



Un SIG collaboratif pour la recherche historique

Conception d'un atlas historique numérique et d'une plateforme de travail collaborative à partir de la méthode *SyMoGIH*.

Partie 1 : Naissance et conception d'un système d'information géo-historique collaboratif

C. BUTEZ, AVEC LA COLLABORATION DE F. BERETTA, S. BOSCHETTO, D. FERHOD ET P. VERNUS

LABORATOIRE DE RECHERCHE HISTORIQUE RHÔNE-ALPES - LARHRA UMR 5190

INSTITUT DES SCIENCES DE L'HOMME - ISH

14, AVENUE BERTHELOT

69 363 LYON CEDEX 07

CHARLOTTE.BUTEZ@ISH-LYON.CNRS.FR

L' introduction des *digital humanities* en sciences historiques a fait naître de nouvelles pratiques de recherche du point de vue des méthodes et des outils de travail. Avec elles, l'historien peut envisager de nouvelles problématiques et perspectives de recherche ; il peut aller plus loin dans l'exploitation de ses données et entrer dans des systèmes d'information collaboratifs et cumulatifs.

Dans ce contexte, pourquoi ne pas créer un *système collaboratif générique* et ouvert traitant l'information historique, prenant en compte, outre sa dimension temporelle, sa dimension géographique, et en faire un outil au service de la recherche en histoire ? C'est le défi que s'est lancé une petite équipe d'historiens,

informaticiens et géomaticiens du *Larhra*, qui a pour cela conçu une méthode de modélisation de l'information historique et une plateforme de travail.

En premier lieu, la construction d'un système d'information collaboratif impliquait de structurer les données issues de la recherche historique selon une méthode et une sémantique commune entre les chercheurs. À partir de ce socle commun de modélisation, les données pouvaient être intégrées dans une plateforme collaborative couvrant les fonctionnalités de saisie, de stockage, d'analyse et de publication sur le *web*. Cette réflexion, commencée en 2007, a servi de base au projet *Système Modulaire de Gestion de l'Information Historique (SyMoGIH)*.

Mise en place d'une méthode de modélisation de l'information géo-historique

Modélisation des objets géo-historiques

Le premier objectif a été de définir une méthode commune de modélisation de l'information historique. Dès le départ la dimension géographique a été prise en compte par les historiens car la contextualisation spatiale joue un rôle capital pour l'interprétation comme pour la représentation de l'information historique. L'ensemble de la méthode *SyMoGIH* repose sur un principe fondamental, qui est d'une certaine manière commun

à la pratique traditionnelle de l'historien et de l'informaticien, celui de l'atomisation de l'information. Il s'agit de décomposer l'information historique en données primaires, *primary data*, c'est-à-dire d'identifier des

le système dans une base de données relationnelle et créer une application *web* de saisie.

L'outil collaboratif imposait également de concevoir un modèle qui soit générique et indépendant

Le premier aspect est relatif à l'identification d'un lieu de façon unique. Ainsi, pour garantir son individualité, l'objet lieu a été défini selon trois composants :

- ▶ Son nom : le lieu est associé à un ou plusieurs toponymes dont un est défini comme standard ;
- ▶ Son type : le lieu a un seul type (lieu habité, géographie naturelle, surface de territoire...) ;
- ▶ Sa localisation : il est associé à une ou plusieurs localisations géographiques dont une est définie comme standard.

Ce sont ces trois éléments qui permettent de l'identifier et de le rendre unique dans le système via un identifiant alphanumérique. La série des lieux ainsi définis forme le *gazetteer*, autrement dit le dictionnaire collaboratif des lieux.

Le lieu et ses localisations

La localisation du lieu peut se faire de deux manières, ponctuelle et associée à un couple de coordonnées, ou bien surfacique et représentée par une forme vecteur géoréférencée. Dans le cas d'une représentation surfacique, considérant qu'un même « lieu » peut connaître plusieurs réalités géographiques au cours de son histoire, il était nécessaire d'introduire une nouvelle entité intermédiaire, celle de la « forme concrète ». Cette entité représente la forme précise du lieu à un moment ou au cours d'une période donnée, alors que le « lieu » représente une classe abstraite regroupant toutes les formes concrètes que ce lieu a possédées.

Le cas du territoire de l'État français peut illustrer cette manière de modéliser. C'est un lieu

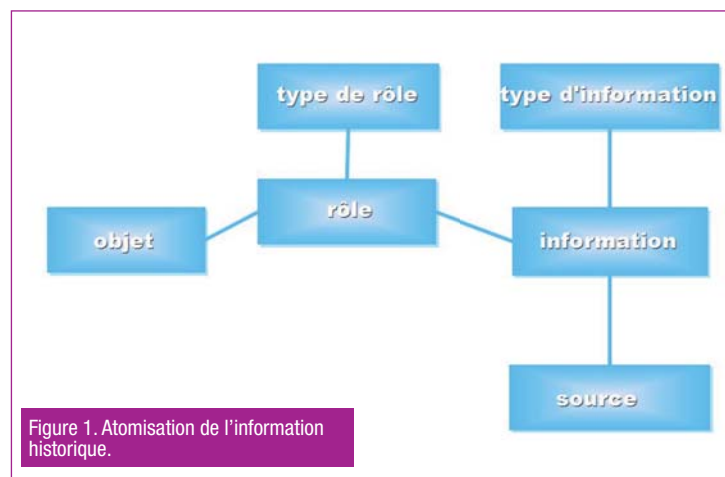


Figure 1. Atomisation de l'information historique.

unités de connaissance atomiques auxquelles on associe tous les objets qu'elles relient, tout en spécifiant quel est le rôle de chaque objet (cf. Figure 1) . Par exemple, le chercheur repère un événement tel que la naissance d'un individu, puis identifie les acteurs concernés (le nouveau-né, la mère, etc.), leur rôle, le lieu dans lequel se produit l'information... Chaque information est authentifiée, ici, par exemple, en citant le registre d'état-civil ou l'acte de naissance.

L'usage de la méthode de modélisation *Merise*, accessible aux non-informaticiens, a joué un rôle essentiel dans la conception du projet. Elle a permis, d'une part, aux historiens impliqués dans ce travail de représenter les informations historiques sous forme *entités-relations*, et, d'autre part, de garantir la communication avec les informaticiens. À partir de ce modèle conceptuel de données fourni par les chercheurs, les informaticiens ont pu implémenter

de toute problématique de recherche. De ce travail de conceptualisation est née une sémantique de référence exposée sur <http://SyMoGIH.org>.

Définition de l'entité « lieu » dans SyMoGIH

La gestion de la dimension géographique des données historiques a constitué un volet majeur dans la conception du système. L'information historique devait pouvoir être géolocalisée et associée à un lieu qui soit identifiable et pouvant connaître une évolution toponymique, typologique et spatiale dans le temps. Ainsi, dans la méthode *SyMoGIH* l'information géographique est traitée à trois niveaux :

- 1) L'identification du lieu ;
- 2) Sa géolocalisation ponctuelle ;
- 3) La ou les représentation(s) surfacique(s) que le lieu prend dans le temps.

Fig 2.1. Le territoire métropolitain de la France : plusieurs formes concrètes pour un même lieu

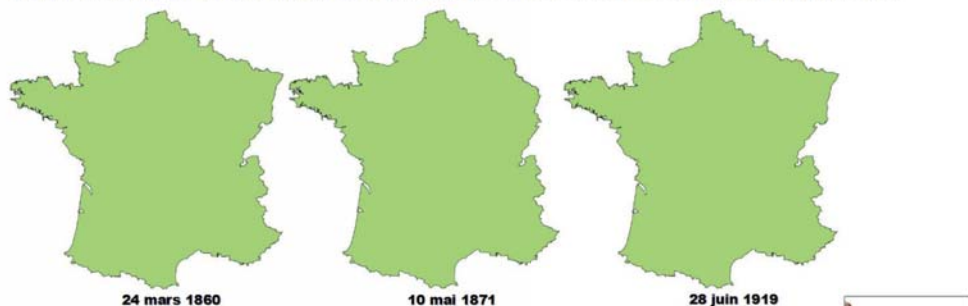


Fig 2.2. La forme concrète de la France le 10 mai 1870 : plusieurs géométries

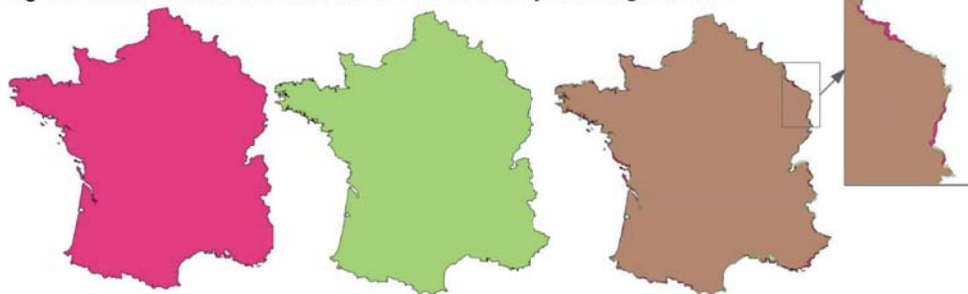


Figure 2 - Illustrations des multiples représentations d'un même lieu.

identifié, unique dans le système *SyMoGIH*, mais s'il faut le représenter de 1860 à nos jours, il possède potentiellement une multitude de représentations (cf. Figure 2).

Certaines de ces représentations constituent des réalités bien différentes, c'est-à-dire plusieurs formes concrètes que le « lieu » a incarné au cours de l'Histoire

(cf. Figure 2.1). Par exemple, après la guerre de 1870-1871 et l'annexion de l'Alsace-Lorraine par la Prusse, la France possède une forme concrète entre 1871 et 1919 différente de celle de 1860, puis encore une autre qui réintègre ce morceau de territoire à partir de 1919. Par ailleurs, chacune de ces formes concrètes peuvent être représentées par différentes géométries (cf.

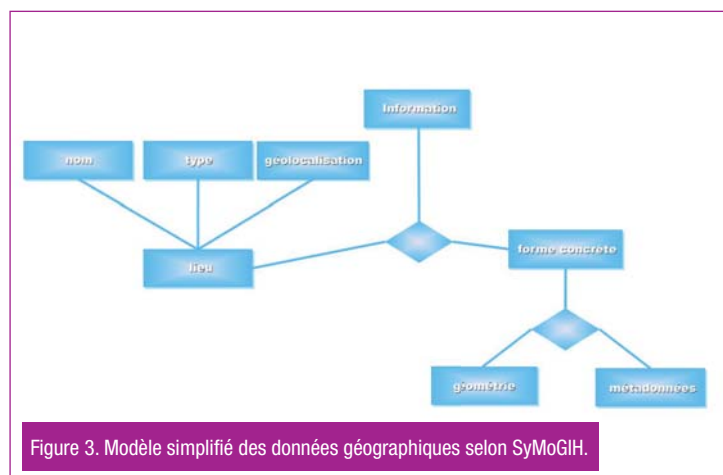


Figure 3. Modèle simplifié des données géographiques selon *SyMoGIH*.

Figure 2.2) selon les niveaux d'échelle, de résolution ou de dimensions géométriques (ligne, polygone).

Cette distinction est apparue comme essentielle dans la modélisation car les différences entre les représentations ne sont pas de même nature : d'une part, il s'agit de séparer le lieu des formes concrètes qu'il prend au cours de sa vie ; d'autre part, de rendre compte de la distinction entre une forme concrète et les différentes géométries produites pour la représenter (cf. Figure 3).

L'information historique comme lien entre lieu, représentations géographiques et géométries

Pour réaliser un système collaboratif adapté à la recherche historique, il fallait aussi prendre en compte la documentation et le « sourçage » des données. Chaque association entre un lieu et une forme concrète est susceptible d'être authentifiée grâce à une ou plusieurs informations. L'historien peut ainsi documenter sa démarche quand, face à une nouvelle réalité dans la représentation d'un lieu, il crée une forme concrète et fixe des bornes chronologiques à celle-ci. Concernant les géométries, les métadonnées liées à la production des géométries ont été intégrées au modèle de données. L'origine et les raisons du choix de la forme de toute géométrie est ainsi documentée. Une association ternaire centrale met en relation la forme concrète, les géométries et les métadonnées de stockage et de production. De cette façon, l'information historique est intégrée comme lien entre le lieu, la forme concrète et les géométries.

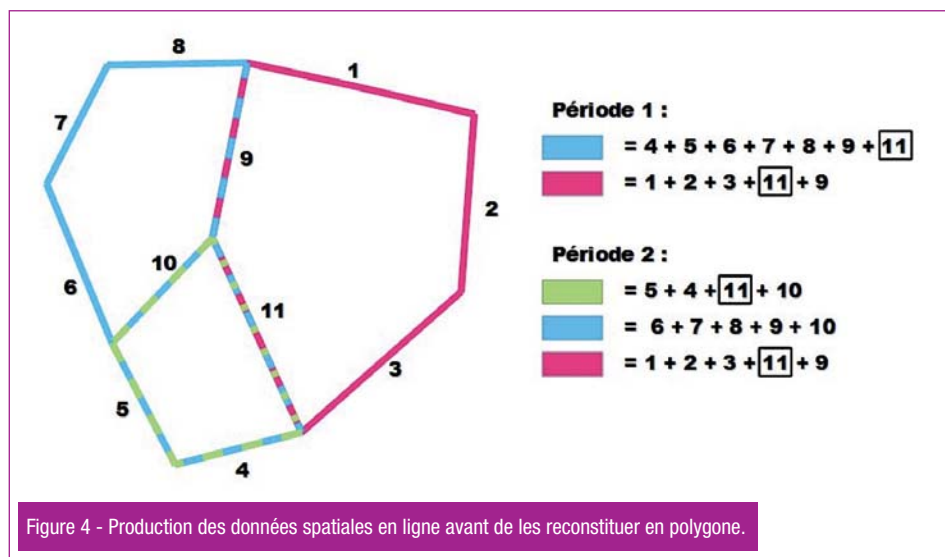
Géométries de production – Géométries de publication

Même si ce n'est pas la seule manière possible, il a paru préférable dans le projet de production d'un atlas historique numérique, prenant en compte l'évolution des territoires dans le temps, d'opter pour une vectorisation des représentations en lignes plutôt qu'en polygones. Cette méthode de production est également présentée dans les travaux de Ian Gregory. Chaque ligne produite peut être associée à une ou plusieurs formes concrètes (exemple de la ligne 11 sur la figure 4) et chacune de ces associations est bornée par des dates chronologiques. Cette méthode de production des géométries garantit le maintien d'une topologie valide. Elle permet également d'obtenir plus de souplesse et de précision lors de la production des données. Toutes les lignes peuvent être stockées dans une seule table, sans problème de performance.

L'utilisation de ce système reliant des lignes est appliqué lors de la production des données. Quand il s'agit de procéder à l'analyse spatiale des données géo-historiques, les lignes sont recomposées en polygones selon une date, par exemple dès que se produit un changement de territoire, ce qui accélère traitements et performance.

Faire de l'analyse spatiale via l'information historique

Dans la méthode *SyMoGIH*, l'information historique intervient à plusieurs niveaux. Elle ne documente pas seulement les associations d'un lieu à ses formes concrètes, elle joue également un rôle quand il s'agit de définir les relations des lieux entre eux.



Par exemple, pour traiter des relations historiques entre territoires, l'historien peut créer des types d'information tels que « rattachement de territoire », qui indique ici que tel lieu de type territoire est rattaché, à telle date, à un autre lieu de type territoire. De cette façon, il peut traiter le fait que le territoire de la Lituanie a été rattaché au territoire de l'Union européenne en 2004. Ces rattachements successifs créent une arborescence composée d'« ascendants » et de « descendants ». Dans le cas de l'Union européenne, ce seront vingt-sept informations, concernant les vingt-sept rattachements de territoires à l'Union européenne, qui constitueront l'Union européenne de 1958 à nos jours. Les territoires des pays auront au minimum une forme concrète, elle-même ayant au moins une géométrie stockée (en ligne ou polygone). Cette façon de gérer les rattachements présente un double avantage : d'une part, si une requête porte sur les données portant sur l'UE à telle date, le résultat sera constitué de tous les territoires dont elle hérite via les informations de rattachement ; d'autre part, si un lieu possédant son arborescence, tel un archipel, bascule d'un territoire à un autre, il suffira de créer une seule information pour que tous ses

« descendants » suivent, dans ce cas les îles qui le composent.

Avec cette méthode, l'historien entame une démarche d'analyse spatiale à travers la base de données historique en amont de la vectorisation ou de la visualisation : il commence par traiter l'évolution historique des territoires en créant une série d'informations. Les géométries n'apparaissent qu'en dernier lieu, souvent sous forme d'un travail d'approximation étant donné que la frontière n'est parfois pas connue rigoureusement : son tracé est déterminé par le chercheur qui choisit, localité après localité, de la faire passer de tel ou tel côté. Cette méthode amène à croiser la pratique de l'atlas historique traditionnel à celle des bases de données historiques et des méthodes SIG.

Reconstituer les évolutions géo-historiques et les visualiser de façon dynamique

À partir de cette procédure de modélisation et de production des données, il est possible de reconstituer la réalité des territoires selon une date allant jusqu'à la précision *du mois et du jour*. La requête de reconstitution va faire appel à plusieurs éléments portant sur le lieu « parent » :



- La requête va d'abord reconnaître de manière récursive tous les « descendants » dont le « parent » hérite à une date donnée ;
- Puis, pour chacun de ces « enfants », identifier la forme concrète standard qui le représente à la même date ;
- Enfin, pour chaque forme concrète sélectionnée à cette date, récupérer les données géométriques en lignes qui la matérialisent et reconstituer le polygone.

Ces requêtes sont contenues dans des fonctions (SQL-PL/pg-SQL) qui facilitent l'utilisation du système par l'historien qui n'a à renseigner que l'identifiant du lieu « parent » qu'il souhaite visualiser.

Le cas de l'atlas historique numérique qui illustre la mise en oeuvre de ce système est publié en

partie sur le site <http://geoLarhra.org>. L'objectif de ce projet est de mettre à disposition des historiens un atlas contenant les frontières des territoires de l'Europe de 1815 à nos jours. L'idée est aussi de pouvoir visualiser de façon dynamique l'évolution territoriale quelle que soit la date et de prendre en compte simultanément les évolutions de chaque espace, même les plus restreints.

Une plate-forme collaborative et ouverte, le choix d'une solution libre

Les choix technologiques de ce projet devaient répondre de fortes contraintes liées à l'aspect collaboratif et générique du système et aux usages multiples auxquels il était destiné.

Une architecture logicielle articulée autour du triptyque PostgreSQL/PostGIS/QGIS

Concernant le SGBD, choisir PostgreSQL présentait plusieurs avantages, celui d'être libre et gratuit, de mettre en place une gestion fine des droits des utilisateurs, de disposer de l'extension PostGIS et d'être un composant facilement interfaçable avec d'autres outils de gestion de base de données pour interroger les données (PGAdminIII, SQL Workbench) ou les analyser et visualiser (logiciel SIG, d'analyse statistique, serveur web SIG).

La principale partie du développement informatique a consisté à créer une interface web de saisie pour alimenter la base de données : après avoir modélisé ses informations, le chercheur les entre dans la base de données

via cette interface tout en indiquant les sources archivistiques ou bibliographiques. Quant aux géométries, elles sont produites à partir du logiciel *Quantum GIS* (QGIS) qui a l'avantage d'être également libre et multiplateforme, idéal dans un milieu où l'on utilise indifféremment *Windows*, *MacOS X* ou *Linux*.

QGIS s'interface directement avec les données stockées dans le SGBD *PostgreSQL/PostGIS*. Il dispose désormais de plusieurs gestionnaires de base de données, dont *DB Manager* qui permet d'exécuter des requêtes sur les données géographiques et attributaires, et de visualiser directement le résultat. Le travail de production et de mise à jour des géométries peut être fait directement sur les tables stockées dans *PostGIS*.

Publication des données géo-historiques – serveur TinyOWS et protocoles WFS et WMS dans Drupal et OpenLayers

Les données géo-historiques sont publiables via le serveur cartographique *TinyOWS* qui permet de les lire en WMS et WFS-T dans *QGIS* ou dans *OpenLayer* (cf. Figure 5).

Concernant la publication des données, le choix du CMS s'est porté sur *Drupal*. Outre le fait que *Drupal* est compatible avec *PostgreSQL*, il est fiable et robuste et dispose de nombreux modules utiles au système tels que le module *Openlayer* intégrant la publication de cartes en WMS et WFS-T depuis le serveur cartographique *TinyOWS* (cf. fiche descriptive de *Drupal* dans <http://projet-plume.org>). Dans ce contexte, d'autres fonctionnalités liées à l'affichage ou à l'usage ont été développées

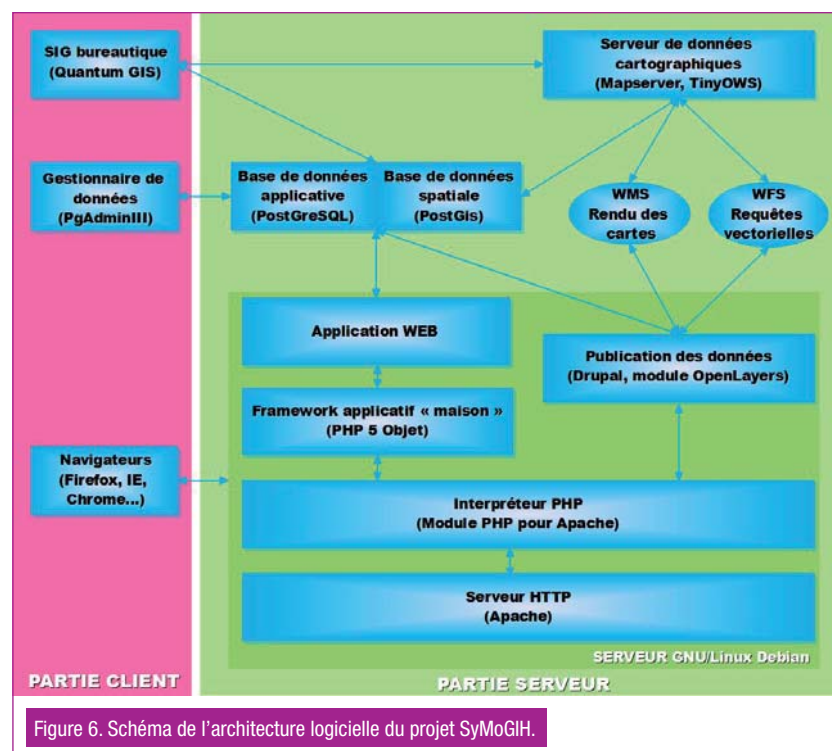


Figure 6. Schéma de l'architecture logicielle du projet SyMoGIH.

comme, par exemple, la visualisation dynamique des évolutions géographiques avec un curseur ou liées à l'interactivité avec les utilisateurs avec l'import de fichiers au format CSV.

L'architecture logicielle repose donc sur des composants libres répondant aux standards d'interopérabilité (cf. Figure 6) : une application *web* codée en PHP, un SGBD *PostgreSQL/PostGIS* et le serveur cartographique *TinyOWS*. Les données sont publiées dans le CMS *Drupal* et le module *Openlayers*. Les standards OGC sont utilisés au maximum notamment avec l'usage des protocoles WMS et WFS-T.

À ce jour cette méthode a fait ses preuves dans plusieurs projets de recherches, individuels et collectifs (cf. <http://patrondefrance.fr>) ; elle a entraîné dans son sillage une réelle prise de conscience du potentiel des outils numériques pour la recherche historique et confronté la production de données par les historiens à la rigueur

des standards informatiques pour en garantir l'interopérabilité et la pérennité.

La deuxième partie de cet article, à venir dans un prochain numéro, illustrera concrètement la démarche du chercheur utilisant la méthode *SyMoGIH*. Pour cela il partira d'un cas concret, celui de la production de l'atlas historique numérique de l'Europe centrale et plus précisément de l'Italie de 1815 à nos jours. |

Références

- <http://SyMoGIH.org>
- <http://geoLarhra.org>
- http://Larhra.ish-lyon.cnrs.fr/Pole_Methodes/index_fr.php
- Francesco Beretta / Pierre Vernus, Le projet SyMoGIH et la modélisation de l'information : une opération scientifique au service de l'histoire, Les Carnets du Larhra 1(2012), 81-107 (cf. HalSHS).
- Baptiste, Jean-Luc, Merise : guide pratique. Modélisation des données et des traitements, langage SQL, 2009
- Gregory I.N. (2003) A place in history: A guide to using GIS in historical research. Oxford: Oxbow Books. p.66