



HAL
open science

Modélisation et analyse de données environnementales à travers une ontologie spatio-temporelle

Ba-Huy Tran, Christine Plumejeaud-Perreau, Alain Bouju, Vincent Bretagnolle

► To cite this version:

Ba-Huy Tran, Christine Plumejeaud-Perreau, Alain Bouju, Vincent Bretagnolle. Modélisation et analyse de données environnementales à travers une ontologie spatio-temporelle. Atelier IN-OVIVE 4ième édition, Jun 2016, Montpellier, France. hal-01329854v2

HAL Id: hal-01329854

<https://inria.hal.science/hal-01329854v2>

Submitted on 30 Sep 2016

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.

Modélisation et analyse de données environnementales à travers une ontologie spatio-temporelle

Ba-Huy Tran¹, Christine Plumejeaud-Perreau²,
Alain Bouju¹, Vincent Bretagnolle³

¹ L3I, Laboratoire Informatique Image et Interaction (L3I), Université de la Rochelle
ba-huy.tran@univ-lr.fr, alain.bouju@univ-lr.fr

² LIENSS, Littoral ENvironnement et Sociétés, U.M.R. CNRS 7266
christine.plumejeaud-perreau@univ-lr.fr

³ CEBC Centre d'Etudes Biologiques de Chizé, U.M.R CNRS 7372 et Université de la Rochelle
breta@cebc.cnrs.fr

Résumé : Cet article présente les travaux pour le projet interdisciplinaire GEMINAT (GéoConnaissances des milieux naturels) qui a pour but d'intégrer et d'exploiter des données environnementales hétérogènes par l'application du web sémantique. À partir d'un cas d'étude mené sur un observatoire environnemental basé à Chizé, nous résumons les besoins d'analyse spatio-temporelle des experts biologistes et écologues envers les bases de données de l'assolement et de la biodiversité. Nous montrons comment la mise en œuvre d'un framework avec une ontologie spatio-temporelle jouant le rôle d'un médiateur sémantique peut résoudre les difficultés d'analyse et de maintenance qu'induisent ces systèmes, amenés à de constantes évolutions de leurs modèles. En particulier, la démonstration de la faisabilité d'un tel système est faite, et nous mesurons sa capacité à répondre à des requêtes complexes mêlant plusieurs sources de données et les dimensions spatiales et temporelles.

Mots-clés : Ontologies, Environnement, Spatio-temporel, Intégration de données.

1 Introduction

La nécessité de collecter des observations sur une longue durée pour la recherche sur les relations entre environnement et anthroposystème a entraîné la mise en place de Zones Ateliers¹ par le CNRS. Elles questionnent les interactions entre un milieu et les sociétés qui l'occupent et l'exploitent. La spécificité des Zones Ateliers réside dans la taille de l'objet d'étude, qui est un territoire de grande dimension.

Ainsi, le Centre d'Etudes Biologiques de Chizé (CEBC) a mis en place un observatoire des assolements et de la biodiversité sur la "Zone Atelier Plaine et Val de Sèvre". Cet observatoire collecte de nombreuses données sur la biodiversité faunistique et floristique, ainsi que l'assolement agricole, avec un suivi de cette même zone sur plus de 20 ans dans diverses bases de données spatio-temporelles. Cependant, l'observation de la rotation des cultures, des insectes de tous types (carabes, libellules, etc.), des plantes, ou des oiseaux ne requièrent ni les mêmes moyens, ni les mêmes méthodes. Ainsi les protocoles de collecte de ces informations diffèrent forcément, ils font l'objet de projets distincts qu'assurent diverses équipes de recherche au CEBC, constituant à terme des bases de données très hétérogènes. Pourtant il existe un fort besoin d'être en mesure de croiser ces informations, de façon assez systématique et souple, pour mener une analyse spatio-temporelle fine des milieux écologiques.

1. <http://www.za-inee.org>

A cet effet, les utilisateurs envisagent la migration des données et schémas dans un système d'information centralisé. Pour cela, il faut mettre en œuvre une médiation entre ces sources de données. En terme de médiation, une des solutions les plus avancées et prometteuse repose sur l'intégration par les techniques du Web sémantique.

La première section expose en détail notre cas d'étude avec les bases de données hétérogènes disponibles et les besoins d'analyse spatio-temporelle envers ces sources de données. La seconde section propose un framework sémantique avec une ontologie spatio-temporelle en tant que médiateur sémantique pour l'intégration et l'exploitation des données hétérogènes. La troisième section démontre la faisabilité et les possibilités qu'offre cette nouvelle approche appliquée à notre cas d'étude. La dernière expose les perspectives de cette recherche.

1.1 Les bases de données de la Zone Atelier Plaine & Val de Sèvre

L'observatoire de la "Zone Atelier Plaine & Val de Sèvre" (Figure 1a) constitue notre cas d'étude. Il couvre une zone de 450 km² au sud de la ville de Niort, dans le département des Deux-Sèvres en Poitou-Charentes, France. Il s'agit essentiellement d'une plaine céréalière intensive : céréales, maïs, tournesol, pois et colza où l'élevage (bovins, caprins) est encore présent mais en forte baisse. Les parcelles agricoles sont encore de taille modeste (4-8 ha) et 15% d'entre elles sont occupées par des prairies (artificielles, permanentes ou temporaires), contre 60% en 1970. Il a pour objet de rechercher la relation entre les pratiques agricoles et les processus écologiques, à travers l'étude de l'évolution de l'organisation spatiale du paysage. L'enjeu de notre recherche est d'offrir les moyens d'une analyse spatio-temporelle fine des données collectées. Depuis plus de vingt ans, plusieurs bases de données ont été recueillies par l'équipe AGRIPOP (CNRS Chizé) (Figure 1b). Ces données peuvent être catégorisées en deux groupes : la base d'assolement et les bases de biodiversité (plus de détails peuvent être trouvés sur le site de l'atelier²).

La base *Assolement*

L'organisation spatiale du paysage évolue dans le temps parce que les agriculteurs modifient l'assolement de leurs parcelles chaque année, mais également recomposent parfois les parcelles entre elles (par des scissions, fusions ou des redistributions), changeant ainsi les formes des parcelles.

Depuis 1994, les occupations du sol sont donc relevées annuellement sur le terrain et numérisées sur les 19 000 parcelles agricoles. Ces données sont centralisées dans une base de données nommée "Assolement".

Une parcelle agricole est définie comme une unité de gestion, un polygone entouré par des entités ayant différentes cultures dans les années successives (généralement quatre). Une parcelle est délimitée par des limites physiques telles qu'une route, une rivière, un chemin de champ ou une limite d'un seul champ. Elle ne contient qu'un seul type de culture. Elle diffère de la parcelle cadastrale, mais également des blocs stockés dans le RPG (Registre Parcellaire Graphique) qui sont mis à jour tous les deux ans et distribués par IGN, disponibles ici³. Le RPG contient des informations pertinentes pour le suivi des propriétaires des parcelles, mais les limites des blocs ne correspondent pas aux parcelles

2. <http://za-geminat.cnrs.fr/>

3. <http://www.geoportail.gouv.fr/donnee/251/registre-parcellaire-graphique-rpg-2012>

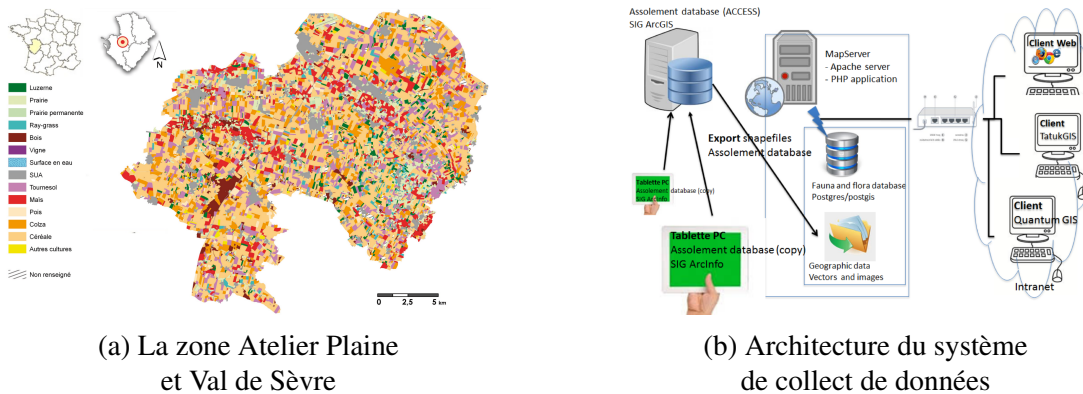


FIGURE 1: Observatoire de la Zone Atelier Plaine & Val de Sèvre

observées sur le terrain. En fait, les agriculteurs souvent divisent leur bloc en plusieurs parcelles, mais seulement la surface de chaque bloc est déclarée avec sa nature de la culture en pourcentage.

Le modèle de données de la base se fonde sur le paradigme Space-Time composite proposé par (Langran & Chrisman, 1998). L'idée consiste à ne pas stocker la géométrie de chaque parcelle pour chaque année, mais à utiliser dans le modèle de petites géométries, appelées ici *microparcelles*, obtenues par l'intersection de toutes les parcelles au cours de la période d'observation. La géométrie de toutes les parcelles peut être reconstruite "à la volée" pour chaque année en utilisant une composition des microparcelles constituant la parcelle.

L'observation sur le long terme nécessite aussi un système opérationnel avec des interfaces conviviales pour les utilisateurs, similaire aux solutions proposées par les Systèmes d'Information Géographique (SIG). Les utilisateurs doivent enregistrer chaque année le nouveau type d'utilisation du sol de chaque parcelle et éventuellement des changements de forme pour chaque parcelle.

Les bases *Biodiversité*

- Parallèlement, des données de faune et flore sont collectées sur le terrain depuis plusieurs années par l'équipe AGRIPPOP de Chizé et centralisées dans une autre base de données qui est structurée suivant un schéma relationnel spatial, implémenté dans PostgreSQL avec PostGIS. Ces données, ponctuelles et datées, proviennent des différents chercheurs qui rapportent leurs observations concernant 600 espèces, principalement des oiseaux, et certaines plantes, grâce à une interface Web. Pour les oiseaux, la base constitue une collection d'observations décrivant le comportement des espèces observées ainsi que leurs nids, et leur contexte (hauteur de végétation, date, heure, localisation, etc.).
- Il existe également les données des micromammifères. Il s'agit d'un ensemble de 10.000 observations de trois espèces de micro-mammifères qui sont de bons indicateurs de pratiques agricoles : Campagnol des champs, Mulot sylvestre et Musaraigne musette. Dans chaque observation, les taux de capture et les caractéristiques individuelles sont enregistrés.
- Il existe par ailleurs d'autres données structurées sur différentes espèces, souvent

dans des tableurs, ou bien des bases de données Access. Il serait souhaitable de pouvoir interroger et croiser aussi ces sources avec la connaissance de l'assolement et de la faune avicole. C'est le cas des données relatives à l'observation des carabes, petits coléoptères auxiliaires des champs et très sensibles à la qualité des milieux, qui bénéficient d'un suivi depuis 9 ans dans une base Access.

1.2 Expression des besoins autour de l'exploitation des données hétérogènes

Avec les données déjà disponibles, un nombre conséquent d'analyses peut-être mené :

1. L'analyse peut être utilisée en premier lieu pour vérifier l'ensemble de données collecté. Lors de la rotation des cultures, les experts peuvent décrire un certain nombre de règles de succession afin d'éliminer ou de corriger les valeurs douteuses. Par exemple, la succession des cultures peu probable comme "Tournesol- Tournesol" ou "Tournesol-Colza", ainsi que la disparition de bois dans la zone de l'atelier peut être détectée et examinée. Principalement, ce type d'analyse a besoin du raisonnement sur les relations temporelles entre les durées relevées de l'assolement.
2. D'autre part, les événements territoriaux, tels que la fusion, l'intégration, la scission, l'extraction, la réaffectation et de rectification (Plumejeaud *et al.*, 2011) sont souhaités être détectés. L'analyse de ces événements permet de découvrir la corrélation entre la prise de l'utilisation des terres et la fragmentation des terres ou de l'agrégation dans la pratique agricole. Ces événements peuvent être détectés à travers des requêtes avec le raisonnement spatio-temporel.
3. Enfin, les experts veulent aussi rechercher la corrélation entre les observations des espèces et les utilisations des terres des parcelles. Ils pourraient concerner les préférences des animaux par le type et la forme de la rotation des cultures. Ils peuvent également vérifier l'apparition d'espèces en fonction de la chaîne alimentaire et de la saison. Les requêtes croisées de différentes bases de données avec le raisonnement sur les relations spatio-temporelles sont nécessaires pour sélectionner les observations qui se reproduisent à l'intérieur des durées relevées de l'assolement.

2 Une ontologie spatio-temporelle pour l'exploitation des données environnementales

Dans le but d'intégrer et d'exploiter l'ensemble de données présentées, nous introduisons une approche ontologie spatio-temporelle globale qui agit en tant que médiateur sémantique. L'ontologie est formée sur la base de l'ontologie du temps et de l'espace. Un mécanisme de raisonnement sur les relations spatio-temporelles entre entités est également proposé.

2.1 Ontologie de temps et représentation du temps dans le web sémantique

L'ontologie OWL-Time⁴ (Hobbs & Pan, 2004) développée au sein du consortium W3C se consacre aux concepts et relations temporelles comme définis dans la théorie d'Allen, et bénéficie d'une spécification précise formalisée en OWL, et est donc certainement appropriée. Cette

4. <http://www.w3.org/2006/time>

ontologie de domaine est tout d'abord destinée à décrire le contenu temporel des pages Web et les propriétés temporelles des services Web. Étant recommandée par le W3C pour la modélisation des concepts temporels, cette ontologie fournit un vocabulaire pour exprimer des faits sur les relations topologiques entre les instants et les intervalles.

Cependant, une ontologie de temps n'est pas suffisante pour représenter l'évolution d'un objet. Ces ontologies sont synchroniques, c'est-à-dire qu'elles se réfèrent à un seul point dans le temps. Afin d'assurer le suivi de l'évolution spatiale et sémantique (relation diachronique) d'un objet, nous avons besoin d'incorporer la dimension temporelle dans l'ontologie. En effet, les philosophes ont établi une distinction entre deux paradigmes : l'endurantisme et le perdurantisme pour représenter les identités diachroniques. L'endurantisme suppose que les objets ont trois dimensions spatiales et existent en totalité à chaque moment de leur vie. Ainsi, ces objets n'ont normalement pas de dimension temporelle. En revanche, l'approche perdurantiste considère que les objets ont quatre dimensions (spatiales et temporelles). Ces objets ont des *tranches de temps* dans leur vie qui composent la dimension temporelle. Cette approche représente donc les différentes propriétés d'une entité dans le temps comme les *fluents* qui ne sont validés que pendant certains intervalles ou à des moments dans le temps. Selon les études de comparaison de (Al-Debei *et al.*, 2012), l'approche perdurantiste permet des représentations plus riches de phénomènes du monde réel grâce à sa flexibilité et son expressivité.

Les deux langages principaux du Web Sémantique, OWL et RDFS, ne permettent que des relations binaires entre les individus, sans aucune considération de la relation temporelle entre eux. Plusieurs méthodes ont été proposées afin de surmonter cette limitation, la plus connue est le modèle perdurantiste 4D-Fluents (Welty & Fikes, 2006). Le modèle considère que l'existence d'une entité peut être exprimée avec plusieurs représentations, chacune correspond à un intervalle de temps défini, appelée un *tranche de temps* (*time slice*). Lorsque la propriété d'un objet change, une nouvelle tranche de temps est établie, tenant la nouvelle propriété de l'objet. De cette manière, les changements se produisent sur les propriétés de la partie temporelle de l'ontologie en gardant les entités de la partie statique inchangée. Cette méthode introduit la classe *TimeSlice* qui représente les parties temporelles de l'entité tandis que *TimeInterval* constitue une classe du domaine temporel. L'entité est liée à une instance de la classe *TimeSlice* par la propriété *tsTimeSliceOf* et cette instance est connectée avec une instance de la classe *TimeInterval* par la propriété *tsTimeInterval*.

Plusieurs approches basées sur le 4D-Fluents ont été introduites pour représenter la dimension temporelle. tOWL (Frasincar *et al.*, 2010) étend le langage OWL avec une dimension temporelle permettant la représentation du temps, des changements et des transitions. SOWL (Batsakis & Petrakis, 2011; Batsakis & Antoniou, 2014) par exemple étend l'ontologie OWL-Time en considérant les relations qualitatives entre les intervalles. Récemment, le modèle *Continuum*, présenté par (Harbelot *et al.*, 2013, 2014) permet l'inférence de l'information qualitative à partir d'informations quantitatives en reliant les diverses représentations dynamiques d'une entité. Le modèle représente des entités dynamiques comme constituées par des tranches de temps avec sémantique, spatiale, composantes temporelles et d'identité. Donc, il est capable de relier les diverses représentations d'une entité et permet l'inférence de l'information qualitative. Les résultats d'inférence sont ensuite ajoutés à l'ontologie afin d'améliorer les connaissances sur le phénomène. Le modèle a été appliqué avec succès dans les études sur l'évolution urbaine (Harbelot *et al.*, 2013) ou d'un processus de décolonisation (Harbelot *et al.*, 2014).

Possédant la capacité de présentation du monde réel des objets dynamiques, le modèle de

4D-fluent avec l'ontologie OWL-Time sont choisis pour le développement de notre ontologie spatio-temporelle qui sera présentée plus tard.

2.2 Ontologie de l'espace

Les entités spatiales sont représentées par des points, des lignes (polygonale, lignes) qui renferment les objets ou les régions et leurs relations. De nombreuses ontologies spatiales ont été introduites pour différentes applications. Selon l'étude de (Ressler *et al.*, 2010), parmi 45 ontologies géospatiales et temporelles, sept ontologies spatio-temporelles sont recommandés pour la réutilisation. Les deux versions de GeoRSS⁵ : GeoRSS Simple et GeoRSS GML sont les plus utilisées. GeoRSS a été conçu comme un standard destiné à inclure les coordonnées géographiques dans un flux RSS pour les applications géographiques comme la cartographie interactive. Dans le standard GeoRSS, les contenus sont des points géographiques, des lignes, des zones d'intérêt et leurs descriptions. Ces emplacements peuvent être codés soit dans une chaîne littérale de latitude et de longitude, appelé GeoRSS Simple, ou avec une représentation plus robuste en utilisant GML, appelé GeoRSS GML, qui est formellement défini comme un profile d'application de GML et prend en charge une plus grande gamme de fonctionnalités, notamment les systèmes de référence autres que la latitude et la longitude du WGS84. Le modèle GeoRSS simple est réutilisé dans le développement de notre ontologie en raison de sa simplicité, ses trois concepts principaux : le point, la ligne et le polygone sont suffisants pour représenter la dimension spatiale de nos entités spatio-temporelles.

2.3 Raisonnement spatio-temporel

Les objets perdurants évoluent dans le temps. En conséquence, les objets spatio-temporels peuvent changer leur position ou leurs occupations au cours de leur vie. Le raisonnement spatio-temporel est utilisé pour détecter les relations spatio-temporelles entre ces entités. Ce dernier est réalisé par la combinaison dynamique du mécanisme de raisonnement spatial et temporel qui sont présentés ci-après.

Les relations qualitatives dans le domaine temporel sont basées sur des relations binaires et mutuellement exclusives. Les travaux d'Allen (Allen, 1983) fondent une algèbre temporelle permettant de définir les relations topologiques entre objets datés. Pour deux intervalles temporels définis par leurs dates de début et fin, il existe les 13 relations suivantes : *before*, *meets*, *overlaps*, *during*, *starts*, *finishes* et leur réciproque, respectivement *after*, *met-by*, *overlapped-by*, *contains*, *started by*, *finished-by*, et *equals*. Ces intervalles peuvent être considérés comme des instances de la classe *ProperInterval* de OWL-Time. Ils sont liés à deux instances de la classe *Instant* par l'attribut *hasBeginning* et *hasEnd* qui déterminent leurs dates de début et fin.

Les 13 relations d'Allen permettent de répondre à des questions sur la proximité temporelle de deux phénomènes, à condition d'employer pour les intervalles la même granularité. En sus, nous devons donc exprimer et modéliser la relation "à l'intérieur de" entre un instant et un intervalle qui est indispensable pour le croisement de nos deux bases de données. Par ailleurs, le langage OWL lui-même n'a pas d'opérateurs temporels pour manipuler des valeurs de temps. Le langage Semantic Web Rule Language (SWRL) est une solution pour ajouter des règles

5. <http://www.georss.org/>

d'inférence générales. Il fournit des prédicats facilitant surtout la comparaison des valeurs temporelles. Ce dernier peut être utilisé par un moteur d'inférence, comme par exemple Pellet⁶ pour déduire des relations temporelles entre les entités. Dans notre cas, le même résultat est obtenu en représentant ces règles par des requêtes SPARQL CONSTRUCT ou SPARQL UPDATE⁷. L'utilisation du SPARQL comme langage de règle est proposée dans SPIN⁸ (Knublauch *et al.*, 2011) ou Corese/KGRAM (Corby *et al.*, 2004). Par exemple, la déduction de la relation à l'intérieur de entre un instant et un intervalle de temps peut se représenter par la requête (Code 1) ou par une règle SWRL comme le Code 2.

La dimension spatiale des objets dans les bases de données environnementales en général, et dans nos bases de données en particulier est représentée par des points et des polygones qui sont définis par un ensemble des points. Afin de découvrir leurs relations spatiales, les relations qualitatives doivent être déduites de ces informations quantitatives. Dans la littérature, l'analyse topologique entre les objets spatiaux est souvent réalisée par le modèle 9IM (Egenhofer & Herring, 1991) ou le Modèle RCC8 (Randell *et al.*, 1992). Dans les deux cas, on obtient un ensemble équivalent de huit relations topologiques disjointes qui sont mutuellement exhaustives : *equals*, *disjoint*, *intersects*, *touches*, *within*, *contains* et *overlaps*. Malheureusement, ces relations ne peuvent être déduites par les règles SWRL simples. Plusieurs études (Karmacharya *et al.*, 2010; Vandecasteele *et al.*, 2012) ont introduit les SWRL built-ins pour le traitement spatial et la représentation des relations spatiales, mais il existe encore des limitations en ce qui concerne principalement la performance et la capacité de réutilisation.

Par conséquent, le raisonnement sur l'information spatiale complexe est réalisé par un triplestore spatial. Ainsi, le raisonnement spatio-temporel est effectué en combinant les relations temporelles déjà déduites et insérées à la base de connaissances et les fonctions spatiales du triplestore.

Code 1: Inférence de la relation à l'intérieur de entre un instant et un intervalle de temps par requête SPARQL Update

```
INSERT {?x time:inside ?a.}
WHERE
{
  ?x rdf:type time:Instant.
  ?x time:inXSDDateTime ?dt.
  ?a rdf:type time:Interval.
  ?a time:hasBeginning ?be.
  ?a time:hasEnd ?end.
  ?be time:inXSDDateTime ?dt1.
  ?end time:inXSDDateTime ?dt2.
  FILTER(?dt >=?dt1 && ?dt <=?dt2)
}
```

Code 2: Inférence de la relation à l'intérieur de entre un instant et un intervalle de temps par règle SWRL

```
Instant(?x), ProperInterval(?a),
  hasBeginning(?a,?b), hasEnd(?a,?c),
  inXSDDateTime(?b,?d),
  inXSDDateTime(?c,?e),
  inXSDDateTime(?x,?y),
  lessThanOrEqual(?y,?e),
  greaterThanOrEqual(?y,?d) ->
  inside(?x,?a)
```

6. <http://clarkparsia.com/pellet/>

7. <https://www.w3.org/TR/sparql11-update/>

8. <http://spinrdf.org/>

2.4 Une ontologie spatio-temporelle pour l'environnement

Nous proposons une ontologie basée sur l'approche 4 D-fluents qui joue le rôle d'un médiateur sémantique pour l'intégration et l'exploitation des données environnementales présentées ci-dessus.

Objet spatio-temporel

Les objets spatio-temporels (*sige :STObj*) principaux dans notre recherche sont les parcelles, les nids et les individus (appelés ici les individus spatio-temporels) appartenant aux différents types d'insectes d'oiseaux ou de micromammifères. Ils peuvent être connus avec un numéro suivi ou inconnus.

Élément spatio-temporel

Les objets spatio-temporels ont un ou plusieurs éléments spatio-temporels (*sige :STElement*) qui correspondent à leurs différentes caractéristiques et occupations spatiales à travers leur vie. La classe *sige :STElement* a deux sous-classes : *sige :Obsv* pour les observations des individus et *sige :TimeSlice* pour les relevés des parcelles. Chaque élément spatio-temporel a trois composants : sémantique (*sig :Description*), temporel (*time :TemporalEntity*) et spatial (*georss :_geometry*). De cette façon, la rotation de cultures de chaque parcelle, les événements territoriaux ou les différentes observations d'une espèce donnée peuvent être représentés et analysés.

Composant temporel

Tandis que les informations des parcelles (leur culture, leur géométrie) sont relevées par des intervalles de validité temporelles, les observations d'espèces sont enregistrées de manière aléatoire par des instants. Notre solution étend alors le modèle 4D-Fluents en généralisant la classe *Interval* à la classe *TemporalEntity* de l'ontologie OWL-Time qui a deux sous-classes *Interval* et *Instant*.

Composant spatial

Dans ce composant, nous proposons deux classes *sig :MPGeom* et *sig :ParcelGeom* qui correspondent respectivement à des micro-parcelles et parcelles comme une sous-classe de la classe *georss :Polygon*, spécialisant *georss :_geometry*. Par conséquent, nous pouvons conserver la structure initiale de nos bases de données construites sur le paradigme *Space-Time composite* et améliorer la performance du système grâce à la géométrie pré-calculée des parcelles.

Composant sémantique

Ce composant vise à décrire les observations des individus ou les relevés des parcelles. Il peut être le contexte de ces observations ou la culture appliquée à ces parcelles.

2.5 Vers un système sémantique pour l'intégration et exploitation des données environnementales

Puisque l'application de règles d'inférence sur des sources de données distribués peut s'avérer extrêmement coûteuse (Seye *et al.*, 2014) et qu'il n'existe pas un bon mécanisme pour raisonner sur les relations spatiales qualitatives, la centralisation de différents sources de données

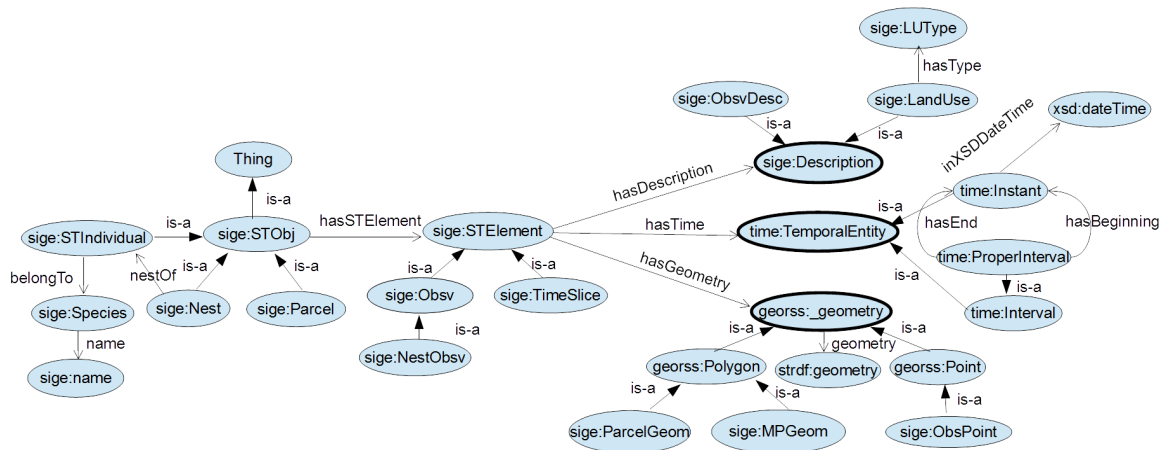


FIGURE 2: Une ontologie spatio-temporelle pour l'environnement.

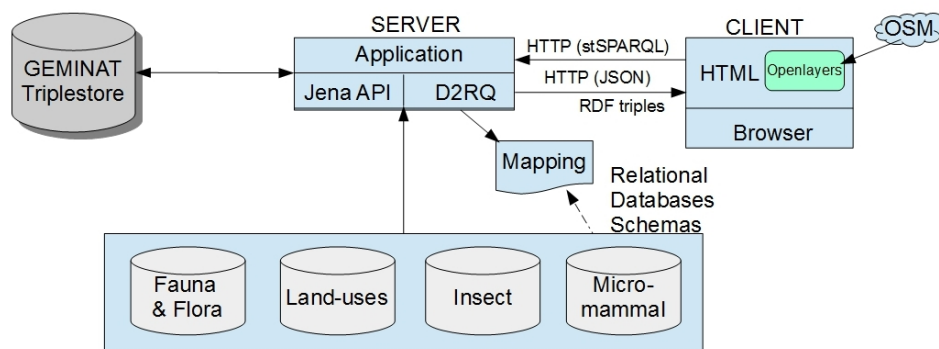


FIGURE 3: Un framework pour l'intégration et l'exploitation des données environnementales

à un triplestore spatial est appliquée. Un framework (Figure 3) est développé dans lequel un serveur web est hébergé pour recevoir les requêtes stSPARQL de l'utilisateur. Le framework se compose de quatre parties : la translation de données, le chargement du triplestore, l'enrichissement spatio-temporel et l'analyse sémantique.

1. **Translation de données** : Afin de peupler l'ontologie avec des sources de données existantes, nous nous appuyons sur la technique de translation qui définit une correspondance entre ces sources de données et l'ontologie. L'outil D2RQ⁹ (Bizer, 2004) est choisi en raison de son support de différents SGBD. Celui-ci transforme les données relationnelles en graphe RDF virtuelle à travers un fichier de mapping qui décrit comment se connecter à ces bases de données et mettre en correspondance l'ontologie au schéma de ces sources

9. <http://d2rq.org/>

de données. Ce graphe RDF est alors géré par le framework *Jena*¹⁰.

2. **Chargement du triplestore** : Les données transformées en RDF sont insérées dans la base de connaissance *GEMINAT* qui est gérée par le triplestore Strabon. Le chargement se réalise grâce à l'interface web, ou aux requêtes SPARQL ou à la fonction du triplestore.
3. **Enrichissement spatio-temporel** : Les relations spatio-temporelles sont inférées par des requêtes SPARQL 1.1 en utilisant des fonctions spatiales incorporées dans le triplestore. Ces nouvelles relations sont insérées dans la base de connaissance pour l'enrichissement.
4. **Analyse sémantique** : Le résultat des requêtes SPARQL retourné par Strabon est traité et ensuite préparé par Jena pour la visualisation. Ce dernier est visualisé à travers la bibliothèque *OpenLayers*¹¹ avec la base de données géographiques d'*OpenStreetMap*¹². Les résultats sont stockés dans plusieurs couches différentes afin de faciliter la présentation et l'analyse.

3 Exploitation des données environnementales hétérogènes

Le framework proposé ainsi que l'utilisation d'une ontologie spatio-temporelle en tant que médiateur sémantique peut remplir les trois principaux besoins de l'analyse spatio-temporelle. En effet, le modèle de données sous forme de graphe RDF sous-jacent facilite l'intégration des différentes bases de données. En outre, grâce au triplestore Strabon, les relations spatio-temporelles entre les objets peuvent être déduites. De nouvelles déclarations peuvent se déduire de la base de connaissances à travers des règles spatio-temporelles et des règles métiers représentées sous forme de requêtes SPARQL Update.

- Pour analyser les corrélations entre la rotation des cultures et la biodiversité, les experts peuvent visualiser la corrélation des espèces par type et la forme de la rotation des cultures. Par exemple, ils peuvent consulter la corrélation entre les positions des Busards cendrés (*Circus pygargus*) et la nature des cultures des parcelles pour une année donnée (Figure 4a). Dans cette analyse, la relation *inside* entre l'instant de l'observation et la durée de déclaration de l'assolement ; et la relation *within* entre le point d'observation et le polygone de la parcelle enregistrée sont utilisées (Code 3).
- Grâce aux relations temporelles qualitatives inférées, les chercheurs peuvent également vérifier la qualité de leurs données enregistrées. En effet, les règles de domaine ou des connaissances d'experts sur la rotation des cultures, l'apparence ou la disparition de certaines plantes cultivées peuvent être représentées par des requêtes SPARQL pour détecter des anomalies dans les données recueillies. Dans cette étude, la relation temporelle *intervallMeets* entre deux intervalles de relevée de l'assolement de la même parcelle est utilisée.

10. <http://jena.apache.org/>

11. <http://openlayers.org/>

12. www.openstreetmap.org

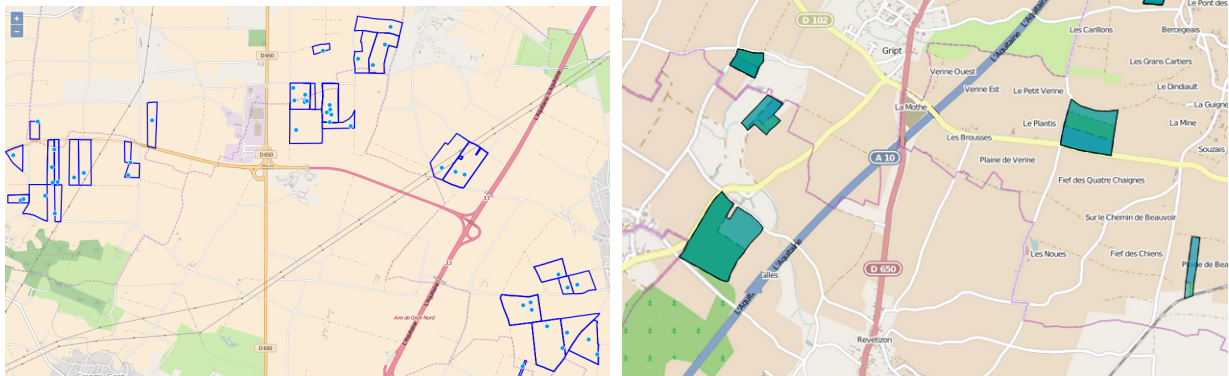
- Les événements territoriaux appliqués sur les parcelles agricoles peuvent être découverts en combinant les relations spatio-temporelles qualitatives. Ils se détectent en incorporant la relation *intervallMeets* entre deux intervalles de temps et la relation *within* entre la géométrie des parcelles de chaque relevée de l'assolement. Par exemple, des événements d'intégration, dans laquelle une parcelle a été absorbée par une autre, en 2009, peuvent être récupérées et affichées dans la carte comme (Figure 4b) grâce à la requête (Code 4). Une analyse de ces événements ainsi que l'utilisation du sol des parcelles concernées peuvent révéler les préférences de la pratique des agriculteurs. Les données sur les sols et les données de journalisation des prix du marché pourraient être utilisés dans les analyses.

4 Conclusion et perspectives

Les travaux exposés s'inscrivent dans le projet interdisciplinaire "GéoConnaissances des milieux naturels" visant à améliorer l'exploitation des informations collectées depuis 1994 par l'observatoire de la "Zone Atelier Plaine & Val de Sèvre". Ce projet reçoit des aides financières de la Fédération de Recherche en Environnement pour le Développement Durable (FREDD¹³) du PRES Limousin Poitou Charentes. Nous cherchons à proposer une plateforme ouverte et libre pour analyser et exploiter des données environnementales hétérogènes. Nous proposons une méthode permettant l'interrogation de ces bases à travers une modélisation intégrant les composantes spatiales, temporelles et thématique des données. Appliquée à notre cas d'étude, cette approche facilite la mise en relation des cultures et des observations, qui répond aux besoins d'analyse et d'exploitation des experts. L'approche introduite pourrait être réutilisée pour effectuer la gestion et l'analyse des données environnementales à long terme pour d'autres observatoires.

Dans nos perspectives, d'un part, nous considérons à intégrer d'autres ensembles de données de la zone d'atelier, tels que les insectes et les données botaniques, ou les données satellites. Il sera alors possible d'utiliser le système pour enrichir et qualifier nos sources de données. D'autre part, nous cherchons à montrer la faisabilité et la réutilisabilité de l'approche proposée en l'appliquant aux autres données spatio-temporelles hétérogènes. Nous prévoyons également de publier une partie de ces données sur le Web sous forme des données liées dans le but de faciliter les échanges avec d'autres ensembles de données disponibles, notamment météorologiques. Enfin, l'application de l'extraction de règles d'association sur la base de connaissances est également envisagée pour découvrir de nouvelles règles d'association et d'examiner celles qui sont déjà connues (par exemple : la relation entre les espèces et les cultures des parcelles et entre les espèces elles-mêmes dans leur chaîne alimentaire).

13. <http://fredd.cue-lpc.fr/>



(a) Corrélations entre les positions des Busards cendrés et la nature des cultures des parcelles en 2009. (b) Événement d'intégration des parcelles en 2009

FIGURE 4: Analyses spatio-temporelles

Code 3: Corrélation entre la nidification des Busards cendrés et la nature des cultures

```
SELECT *
WHERE
{
?nest sige:nestOf ?indv.
?indv sige:belongsTo ?species.
?species sige:name ?name.
?nest sige:hasNestObsv ?nobsv.
?nobsv sig:hasTime ?inst.
?inst time:inside ?intv.
?ts sige:hasTime ?intv.
?ts sige:hasLandUse ?lu.
?nobsv sige:hasGeometry ?geon.
?geon sige:geometry ?geomn.
?ts sige:hasGeometry ?geo.
?geo sige:geometry ?geom.
FILTER (strdf:within(?geomn,?geom)
&& ?name="Busard cendre")
}
```

Code 4: Détection des événements d'intégration des parcelles

```
SELECT *
WHERE
{
?tsa sige:hasTimeSlice ?pc.
?tsb sige:hasTimeSlice ?pc2.
?tsc sige:hasTimeSlice ?pc.
FILTER(?pc!=?pc2)
?tsa sige:hasTime ?intva.
?tsb sige:hasTime ?intvb.
?tsc sige:hasTime ?intvc.
?intva time:intervalMeets ?intvc.
?intvb time:intervalMeets ?intvc.
?tsa sige:hasPGeometry ?geoa.
?geoa sige:geometry ?geoma.
?tsb sige:hasPGeometry ?geob.
?geob sige:geometry ?geomb.
FILTER(strdf:intersects(?geoma,?
geomb))
?tsc sige:hasPGeometry ?geoc.
?geoc sige:geometry ?geomc.
FILTER(strdf:within(?geoma,?geomc)
&& strdf:within(?geomb,?geomc))
}
```

Références

AL-DEBEI M. M., AL ASSWAD M. M., DE CESARE S. & LYCETT M. (2012). Conceptual modelling and the quality of ontologies : Endurantism vs. perdurantism. *CoRR*.
 ALLEN J. F. (1983). Maintaining knowledge about temporal intervals. *Commun. ACM*, **26**, 11.

- BATSAKIS S. & ANTONIOU G. (2014). Representing and reasoning over spatial relations in owl : A rule-based approach. *Lightning Talks, W*, 3.
- BATSAKIS S. & PETRAKIS E. G. M. (2011). Sowl : A framework for handling spatio-temporal information in owl 2.0. In *Proceedings of the 5th International Conference on Rule-based Reasoning, and Applications, RuleML'2011 : Programming*.
- BIZER C. (2004). D2rq - treating non-rdf databases as virtual rdf graphs. In *In Proceedings of the 3rd International Semantic Web Conference (ISWC2004)*.
- CORBY O., DIENG-KUNTZ R. & FARON-ZUCKER C. (2004). Querying the semantic web with corese search engine. In R. L. DE MÁNTARAS & L. SAIITA, Eds., *ECAI*, p. 705–709 : IOS Press.
- EGENHOFER M. & HERRING J. (1991). *Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases*. Department of Surveying Engineering, University of Maine.
- FRASINCAR F., MILEA V. & KAYMAK U. (2010). towl : Integrating time in owl. In R. DE VIRGILIO, F. GIUNCHIGLIA & L. TANCA, Eds., *Semantic Web Information Management*, p. 225–246. Springer Berlin Heidelberg.
- HARBELOT B., ARENAS H. & C. C. (2013). Continuum : A spatio-temporal data model to represent and qualify filiation relationships. In 85. *ACM : Proceedings of the 4th acm sigspatial international workshop on geostreaming*.
- HARBELOT B., ARENAS H. & CRUZ C. (2014). Un modèle sémantique spatio-temporel pour capturer la dynamique des environnements". 14^{ème} conférence extraction et gestion des connaissances, rennes, france.
- HOBBS J. R. & PAN F. (2004). An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing*, 3.
- KARMACHARYA A., CRUZ C., BOOCHS F. & MARZANI F. (2010). Use of geospatial analyses for semantic reasoning. In I. J. SETCHI, R. HOWLETT & L. JAIN, Eds., *R*, p. 576–586. Vol. 6276, p.. Springer Berlin Heidelberg : Knowledge-based and intelligent information and engineering systems.
- KNUBLAUCH H., JAMES A. H. & KINGSLEY I. (2011). Spin - overview and motivation.
- LANGRAN G. E. & CHRISMAN N. R. A. (1998). framework for temporal geographic information. *Cartographica : The International Journal for Geographic Information and Geovisualization*, 25(3), 1–14.
- 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*, 25(10), 1597–1612.
- RANDELL D. A., CUI Z. & COHN A. G. (1992). A spatial logic based on regions and connection. In *Proceedings 3rd International Conference On Knowledge Representation And Reasoning*.
- RESSLER J., DEAN M. & KOLAS D. (2010). Geospatial ontology trade study. In *Proceedings of the 2010 Conference on Ontologies and Semantic Technologies for Intelligence*, p. 179–211, Amsterdam, The Netherlands, The Netherlands : IOS Press.
- SEYE O., FARON-ZUCKER C., CORBY O. & GAIGNARD A. (2014). Publication, partage et réutilisation de règles sur le Web de données. In C. FARON-ZUCKER, Ed., *IC - 25^{èmes} Journées francophones d'Ingénierie des Connaissances*, p. 237–248, Clermont-Ferrand, France.
- VANDECASTEELE A., & NAPOLI A. (2012). Spatial ontologies for detecting abnormal maritime behaviour. In S. K. YEOSU, Ed., *OCEANS 2012 MTS/IEEE Yeosu Conference : The Living Ocean and Coast - Diversity of Resources and Sustainable Activities*.
- WELTY C. & FIKES R. A. (2006). reusable ontology for fluents in owl. In *Proceedings of the conference on formal ontology in information systems*, p. 226–236 : p.. IOS Press.