deux versions (Del Mondo, 2011). Deux prédicats permettent d'exprimer un ordre dans la succession de versions et circuler directement dans le graphe, d'une version à une autre, dans des cas de filiation du type *continuation* observés entre ces deux versions (cas de la préservation de l'identité) :

- `isPredecessorOf(x,y)` qui signifie "la Version x précède la Version y".
- `isSuccessorOf(x,y)` qui signifie "la Version x succède à la Version y".

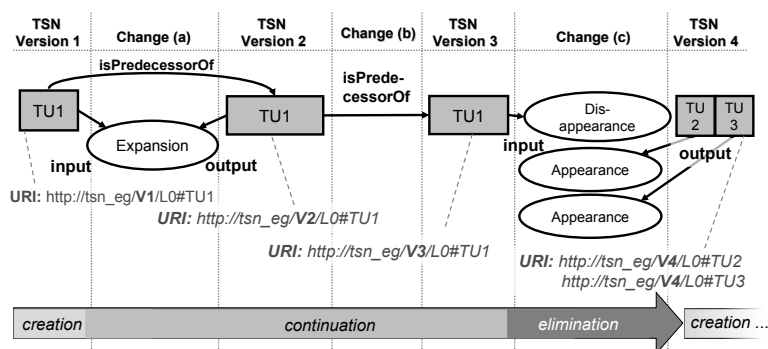


FIGURE 4. TSN-Change Ontology Graphe simplifié - ligne de vie d'une TU

Nous choisissons, contrairement à (Harbelot *et al.*, 2015), de ne pas créer de prédicats (de liens directs) qui permettraient d'unir une TU à une autre dans des cas de *dérivation* (où l'identité de la TU change) car, dans bien des cas, ces liens risqueraient de ne pas être pertinents, dans le contexte de la statistique. Par exemple, dans des cas de redistribution, modifiant les frontières de plusieurs TU (sans en respecter le tracé d'origine) mais aussi leurs identifiants, il est préférable de déterminer l'ensemble des TU impactées en entrée du changement et l'ensemble des TU créées en sortie de ce changement et de les relier toutes via un nœud `tsnchange:Reallocation` (voir Figure 5). Ici, l'approche *Change Bridge* de (Kauppinen, Hyvönen, 2007) est donc plus adaptée au domaine des TSN et à la nécessité de décrire le changement. La Figure 4 montre les trois types de liens *Change (a), (b) et (c)* que nous pouvons créer entre des TU, d'une version à une autre de la nomenclature. *(a)* montre un lien de filiation (`tsn:isPredecessorOf`) et la description du changement d'un attribut non porteur de l'identité de l'objet ; *(b)* illustre un lien de filiation (`tsn:isPredecessorOf`) uniquement car aucun autre attribut n'a changé ; *(c)* montre des liens indirects (lien de *dérivation*) via un graphe décrivant les changements subis par les TU qui perdent leur identité au cours du changement. Cette figure isole, au sein

du graphe des changements, les nœuds et les liens qui décrivent la ligne de vie d'une TU en particulier et fournissent ainsi une vue des trois états principaux *creation*, *continuation*, *elimination* attendus dans un modèle traitant le problème d'évolution de l'identité (Hornsby, Egenhofer, 2000).

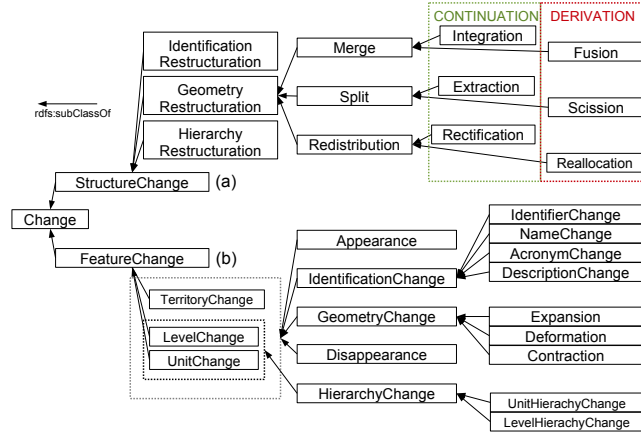


FIGURE 5. *TSN-Change Ontology principales classes et prédicats*

L'ontologie *TSN-Change Ontology* permet la description de graphes des changements territoriaux multi-niveaux. Les principaux éléments de cette ontologie sont présentés Figure 5. Les changements sont décrits à chaque niveau de la TSN impacté et un changement ayant lieu à un niveau de la TSN est relié à des changements ayant lieu plus bas ou plus haut dans la hiérarchie TSN, s'il fait sens de relier ces changements (via les prédicats *causes*, *isCausedBy*). Nous adoptons ici l'approche développée dans (Plumejeaud *et al.*, 2011) et considérons qu'un changement est rarement isolé et indépendant des autres changements qui se produisent simultanément dans d'autres unités, dans une zone donnée dont les limites extérieures sont les mêmes avant et après le changement. Nous étendons le modèle *Typology of Territorial Changes* (Plumejeaud *et al.*, 2011) afin de décrire des changements affectant les frontières extérieures du territoire observé. Ces nouvelles étiquettes permettent aux analystes de dissocier les événements survenus dans un territoire dont les limites n'ont pas changé, des événements qui modifient l'étendue spatiale de la zone d'étude (apparaissant sous un nœud *TerritoryChange*). Nous ajoutons à cette approche une représentation multi-niveaux des changements territoriaux du fait de la nature même d'une TSN, qui divise les territoires sur plusieurs niveaux.

Prenons l'exemple d'une fusion de deux territoires qui entrainerait une modification des TU, existant avant la fusion, et ce, sur plusieurs niveaux de la TSN. Le "pont" du changement présente la structure suivante (décrite également Figure 6 où le territoire couvert par la nomenclature apparaît en vert et le territoire annexé, en gris avant l'intégration) : (1) le nœud père de ce changement est un nœud de type *Integration* (voir la typologie (a) *Structure-Change*, Figure 5). Cette étiquette indique que plusieurs *TSNComponent* (les

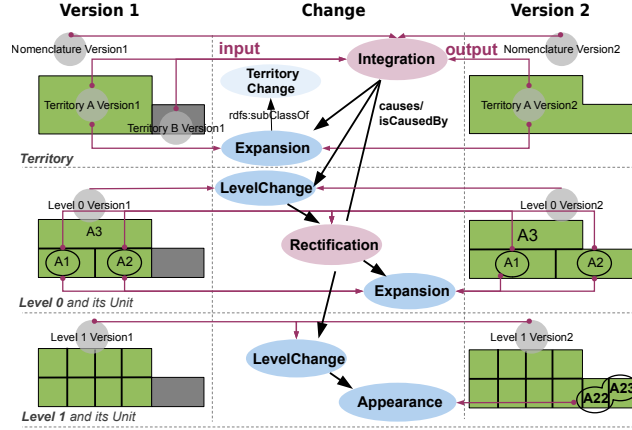


FIGURE 6. TSN-Change : graphe multi-niveaux simplifié - fusion territoires

territoires *Territoire A* et *B* en *INPUT* du changement) fusionnent et qu'un des territoires, le *Territoire A* en *OUTPUT*, garde son identité suite au changement (i.e., changement de type *Continuation*); (2) chacune des modifications provoquées par cette fusion est ensuite représentée, à chaque niveau de la TSN impacté; (3.1) l'impact sur l'objet *Territoire A Version 1* est d'abord décrit en terme d'*Expansion* (étiquette appartenant à la typologie (b) *FeatureChange* présentée Figure 5); (3.2) le niveau 0 de la nomenclature subit un changement (nœud *LevelChange* entraînant une *Rectification* du tracé de ses TU); (3.3) le niveau 1 de la nomenclature subit lui aussi un changement, mais contrairement au niveau 0, ses TU appartenant à la version 1, ne subissent pas de modification de frontières; (4) enfin, chaque changement subit par une TU est décrit; (4.1) au niveau 0 de la nomenclature, cette fusion des territoires provoque une *Expansion* des TU *A1* et *A2*; (4.2) et au niveau 1, la création des TU *A22* et *A23* (*Appearance*).

Un nœud **Change** est défini par le quintuplet :

$$\langle t, INPUT, OUTPUT, causes, isCausedBy \rangle$$

où t assigne un point au changement dans le temps, *INPUT* est l'ensemble des ressources en entrée du changement et impactées par celui-ci, *OUTPUT* est l'ensemble de ressources créées ou modifiées suite au changement, *causes* est l'ensemble des changements causés par le changement *Change* actuellement décrit et, *isCausedBy* indique le changement *Change* supérieur qui a provoqué le changement actuellement décrit. Avec ce quintuplet, tous les sous-concepts de *Change* peuvent être décrits. Nous montrons ici comment il nous permet de dissocier les deux sous-types du changement **Split**.

Une **Scission** est définie par :

$$\langle t, \{input\}, \{output_1, \dots, output_n\}, \{Disappearance(input), Appearance()_1, \dots, Appearance()_n\}, undetermined \rangle \quad (1)$$

où $n \geq 2$; $input = output_1 \cup output_2 \cup \dots \cup output_n$; $Disappearance(input) \rightarrow \emptyset$;
 $Appearance() \rightarrow output$ avec $id_{output} \neq id_{input}$ ²⁰.

Une **Extraction** est définie par :

$$\begin{aligned} &< t, \{input\}, \{output_1, \dots, output_n\}, \\ &\{Contraction(input), Appearance()_1, \dots, \\ &Appearance()_m\}, undetermined > \end{aligned} \quad (2)$$

où $n \geq 2$; $m \geq 1$; $input = output_1 \cup output_2 \cup \dots \cup output_n$;
 $Contraction(input) \rightarrow output$ avec $id_{input} \text{ owl:sameAs } id_{output}$,
 $area_{input} > area_{output}$; $Appearance() \rightarrow output$ avec $id_{output} \neq id_{input}$ ²¹.

Faute de place, cet article ne présente pas les requêtes que nous sommes en mesure de construire pour exploiter notre modèle (par exemple pour obtenir la liste de tous les nœuds pères d'un changement ayant lieu entre deux versions de nomenclature ou pour identifier la ligne de vie d'une TU ou d'un territoire). De même, nous ne montrons pas comment notre modèle s'articule avec l'ontologie RDF-Data-Cube ²² pour la représentation d'observations statistiques.

4. Conclusion

Nous avons présenté les ontologies *TSN* et *TSN-Change* permettant la description de versions de découpages territoriaux utilisés comme support de collecte et d'estimation de données statistiques. Ces ontologies participent d'une approche nouvelle pour la description des changements dans ces découpages territoriaux statistiques : la combinaison de l'approche *Change Bridge* à l'approche *fluent*, nous permet de construire des graphes du changement multi-niveaux dans le Web des LD. En proposant un vocabulaire harmonisé, aussi bien pour la description des éléments qui composent une TSN que pour la description des changements de ces éléments, nous avons pour ambition d'améliorer l'interopérabilité entre des systèmes échangeant des données statistiques territoriales.

Nous travaillons actuellement sur la publication de plusieurs TSN (parmi celles d'Eurostat et de l'INSEE) et leurs versions dans le Web des LD afin de tester notre approche sur des cas concrets. Des jeux de données statistiques, concernant les territoires ainsi décrits, seront également publiés par le biais du vocabulaire *RDF Data Cube Vocabulary* (Cyganiak *et al.*, 2014) et de son extension *QB4ST* pour une déclaration canonique de la dimension spatiale d'une observation statistique (Atkinson, 2017). Disposant ainsi dans le Web LD, de plusieurs TSN, de plusieurs versions de ces TSN et de jeux de données statistiques pour les territoires décrits, nous nous consacrerons à la recherche de *patterns* d'évolution des territoires dans le temps. Par ailleurs, l'automatisation

20. *id* pour *identifier*

21. *area* pour *surface area* d'une TU

22. <https://www.w3.org/TR/vocab-data-cube/>

de la création des graphes du changement (via des raisonnements spatiaux et sémantiques sur les données) sera explorée. Nous souhaitons construire un outil capable d'exploiter des versions de TSN (sous la forme de shapefile) pour créer le graphe RDF de ces versions et des changements entre elles.

5. Remerciements

Nos remerciements vont à la Région Française **Auvergne-Rhône-Alpes** ayant financé ce travail [grant number REGION 2015-DRH-0367]. ARC7 : Innovations, mobilités, territoires et dynamiques urbaines.

Bibliographie

- Atkinson R. (2017). *QB4st: RDF Data Cube extensions for spatio-temporal components*.
- Bachimont B., Isaac A., Troncy R. (2002). Semantic commitment for designing ontologies: a proposal. In *International Conference on Knowledge Engineering and Knowledge Management*, p. 114–121. Springer.
- Batsakis S., Petrakis E. G. (2011). SOWL: a framework for handling spatio-temporal information in OWL 2.0. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, p. 242–249. Springer.
- Bernard C., Villanova-Oliver M., Gensel J., Le Rubrus B. (2017, mars). Spatio-Temporal evolutive Data Infrastructure: a Spatial Data Infrastructure for managing data flows of Territorial Statistical Information. *International Journal of Digital Earth*, vol. 10, n° 3, p. 257–283.
- Berners-Lee T., Hendler J., Lassila O., others. (2001). The semantic web. *Scientific american*, vol. 284, n° 5, p. 28–37.
- Conradi R., Westfechtel B. (1998). Version models for software configuration management. *ACM Computing Surveys (CSUR)*, vol. 30, n° 2, p. 232–282.
- Correndo G., Granzotto A., Salvadores M., Hall W., Shadbolt N. (2010). A linked data representation of the nomenclature of territorial units for statistics.
- Correndo G., Shadbolt N. (2013). Linked nomenclature of territorial units for statistics. *Semantic Web*, vol. 4, n° 3, p. 251–256.
- Cyganiak R., Wood D., Lanthaler M. (2014). *RDF 1.1 Concepts and Abstract Syntax*. Consulté sur <https://www.w3.org/TR/rdf11-concepts/#dfn-rdf-source>
- Del Mondo G. (2011). *Un modèle de graphe spatio-temporel pour représenter l'évolution d'entités géographiques*. Theses, Université de Bretagne occidentale - Brest. Consulté sur <https://tel.archives-ouvertes.fr/tel-00651682>
- Flowerdew R. (1991). Data integration: statistical methods for transferring data between zonal systems. In *Handling geographical information: methodology and potential applications*, p. 39–54.

- Goodchild M. F., Lam N. S. N., Geography U. of Western Ontario. Dept. of. (1980). *Areal interpolation: a variant of the traditional spatial problem*. London, Ont.: Department of Geography, University of Western Ontario.
- Gotway C. A., Young L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, vol. 97, n° 458, p. 632–648.
- Grasland C. (1997). Les maillages territoriaux: niveaux d’observation ou niveaux d’organisation? *Les Découpages du territoire*, p. 115–132.
- Harbelot B., Arenas H., Cruz C. (2015). LC3: A spatio-temporal and semantic model for knowledge discovery from geospatial datasets. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 35, p. 3–24. Consulté sur <http://www.sciencedirect.com/science/article/pii/S1570826815000840>
- Hornsby K., Egenhofer M. J. (2000, avril). Identity-based change: a foundation for spatio-temporal knowledge representation. *International Journal of Geographical Information Science*, vol. 14, n° 3, p. 207–224.
- Howenstine E. (1993, novembre). Measuring demographic change: The split tract problem. *Professional Geographer*, vol. 45, n° 4, p. 425.
- Kauppinen T., Hyvönen E. (2007). Modeling and reasoning about changes in ontology time series. In *Ontologies*, p. 319–338. Springer.
- Kauppinen T., Väätäinen J., Hyvönen E. (2008). *Creating and using geospatial ontology time series in a semantic cultural heritage portal*. Springer.
- Klien E., Probst F. (2005). Requirements for geospatial ontology engineering. In *8th conference on geographic information science (AGILE 2005)*, p. 251–260.
- Lacasta J., Lopez-Pellicer F. J., Florczyk A., Zarazaga-Soria F. J., Nogueras-Iso J. (2014, septembre). Population of a spatio-temporal knowledge base for jurisdictional domains. *International Journal of Geographical Information Science*, vol. 28, n° 9, p. 1964–1987.
- Maedche A., Staab S. (2001). Learning ontologies for the semantic web. In *Proceedings of the Second International Conference on Semantic Web-Volume 40*, p. 51–60. CEUR-WS. org.
- Noy N. F., Klein M. (2004, juillet). Ontology Evolution: Not the Same as Schema Evolution. *Knowledge and Information Systems*, vol. 6, n° 4, p. 428–440.
- Openshaw S., Taylor P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences*, vol. 21, p. 127–144.
- Plumejeaud C., Mathian H., Gensel J., Grasland C. (2011). Spatio-temporal analysis of territorial changes from a multi-scale perspective. *International Journal of Geographical Information Science*, vol. 25, n° 10, p. 1597–1612.
- Sack R. D. (1983, mars). Human Territoriality: A Theory. *Annals of the Association of American Geographers*, vol. 73, n° 1, p. 55–74.
- Stefanidis K., Chrysakis I., Flouris G. (2014). On Designing Archiving Policies for Evolving RDF Datasets on the Web. In E. Yu, G. Dobbie, M. Jarke, S. Purao (Eds.), *Conceptual Modeling: 33rd International Conference, ER 2014, Atlanta*,

GA, USA, October 27-29, 2014. Proceedings, p. 43–56. Cham, Springer International Publishing. (DOI: 10.1007/978-3-319-12206-9_4)

Tichy W. (1988). *Software Configuration Management Overview*. Citeseer.

Tran B.-H., Plumejeaud-Perreau C., Bouju A., Bretagnolle V. (2015). A semantic mediator for handling heterogeneity of spatio-temporal environment data. In *Research Conference on Metadata and Semantics Research*, p. 381–392. Springer.

Welty C., Fikes R., Makarios S. (2006). A reusable ontology for fluents in OWL. In *FOIS*, vol. 150, p. 226–236.