

---

# Oronce Fine : une plateforme pour l'annotation sémantique de ressources cartographiques sur le Web de données\*

Pandolfi Benoît<sup>1</sup>, Brando Carmen<sup>1</sup>, Mermet Eric<sup>1</sup>, Verdier Nicolas<sup>1</sup>, Léa Hermenault<sup>1</sup>

1. Plateforme Géomatique, EHESS  
54 boulevard Raspail 75006 Paris, France  
{prenom.nom}@ehess.fr

---

*RÉSUMÉ.* Ce texte présente la plateforme collaborative en ligne Oronce Fine. Cette plateforme conforme aux technologies et standards du Web sémantique a pour objectif d'être le lieu d'accueil des données des projets scientifiques issus des disciplines, terrains et temporalités divers en sciences humaines et sociales. A travers cet outil le projet Oronce Fine se propose principalement de faciliter l'intégration et le partage de documents anciens intégrant de multiples dimensions spatio-temporelles (cartes et plans anciens, données géographiques format vecteur, images satellite, photographies, textes) et enrichis par des vocabulaires du Web de données. Elle a enfin pour objectif d'être un outil de travail collaboratif pour entre autres des historiens et des archéologues, ainsi que d'être consultable par un large public.

---

*Abstract.* This proposal presents the Oronce Fine online collaborative platform. This platform complies with Semantic Web technologies and standards and aims to accommodate data for scientific projects on various disciplines, fields and time horizons in the humanities and social sciences. Through this tool, we intend to facilitate the integration and sharing of historical documents integrating multiple spatial and temporal dimensions: old maps and plans, geographical vector data, satellite images, photographs, texts, semantically enriched with vocabularies of the Web of data. Finally, it aims to be a collaborative working tool for historians and archaeologists, among others, as well as to be accessible to a wide audience.

---

**Mots-clés :** annotation sémantique, web de données, données anciennes.

**Keywords :** semantic annotation, web of data, geo-historical data.

---

---

<sup>1</sup>\* Avec le soutien financier de l'EHESS et du projet PSL Oronce-Fine, Semantic-enabled platform for the publication, integration and exploration of geo-historical resources-Oronce Fine

## 1. Introduction

La production et la valorisation de données de la recherche est un enjeu pour la reproductibilité et la visibilité des travaux en sciences humaines et sociales (SHS). Dans ce contexte, notre projet vise à constituer et déployer une plateforme collaborative en ligne, baptisée Oronce Fine, construite conformément aux technologies et standards du Web sémantique afin de faciliter l'intégration et le partage de documents anciens hétérogènes intégrant des dimensions spatio-temporelles multiples.

La plateforme Oronce Fine sera le lieu d'hébergement des données des projets scientifiques sur disciplines, terrains et temporalités divers, il est d'ores et déjà un outil de travail collaboratif pour des historiens du Centre de recherche historique (CRH UMR 8558 CNRS/EHESS), ainsi que pour des archéologues rattachés à la Maison de l'Archéologie et de l'Ethnologie (Nanterre). A ce stade, nous nous concentrons particulièrement sur trois cas d'étude : les itinéraires et lieux habités du Chemin de Saint-Jacques dans le nord de l'Espagne, les paysages de l'agriculture d'hier et d'aujourd'hui en Île-de-France, et les études africaines et l'histoire de la "Sémiologie graphique" vues au travers des fonds du Laboratoire de graphique dirigé par Jacques Bertin.

La valeur ajoutée de la plateforme Oronce Fine est de fournir un service de dépôt de ressources cartographiques géoréférencées, de description à partir des métadonnées normalisées (Dublin Core, INSPIRE) et d'annotation sémantique compatible avec le modèle Web Annotation Data Model (W3C). La description et l'annotation s'appuient sur les vocabulaires du Web des données et relie automatiquement à des sources et référentiels externes de données spatialisées sémantisées (Geonames, Getty TGN). La présentation décrira l'entreprise de conception et de développement de notre prototype qui s'appuie sur l'outil ouvert de gestion et de publication de contenus, Omeka-S<sup>2</sup>.

## 2. Outil de gestion de contenus patrimoniaux

Le choix a été fait de ne pas développer un système entier, mais d'utiliser des logiciels existants et ouverts. Ces logiciels répondent chacun à une partie des besoins des différents cas d'étude. Les différentes fonctionnalités sont reliées dans un système centralisé, ici Omeka S. Celui-ci est une application inspirée de *Content Management Systems* et il est développé en code ouvert par le Centre Roy Rosenzweig de l'Université George Mason (Etats-Unis). Omeka-S centralise les données géoréférencées y compris les métadonnées, les données sont également

---

<sup>2</sup> <https://omeka.org/s/>

diffusées en flux WMS (*Web map service*) grâce à un serveur cartographique comme GeoServer<sup>3</sup> et via le protocole OAI-PMH<sup>4</sup>, très répandu dans le monde des bibliothèques numériques. Une seule interface administrateur permet la publication de sources, enrichies sémantiquement par l'utilisateur, et rend possible la gestion de plusieurs sites. Cette fonctionnalité convenait parfaitement au type de plateforme envisagé dans le projet.

### 3. Géoréférencement et intégration

Les sources cartographiques suivent un processus prédéfini de géoréférencement et d'intégration à Omeka S. Ce processus est présenté en figure 1. Ces sources regroupent des documents se présentant sous diverses formes : plans, cartes, images, diapositives, articles, cartes postales, témoignages, etc. Le processus de géoréférencement et d'intégration se déroule en trois phases. Durant la première phase les documents sont numérisés et mis en forme en mode hors ligne par les équipes de recherche, puis envoyés sur un serveur externe d'hébergement des données. Lors de la seconde phase, les données sont normalisées et intégrées à Omeka S ; un identifiant est attribué par document au sein du système, de type ARK<sup>5</sup> dans les meilleurs des cas. La troisième et dernière phase consiste à employer le module d'annotation sémantique d'Omeka S.

Dans le détail, les documents sont, dans un premier temps, numérisés localement. Dans le cadre des trois cas d'études précédemment cités, il s'agit de plus de 250 cartes des fonds de l'ancien Laboratoire de graphique de l'EHESS qui ont été numérisés et plus de 3000 plans, cartes et diapositives du fonds de J. Passini (Passini, 1984), dans le cadre du projet Chemin de Saint Jacques. L'ensemble des documents sont alors décrits par leurs métadonnées dans un tableur préformaté. Celui-ci contient des champs correspondant à la description normée des documents et de leur contenu. Les données pouvant être géoréférencées sont traitées sous le logiciel QGIS<sup>6</sup>, sur des fonds disponibles sur le Web comme OpenStreetMap en WGS84. Une fois géoréférencée, l'emprise des contenus est extraite par vectorisation. Le vecteur ainsi obtenu permet de découper les images géoréférencées pour ne garder que le contenu spatialisé. Cette emprise est aussi stockée en WKT (*well-known text*)<sup>7</sup> pour servir plus tard au processus d'annotation spatiale sous Omeka S.

Toutes les données sont ensuite transférées sur le serveur et font l'objet d'un retraitement. Les données-images sont stockées à la fois sous forme brute, afin d'être disponible au téléchargement, et sous forme normalisée afin d'être intégrées

---

<sup>3</sup> <http://geoserver.org>

<sup>4</sup> Acronyme anglais pour *Open Archives Initiative Protocol for Metadata Harvesting*

<sup>5</sup> [http://www.bnf.fr/fr/professionals/issn\\_isbn\\_other\\_identifiers/a.ark\\_en.html](http://www.bnf.fr/fr/professionals/issn_isbn_other_identifiers/a.ark_en.html)

<sup>6</sup> <https://www.qgis.org>

<sup>7</sup> Il s'agit d'une représentation textuelle de la géométrie d'une entité spatiale

aux sites diffusés par Omeka S. Les images géoréférencées sont, de leur côté, tuilées pour ensuite être intégrées à un serveur Geoserver et diffusées en WMS (*Web Map Service*). Le tableau contenant les métadonnées subit, lui aussi, des modifications car les champs préformatés sont mappés sur des ontologies choisies : Dublin Core, Bibliographic Ontology et INSPIRE pour les cas d'études précédemment cités. Le tableau est ensuite transformé en CSV et stocké sur le serveur.

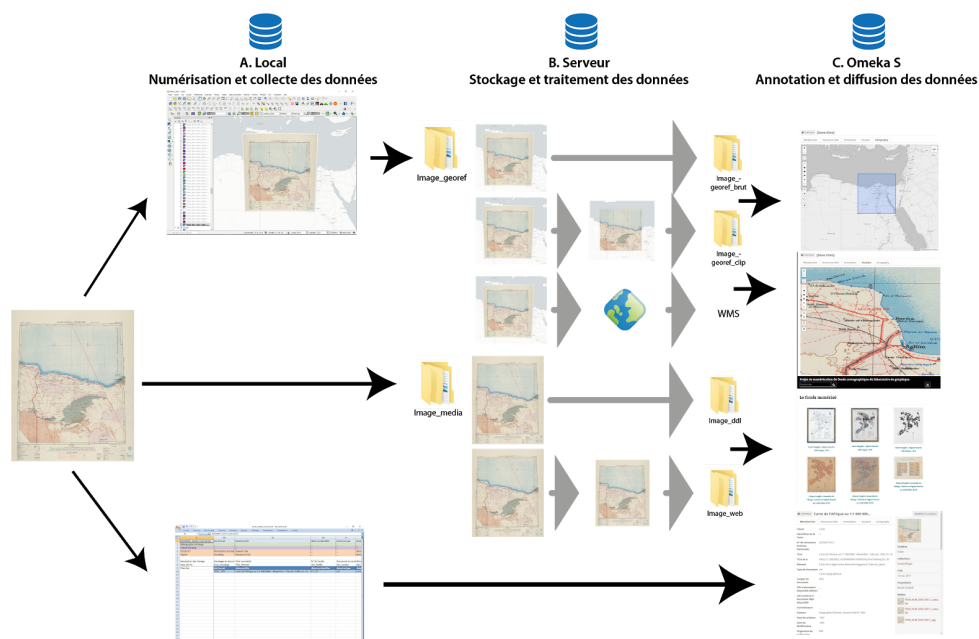


FIGURE 1. *Processus de géoréférencement et d'intégration des sources à la plateforme Oronce Fine*

À l'issue de cette phase de traitement ces éléments sont intégrés à Omeka S. Les métadonnées du CSV deviennent des items décrits par des métadonnées. Les images retraitées sont associées à ces items sous forme de media, et les images brutes rendues disponibles par des liens de téléchargement. Sous Omeka S grâce à deux modules développés dans le cadre du projet Oronce Fine il est désormais possible d'annoter à la fois le contenu des images (media) lui-même et d'annoter spatialement les données. Pour finir Omeka S, diffuse sous forme de Sites web les items, les medias, les annotations et les WMS dans les sites propres à chaque cas d'étude. Les métadonnées sont aussi diffusées selon le protocole OAI-PMH.

#### 4. Module d'annotation sémantique

Un module d'annotation sémantique a été développé sous forme de plugin Omeka S<sup>8</sup> par D. Berthureau. Cette implémentation s'appuie sur les recommandations de la W3C qui conseille l'utilisation du *Web annotation Data model* (WADM) ainsi que l'ontologie associée, le *Web Annotation Vocabulary* (WAC). L'annotation est plus précisément définie comme une ressource composée d'un *body* qui correspond au commentaire ou à la ressource descriptive ainsi que d'une *target* qui constitue la ressource concernée par ce qui est décrit dans le *body* de l'annotation. L'annotation peut également être caractérisée par des propriétés additionnelles comme un type parmi celles répertoriées dans la WAC et un identifiant IRI (*Internationalized resource identifier*). En ce qui concerne la dimension spatiale, la *target* de l'annotation fait souvent référence à une géométrie exprimée en WKT afin de démarquer la zone concernée par l'annotation. Le *body* d'une annotation peut aussi pointer à une IRI d'un gazetier sémantisé externe comme Géonames ou bien interne au sien d'Omeka S, afin d'identifier le lieu dont il est question dans l'annotation. La figure 2 illustre deux annotations sur le plan du cadastre d'Estella, qui fait partie du fonds Chemin de St Jacques sur l'interface d'annotation que nous proposons.



FIGURE 2. Deux annotations sur le Plan du cadastre de Estella (Chemin de St. Jacques), l'une porte sur un extrait du Chemin de St Jacques (en bleu) et l'autre désigne l'entité spatiale Église du Saint Sépulcre (en violet).

Un extrait de ces mêmes annotations sérialisées en JSON-LD (oa constitue le préfixe de la WAC) sont présentées ci-dessous.

```
motivation: "linking",
body: [ { type: "o:Item",
value: "http://psig.huma-num.fr/omeka-s/api/items/634145" } ... ] , ← Église du
Saint Sépulcre
target: { ... type: "oa:Selector", format: "application/wkt", value: "POLYGON
```

<sup>8</sup> <https://github.com/Daniel-KM/Omeka-S-module-Annotate> et <https://github.com/Daniel-KM/Omeka-S-module-Cartography>

((1746.719783 3882, 1744.71958 3788, 1952.740654 3776, 1958.741262 3870, 1746.719783 3882))"}}

Cet ensemble de spécifications décrit un modèle structuré permettant aux annotations d'être partagées et utilisées entre applications via un API REST (*representational state transfer*). L'avantage de l'outil Omeka S est la possibilité d'ajouter simplement une ontologie quelconque si l'on souhaite étendre le modèle d'annotation avec des nouveaux champs de description. Dans d'autres systèmes d'annotation de cartes comme Recogito<sup>9</sup>, la géométrie des entités n'est pas une information qui appartient à l'annotation. Au contraire, ce choix est fait dans un outil d'annotation (non sémantique) et de géoréférencement de cartes comme Maphub<sup>10</sup>. Pour Recogito, c'est le gazetier (ou référentiel) qui fournit l'accès aux descriptions des entités géographiques et permet de tracer leur évolution dans le temps, à condition néanmoins que le gazetier pertinent existe. Ceci représente un choix intéressant qui mérite d'être discuté largement avec la communauté en humanités spatiales.

## Conclusion

La plateforme Oronce Fine est une plateforme capable de fournir un service de dépôt de données. Son originalité tient en ce que ces dernières sont décrites à partir de métadonnées normalisées et qu'elles peuvent être annotées grâce à un module développé dans le cadre de ce projet. La démonstration du fonctionnement de ce nouvel outil permet de réaliser une première évaluation critique des fonctionnalités de la plateforme et de quelle manière elle répond concrètement aux besoins de production et de valorisation de données de la recherche en SHS.

## Bibliographie

- Grassi M., Morbidoni C., Nucci M., Fonda S., Piazza F., « Pundit: augmenting web contents with semantics ». *Literary Linguist Computing*, 28 (4): 640-659. doi: 10.1093/lc/fqt060, 2013.
- Passini J., 1984, *Villes médiévales du chemin de Saint-Jacques-de-Compostelle (de Pampelune à Burgos) Villes de fondation et villes d'origine romaine*. Éditions Recherche sur les Civilisations, "Mémoire" n° 47, 183 p.
- Simon R., Barker E., Isaksen L., De Soto Canamares P., « Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2 », *Journal of Map & Geography Libraries* Vol. 13 , Issue 1, 2017

---

<sup>9</sup> <https://recogito.pelagios.org/>

<sup>10</sup> <http://maphub.github.io/>