

Modèles et outils pour la publication de métadonnées d'archives géographiques et de leurs données dérivées

Melvin Hersent ⁽¹⁾, Nathalie Abadie ⁽¹⁾, Bertrand Dumenieu ⁽²⁾, Julien Perret ⁽¹⁾,

Humanistica - juin 2023



(1)



(2)



SoDUCo ANR-18-CE38-0013



Licence Creative Common BY-SA 4.0

Plan

Partie I : Les spécificités des projets géohistoriques

- Les sources géohistoriques

- Leur traitement

Partie II : Modélisation

- La description de l'information géographique

- La description des archives

- La description de sources géohistoriques

- La description de la provenance

Partie III : Les outils pour déployer notre modèle

- Notre catalogue

- Notre outil pour le web sémantique

- Les outils annexes

Partie I : Introduction

Les spécificités des projets géohistoriques

SoDUCo (Social Dynamics in Urban Context)

Open tools, models, and data - Paris and its suburbs, 1789-1950

<https://soduco.github.io/>

<https://anr.fr/Projet-ANR-18-CE38-0013>

2019-2023



I - Le projet ANR SoDUCo



Information
géographique



Histoire



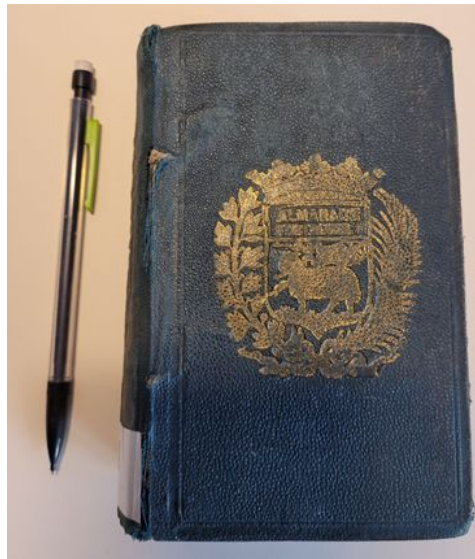
Informatique



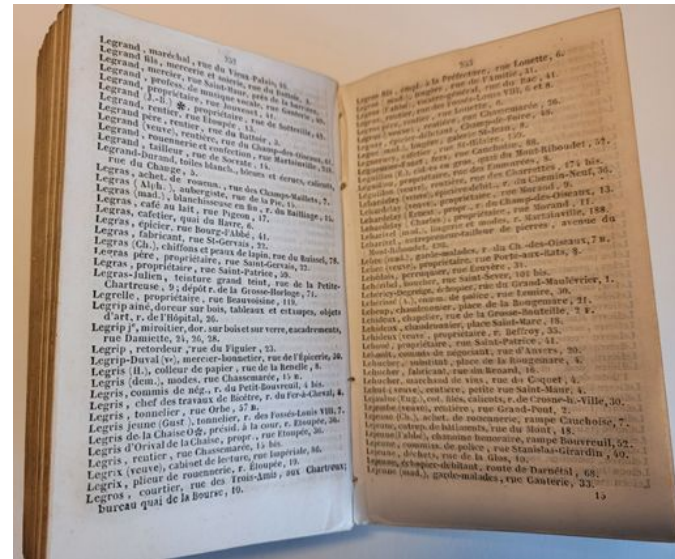
Archivistique

Partie I.1 : Les sources

I.1 - Les sources du projet



Exemple d'un
"Répertoire du commerce"
du XIX^e s.



Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibl. que protestante (Société), Moulins, 16.
Bibron, aide-natural., au Muséum d'hist. nat.
Bibus, tailleur, Roule, 21.
Bibus, tailleur, Richelieu, 31.
Bical et Dorre, fab. de socques, Vertbois, 14.
Bican (Mme), fondeur en cuivre, cour de la Corderie-du-Temple, 26.
Richard (Mme), Nve-de-Luxembourg, 17.
Richard, tabacs et eau-de-vie, Faub.-St-Martin, 45.

Didot 1841a - page 95

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibl. que protestante (Société), Moulins, 16.
Bibron, aide-natural., au Muséum d'hist. nat.
Bibus, tailleur, Roule, 21.
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bical et Dorre, fab. de socques, Vertbois, 14.
Bican (Vve) et fils, fondeur en cuivre, place de la Corderie-du-Temple, 26.
Bicel, épicier, marché d'Aguesseau, 15.
Richard (Mme), Nve-de-Luxembourg, 17.
Richard, tabacs et eau-de-vie, Faub.-St-Martin, 45.

Didot 1842a - page 117

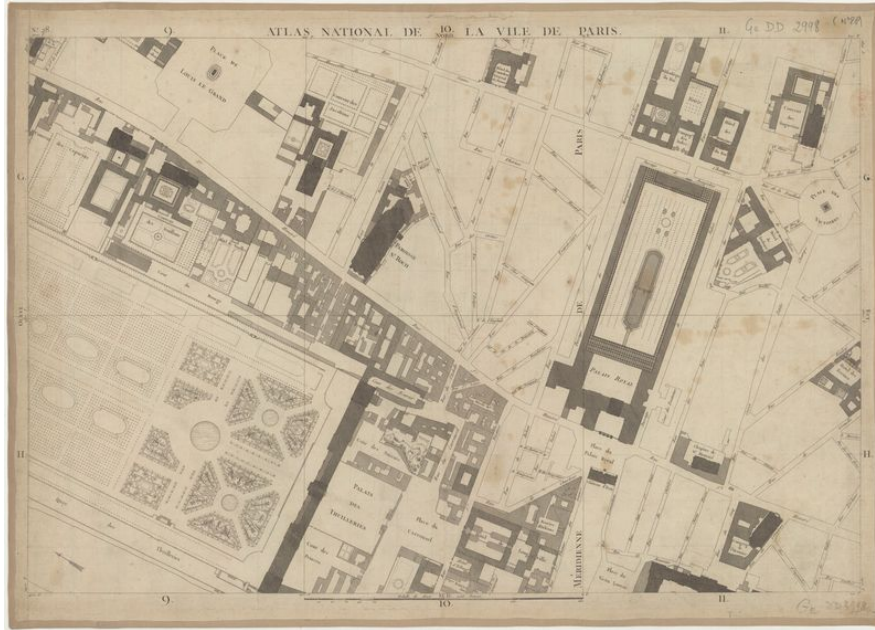
Bibliothèque Ste-Geneviève, rue des Sept-Voies et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bical et Doire, fab. de socques, Vert-Bois, 14.
Bican (Vve) et fils, fondeurs en cuivre, place de la Corderie-du-Temple, 26.
Bicel, épicier, Marché-d'Aguesseau, 15.
Richard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1843a - page 129

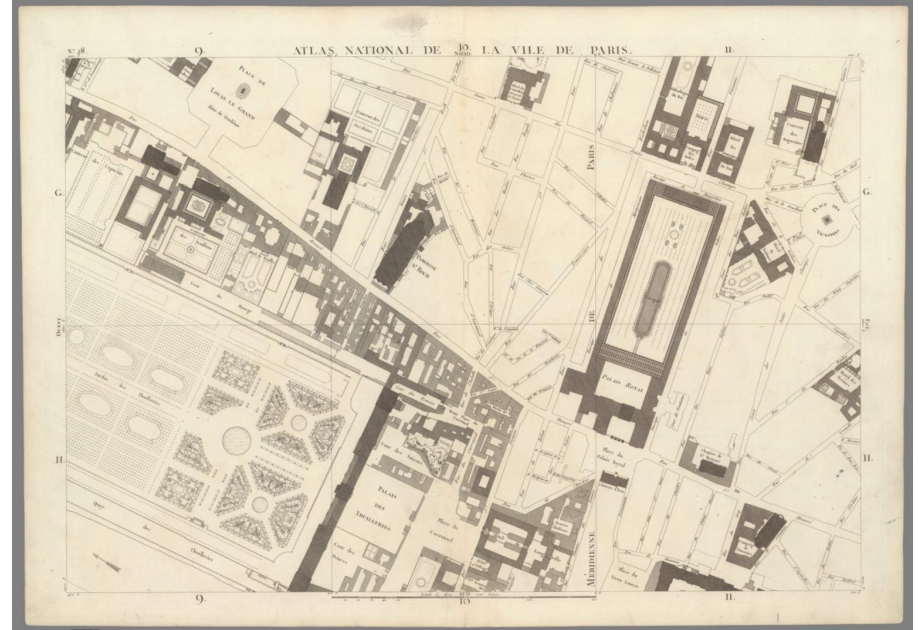
Bibliothèque Ste-Geneviève, rue des Sept-Voies et place du Panthéon.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibolet, relieur, passage Sainte-Marie-Saint-Germain, 10.
Bibonne, architecte, Magasins, 12.
Bibron, aide-naturaliste au Jardin-des-Plantes, Cuvier, 29.
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bican (Vve) et fils, fondeurs en cuivre, place de la Corderie-du-Temple, 26.
Bicel, épicier, Marché-d'Aguesseau, 15.
Richard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1844a - pages 125-126

I.1 - Les sources du projet



Bibliothèque nationale de France,
ark:/12148/cb40708746v



David Rumsey Map Center,
G1844.P3 V4 1795 F.

Atlas du plan général de la ville de Paris, planche n°28

I.1 - Les sources du projet

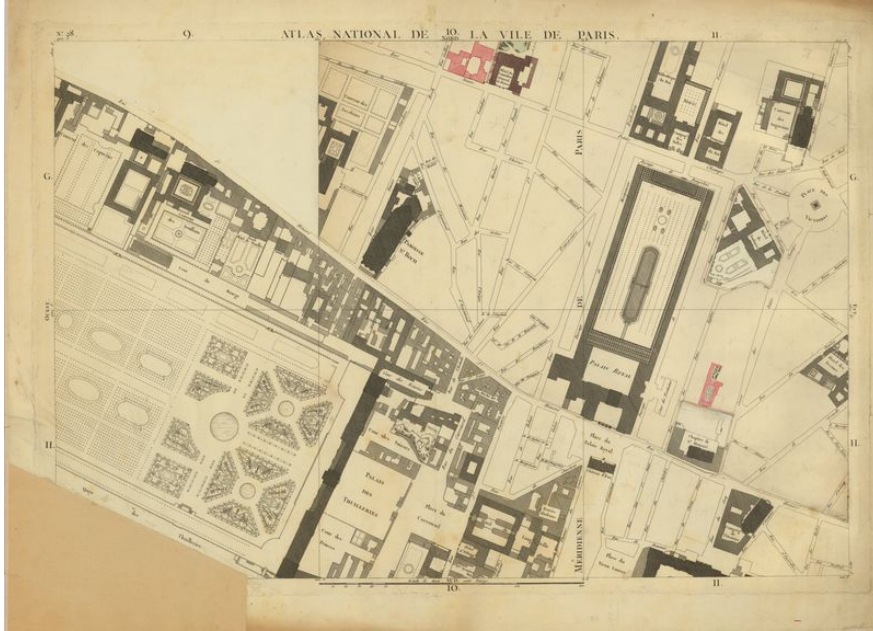


Planche n°28, rabats fermés

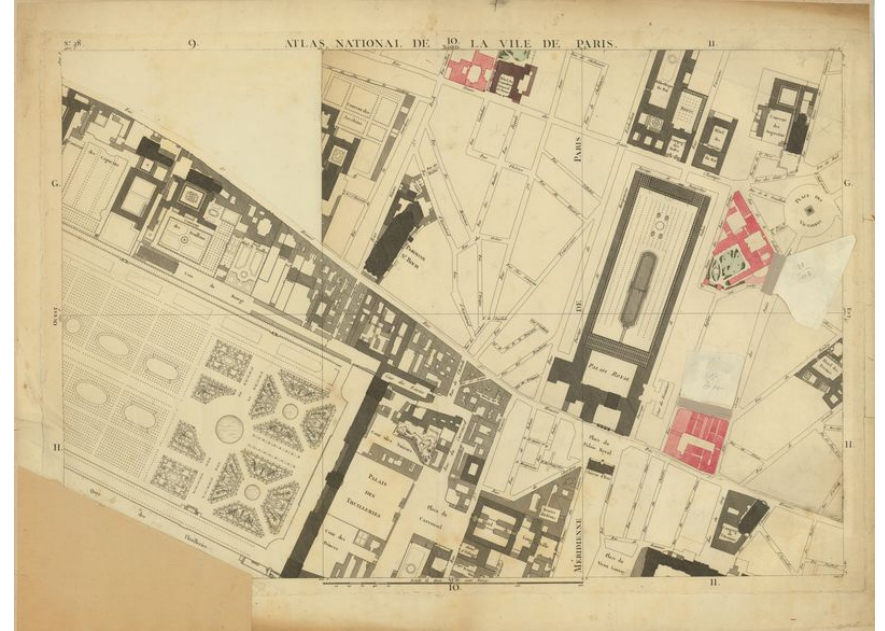


Planche n°28, rabats ouverts

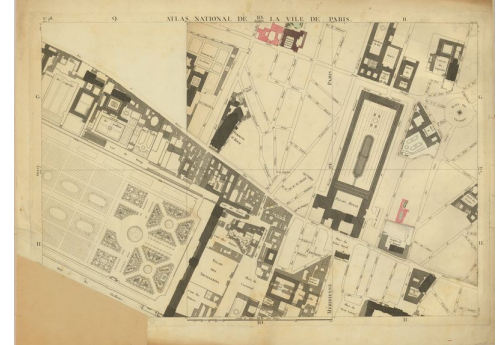
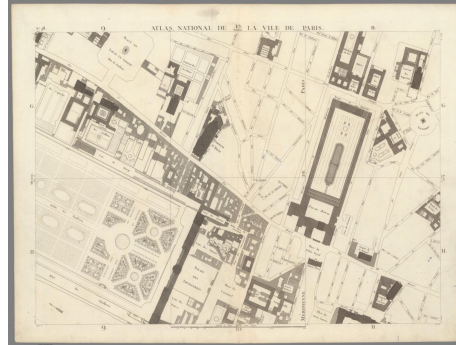
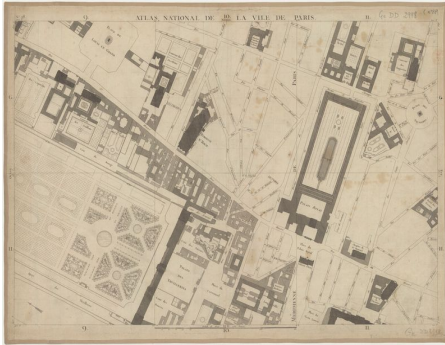
Réemploi dans le cadre de l'élaboration de la carte d'État-Major (première moitié du XIX^e s.), Fonds IGN

Atlas du plan général de la ville de Paris, planche n°28

I.1 - La notion d'Item et d'Instance

Item = contenu intellectuel d'une oeuvre

Instance = Expression physique d'une oeuvre



Trois instances de l'item planche n°28 de *l'Atlas du plan général de la ville de Paris*

Partie I.2 : Le traitement des sources géohistoriques

I.2 - Des sources géographiques anciennes à l'analyse géohistorique

Reconnaissance

Classification

Géo-référencement

Structuration

Image → Géométries



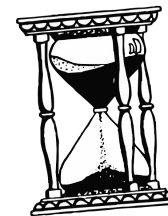
Géométries → Annotations



Données géométriques → Données géographiques



Données géographiques → Données géohistoriques



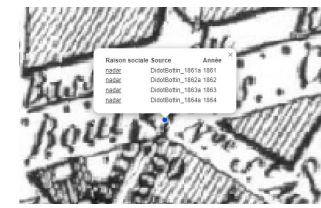
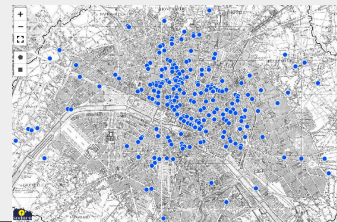
Scan → Texte

Texte → Entités nommées spatiales

Entités nommées spatiales → Données géographiques → Données géohistoriques

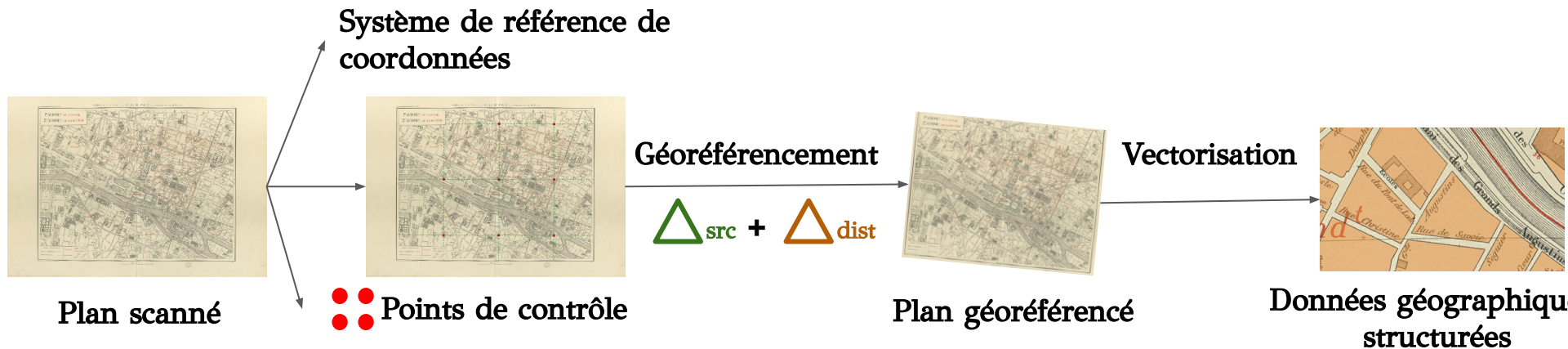
Ravrio et comp., fabr. de bronzes et curiosités, r. Richelieu, 93; la fabrique rue Montmartre, 161.

Ravrio et comp. PESE , fabr. de bronzes et curiosités ACT
r. Richelieu LOC , 93 CARDINAL ; la fabrique FT
rue Montmartre LOC , 161 CARDINAL



Nom	Source	Année
LOC	DroBDrn_1051a	1961
LOC	DroBDrn_1052a	1962
LOC	DroBDrn_1053a	1963
LOC	DroBDrn_1054a	1964

1.2 - De nombreuses ressources géographiques à documenter



Associer à chaque pixel du plan des coordonnées exprimées dans un système de référence de coordonnées.

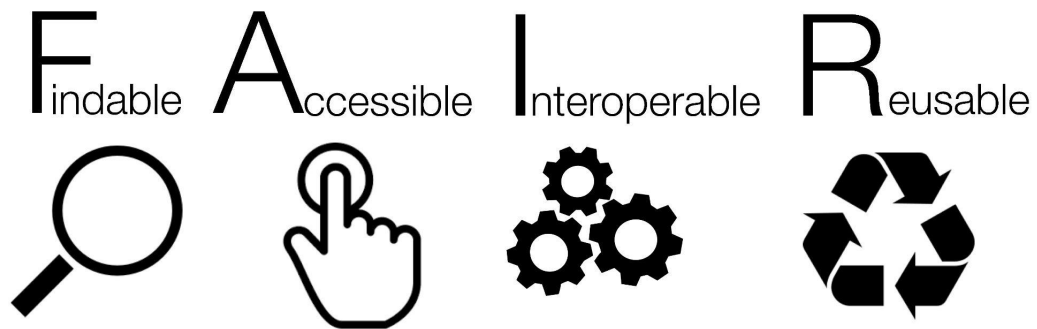
I.2 - Un catalogue pour documenter le projet

- Documenter les sources utilisées au sein du projet (“étude critique”)
- Documenter les données dérivées des sources
- Plateforme entre les chercheurs

Partie II : Modélisation

Standards et vocabulaires pour décrire les archives géographiques

II - Open Science et diffusion des données : les bonnes pratiques



- **F**acile à trouver : Identifiant unique et pérenne, liées à la donnée source et sauvegardées ou indexées
- **A**ccessible : Récupérable par l'identifiant via un protocole standardisé, même si la donnée source disparaît
- **I**nteropérable : Utilise un langage défini et partagé, **utilise un vocabulaire FAIR** et pointe vers d'autres métadonnées
- **R**éutilisable : Licence permettant l'usage des données, une origine claire et un vocabulaire adapté au domaine

Partie II.1 : La description d'information géographique

II.1 - L'ISO 19115

Standard européen pour la description des métadonnées en information géographique



La norme **ISO 19115** est définie par l'ISO/TC211 (comité technique en information géographique)

La commission européenne **Inspire** impose son utilisation

L'**ISO 19115** est un modèle et possède son schéma d'application XML : l'**ISO 19139**

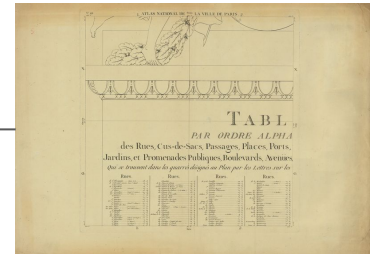
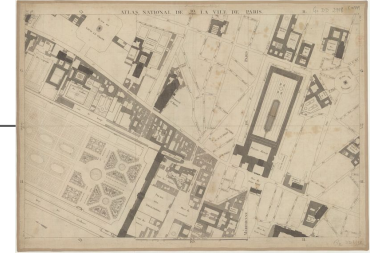
II.1 - Description des sources en suivant l'ISO 19115



Dossier
Verniquet



Source: gallica.bnf.fr / Bibliothèque nationale de France



DS_Aggregate

DS_Dataset

DS_Resource

II.1 - Les avantages de l'ISO 19115

- Particulièrement adaptée pour décrire des données institutionnelles
- Fourni tous les éléments nécessaire pour décrire nos données :
 - Système de coordonnées de référence
 - Emprise et résolution spatiales
 - Type de géométrie
 - Format des données
- Peut être étendu à des besoins spécifiques d'autres domaines

II.1 - Les inconvénients de l'ISO 19115

- La documentation est payante
- Ne respecte pas les principes FAIR. Il manque pour chaque jeu de données:
 - Une URI unique et pérenne (F)
 - D'être indexable par les moteurs de recherche (F)
 - De lier les ressources entre elles (I)
 - D'exposer les données via un protocole de communication standardisé (A)
- Inadapté pour des sources historiques
- Manque la notion d'Instance

Partie II.2 : La description des archives

II.2 - Records in Context - Conceptual Model (RiC-CM)

Norme recommandée par les archives nationales pour décrire les métadonnées des documents d'archives



La norme **RiC-CM** est définie par l'*Expert Group on Archival Description* (EGAD), dont font partie les **Archives nationales**

RiC-CM possède son vocabulaire : RiC Ontology (**RiC-O**)



II.2 - Description des sources en suivant RiC - CM

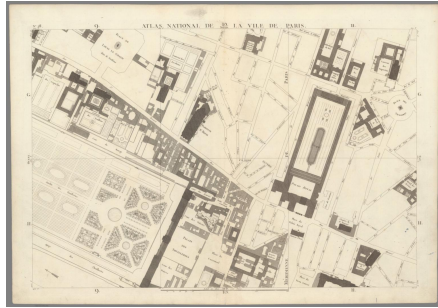


Planche n°28

Record

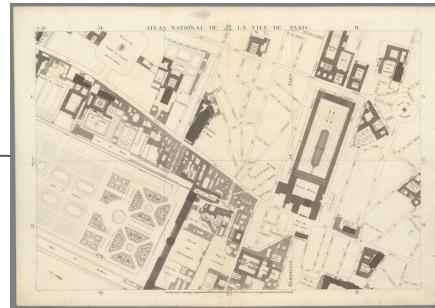


Planche n°28
Imprimée

Instantiation

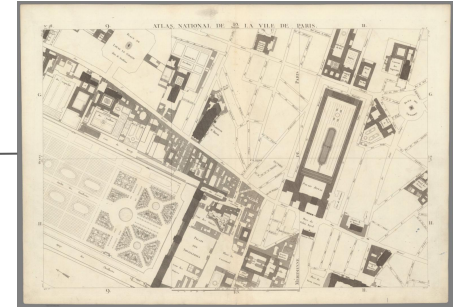


Planche n°28
scannée

Instantiation
dérivée

II.2 - Les avantages de RIC-CM

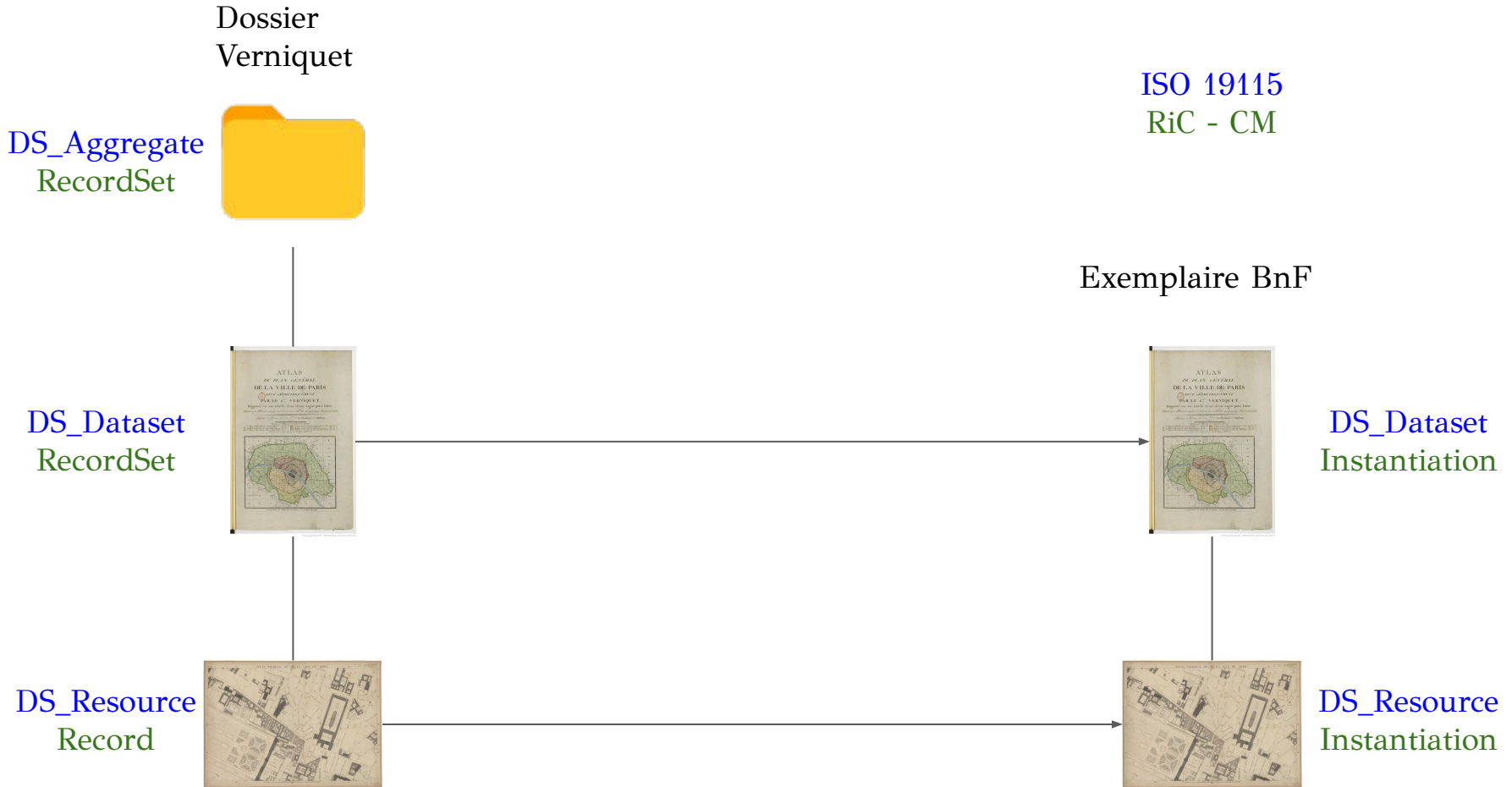
- Modèle spécifiquement conçu pour décrire des archives
- Expérience existante au sein de l'équipe avec ce modèle
- Extension pour décrire les archives géographiques réalisée au sein d'un précédent projet

II.2 - Les inconvénients de RIC-CM

- Pas compatible nativement avec l'ISO 19115

Partie II.3 : L'alignement de l'ISO 19115 et de RiC-CM

II.3 - Alignement entre les deux modèles



II.4 : La description de la provenance

IV - La documentation de la production scientifique

Dossier
Verniquet

DS_Aggregate
RecordSet



DS_Dataset
RecordSet



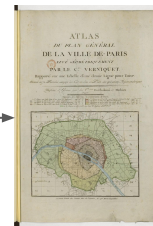
DS_Resource
Record



Que se passe-t-il ici ?

Exemplaire BnF

DS_Dataset
Instantiation



DS_Resource
Instantiation



IV - La documentation de la production scientifique

Que se passe-t-il ici ?

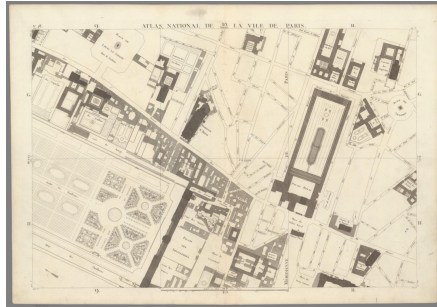


Planche n°28
imprimée

Instantiation

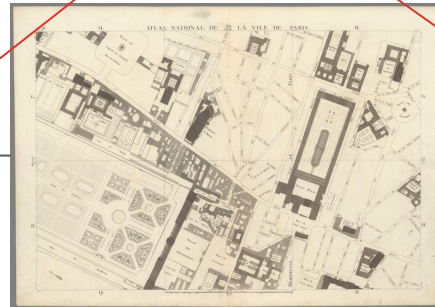


Planche n°28
scannée

Instantiation
dérivée

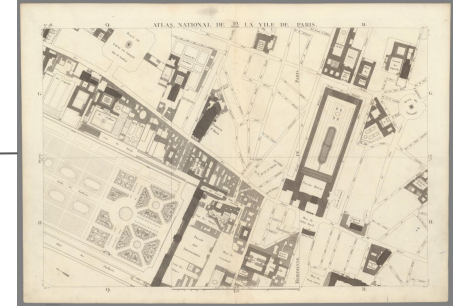


Planche n°28
géoréférencée

Instantiation
dérivée

IV - PROV Data Model (PROV-DM)

Norme recommandée pour décrire les entités, activités et personnes participant à la production de quelque chose



La norme **PROV-DM** est définie par le *World Wide Web Consortium* (W3C)

PROV-DM possède son vocabulaire : PROV Ontology (**PROV-O**)

IV - Alignement avec ISO 19115 et RiC-CM

Vocabulaire existant pour décrire les activités :

- ISO 19115 : classes *Lineage* (*LI* et *LE*), avec *ProcessStep* et *Processing*
- RiC-CM : notion de *Mechanism*

Partie III : Les outils pour déployer notre modèle

Proposition d'une infrastructure de publication de nos
métadonnées et de leurs données dérivées

III - ISO 19115 : rappel des ajouts nécessaires

- Ne respecte pas les principes FAIR. Il manque pour chaque jeu de données:
 - Une URI unique et pérenne (F)
 - D'être indexable par les moteurs de recherche (F)
 - De lier les ressources entre elles (I)
 - D'exposer les données via un protocole de communication standardisé (A)

Partie III.1 : Le catalogue

III.1 - Geonetwork, catalogue de ressources géographiques


- Utilise nativement la norme ISO 19115
- Propose un identifiant unique pour chaque ressource (F)
- Utilisé par d'autres structures (GeOrchestra)
- Interface de saisie et de manipulation des métadonnées

catalog.geohistoricaldata.org


Retour à la carte

Télécharger Mode affichage







Atlas national de la Ville de Paris, feuille N. [uméro] 28 [exemplaire BnF LK7-6043]



Découvrir les données



Links

	annotation_verniquet_bnf_28.json	Ouvrir le lien	
	annotation_verniquet_bnf_28.txt	Ouvrir le lien	
	iiif image api	Ouvrir le lien	

Catalogue GeoNetwork pour publier les métadonnées ISO 19115

III.1 - Recréer notre modèle dans Geonetwork

Dossier
Verniquet

DS_Aggregate
RecordSet



1 ressource = 1 fiche de métadonnée

DS_Dataset
RecordSet



Exemplaire BnF

DS_Dataset
Instantiation



DS_Resource
Record



DS_Resource
Instantiation



Partie III.2 : Alignement des modèles

III.2 - Ontop, notre réponse pour se conformer aux principes FAIR

- Mise en place d'un graphe virtuel
- Mise en place d'un *Endpoint* SPARQL
- Publication des métadonnées selon les bonnes pratiques du Web de données
:
 - Indexables par les moteurs de recherche (F)
 - Liées entre elles (I)
 - Exposées via une API (A)

III.2 - Ontop, porteur des *mappings* pour l'alignement

Les *mappings*, une solution pour “traduire” un vocabulaire vers l'autre :



target cat:recordset/{res1} rico:hasInstantiation cat:instantiation/{res2} .
source SELECT REC.uuid as res1, REL.uuid as res2
 FROM "public"."rico_recordsets_mapping" AS REC
 JOIN "public"."rico_horizontal_relations" AS REL
 ON REC."uuid"=REL."parent";

} Requête SQL

III.2 - Exemple de requête SPARQL avec Ontop

Ontop SPARQL endpoint

endpoint address: <https://ontop.geohistoricaldata.org/sparql> | ontop v4.2.0

Query x Query 1 x +

```
1 PREFIX rico: <https://www.ica.org/standards/Ric/ontology#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 SELECT ?document ?keyword ?creatorName ?RecordSet ?Instantiation ?crs ?bbox
5 WHERE {
6   <http://catalog.geohistoricaldata.org/rdf/id/recordset/d4b335d0-820c-4530-941e-26f3bae60fd1>
7   a ?type;
8   rico:hasCreator/rdfs:label ?creatorName;
9   #rico:isAssociatedWithDate/ ?date;
10  <http://data.ign.fr/def/geometrie#crs> ?crs;
11  rico:isOrWasIncludedIn/rico:name ?RecordSet;
12  rico:hasInstantiation/rico:name ?Instantiation;
13  rico:name ?document;
14  rico:descriptiveNote ?keyword.
15  <http://catalog.geohistoricaldata.org/rdf/id/recordset/d4b335d0-820c-4530-941e-26f3bae60fd1>
16  <http://data.ign.fr/def/xysemantics#extent>/<http://data.ign.fr/def/xysemantics#geographicElement>/<http://data.ign.fr/def/xysemantics#polygon>/<http://www.opengis.net/ont/geosparql#asWKT> ?bbox.
17 }
```



Table Response Pivot Table Google Chart Geo

Showing 1 to 1 of 1 entries

Search: Show 50 entries

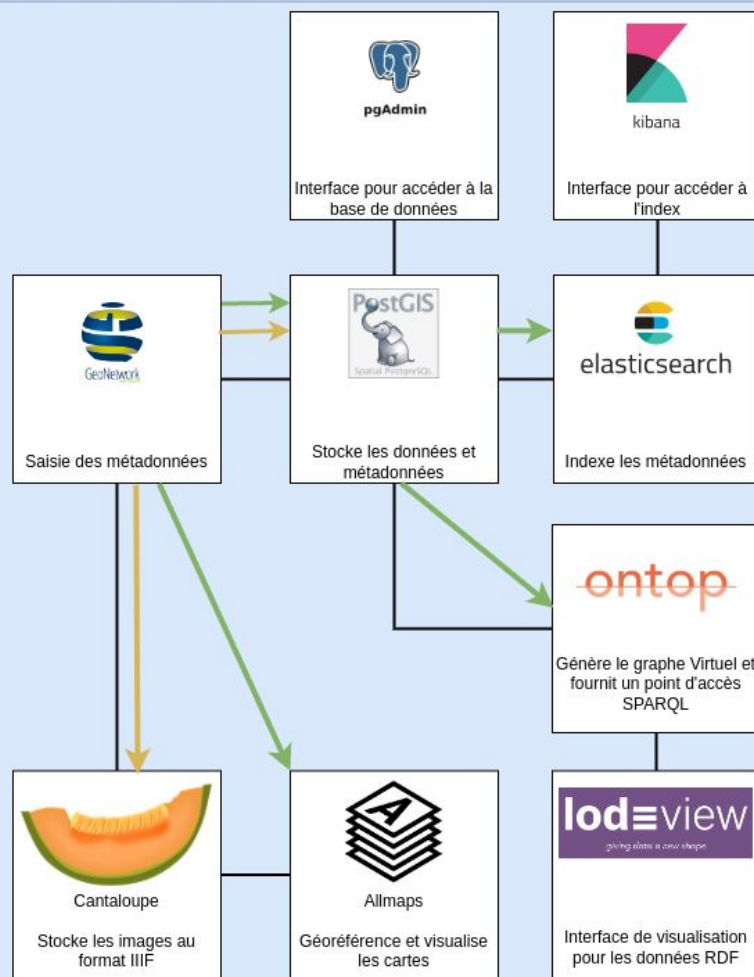
document	keyword	creatorName	RecordSet	Instantiation	crs	bbox
1 Atlas national de la Ville de Paris, feuille N.[uméro] 28	RecordSet	Edme Vermaquet	Atlas du plan général de la Ville de Paris	Atlas national de la Ville de Paris, feuille N.[uméro] 28 [exemplaire BnF LK7-6043]	http://data.ign.fr/id/ign/crs/WGS84G	"POLYGON((2.32590101964404 48.8608923858877,2.32590101964404 48.8679031817315,2.34183933623433 48.8679031817315,2.34183933623433 48.8608923858877,2.32590101964404 48.8608923858877))"

Showing 1 to 1 of 1 entries

Partie III.3 : Les outils annexes

Une architecture en microservices


- Utilisation de Docker pour simplifier le déploiement :
 - Simplifie l'ajout ou le retrait de service
 - Simplifie les sauvegardes et mises à jour
- Hébergement d'image IIIF
- Outils de visualisation



Soduko / Geohistoricaldata catalogue Search Map Sign in English

Back to search Download Display mode


Atlas du plan général de la ville de Paris, feuille N.[uméro] 28 [exemplaire BnF, GE DD-2998 & IFN-53243704]



Overview

Temporal extent
01-01-1784 → 31-12-1799

Discover data



API

Format

WMS Add to map

verniquet_bnf_28
Verniquet BnF - Feuille 28

JPEG

Download

Powered by GeoNetwork 4.2.5.SNAPSHOT About GitHub API

Conclusion

Conclusion

- Proposition d'alignement d'une norme en information en géomatique et en archivistique (**ISO 19115 / RiC-CM**)
- Travail à poursuivre avec l'ajout de la provenance dans nos métadonnées
- Proposition d'une infrastructure pour implémenter ce modèle

Merci pour votre attention !

Liens :

SoDUCo : soduco.github.io/

Catalogue : catalog.geohistoricaldata.org

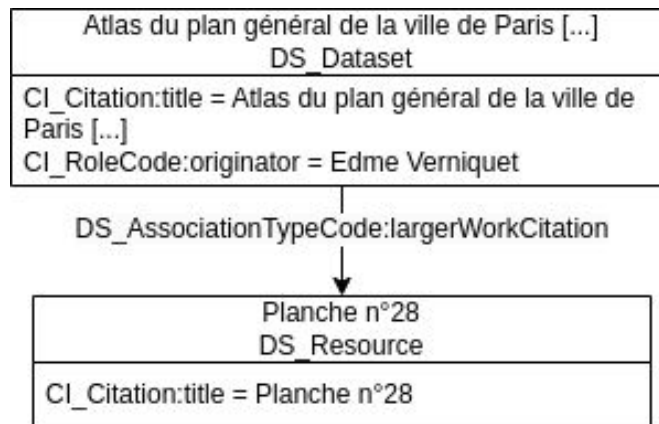
Ontop : ontop.geohistoricaldata.org/

Viewer : soduco.github.io/AllmapsAnnotations/

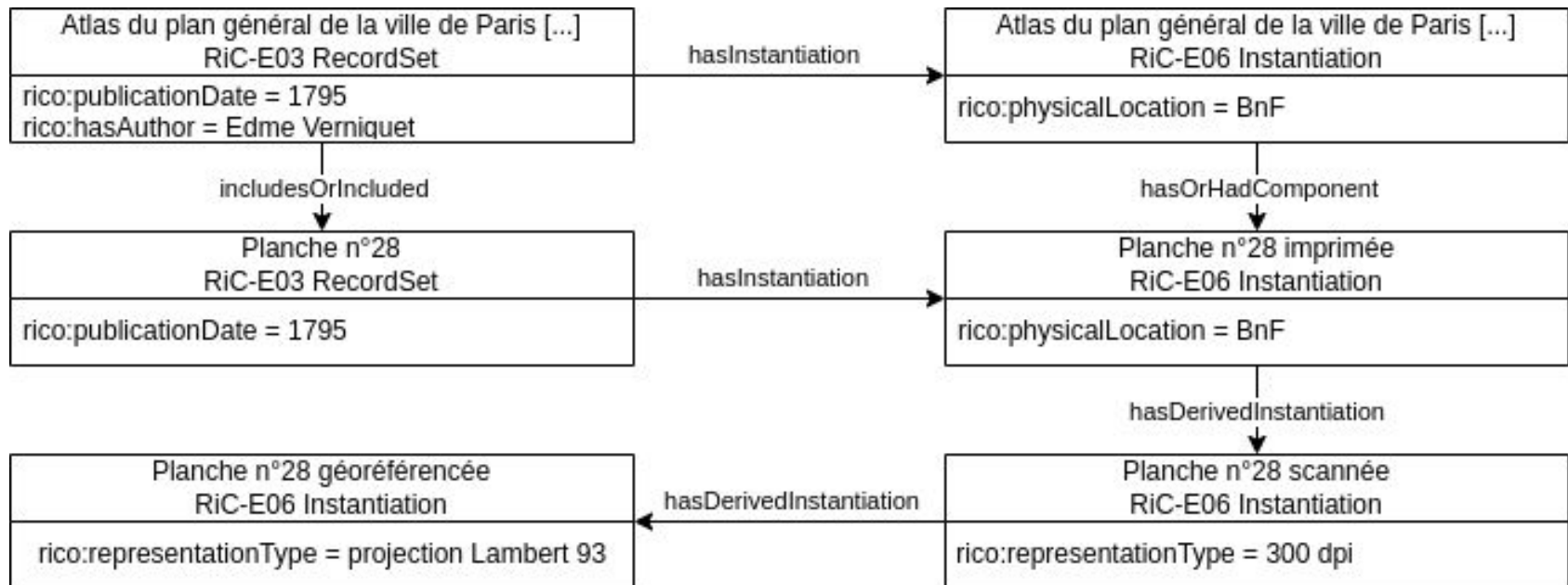
Bibliographie :

- Plan de la ville de Paris, avec sa nouvelle enceinte, levé géométriquement sur la Méridienne de l'Observatoire par le Cen [Edme] Verniquet. Parachevé en 1791. Dessiné et gravé [an II -VII] par les Cens P.[aul] T.[homas] Bartholomé et A.[ntoine] J.[oseph] Mathieu. Écrit par [Pierre Toussaint] Bellanger in : Atlas recomposé du plan général de la ville de Paris, levé géométriquement par le Cen Verniquet [(ed.)]. Rapporté sur une échelle d'une demie ligne pour toise. Divisé en 72 Planches, compris les cartouches et plan des opérations trigonométriques, plus une planche de couverture entoillée, et un tableau d'assemblage tardif. L'an iv, s.d., s.l. [1793-1799, ?], Bibliothèque Nationale de France, IFN-53243704. <http://ark.bnf.fr/ark:/12148/cb40708746v>
- Plan de la ville de Paris, avec sa nouvelle enceinte, levé géométriquement sur la Méridienne de l'Observatoire par le Cen [Edme] Verniquet. Parachevé en 1791. Dessiné et gravé [an II -VII] par les Cens P.[aul] T.[homas] Bartholomé et A.[ntoine] J.[oseph] Mathieu. Écrit par [Pierre Toussaint] Bellanger in : Atlas recomposé du plan général de la ville de Paris, levé géométriquement par le Cen Verniquet [(ed.)]. Rapporté sur une échelle d'une demie ligne pour toise. Divisé en 72 Planches, compris les cartouches et plan des opérations trigonométriques, plus une planche de couverture. L'an iv, s.d. [1793-1799], David Rumsey Map Center, G1844.P3 V4 1795 F.
- Atlas municipal des vingt arrondissements de Paris. 1925. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935524>
- Yizi Chen, Edwin Carlinet, Joseph Chazalon, Clément Mallet, Bertrand Dumenieu, et al.. Vectorization of historical maps using deep edge filtering and closed shape extraction. 16th International Conference on Document Analysis and Recognition (ICDAR'21), Sep 2021, Lausanne, Switzerland. pp.510-525, <10.1007/978-3-030-86337-1_34>. <hal-03256073>
- Diego Calvanese, Benjamin Cogrel, Sarah Komla-Ebri, Roman Kontchakov, Davide Lanti, Martin Rezk, Mariano Rodriguez-Muro, and Guohui Xiao. [Ontop: Answering SPARQL Queries over Relational Databases](#). In: Semantic Web Journal 8.3 (2017), pp. 471–487.

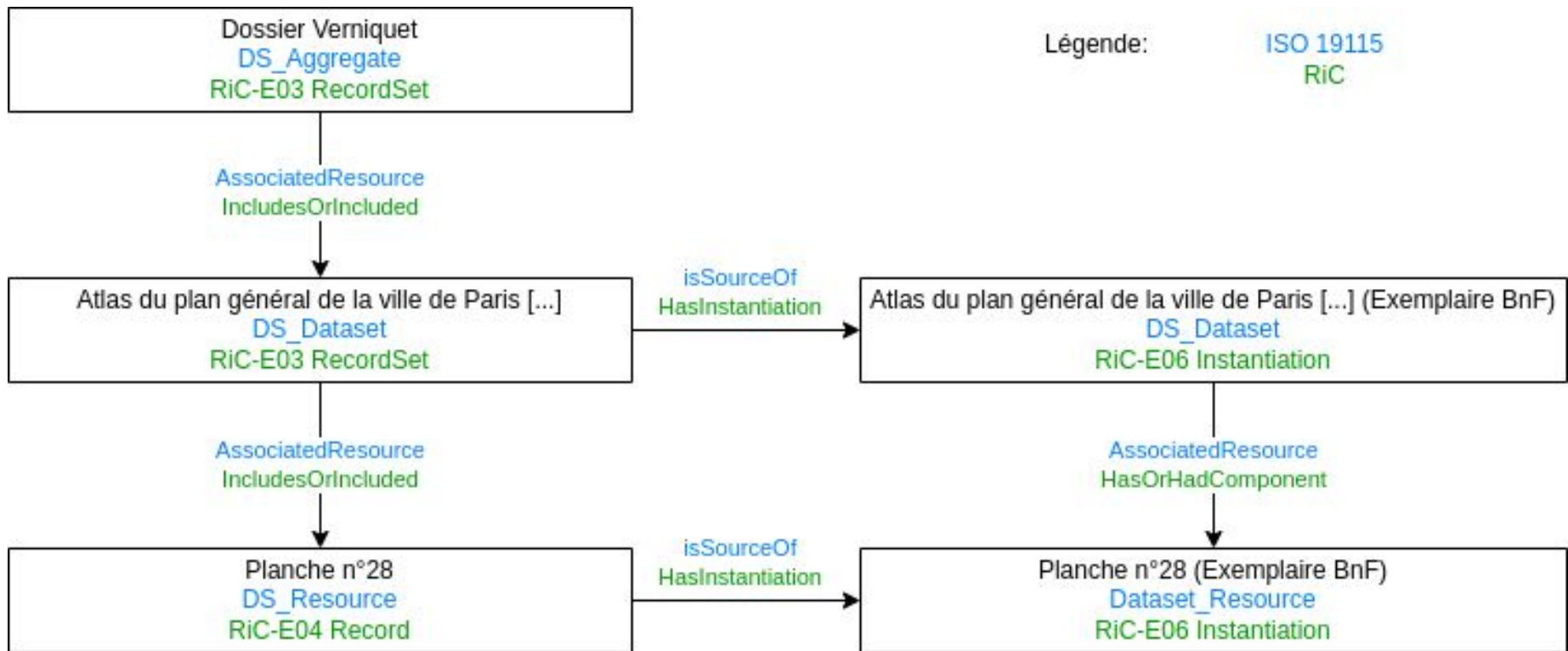
Modélisation suivant l'ISO 19115



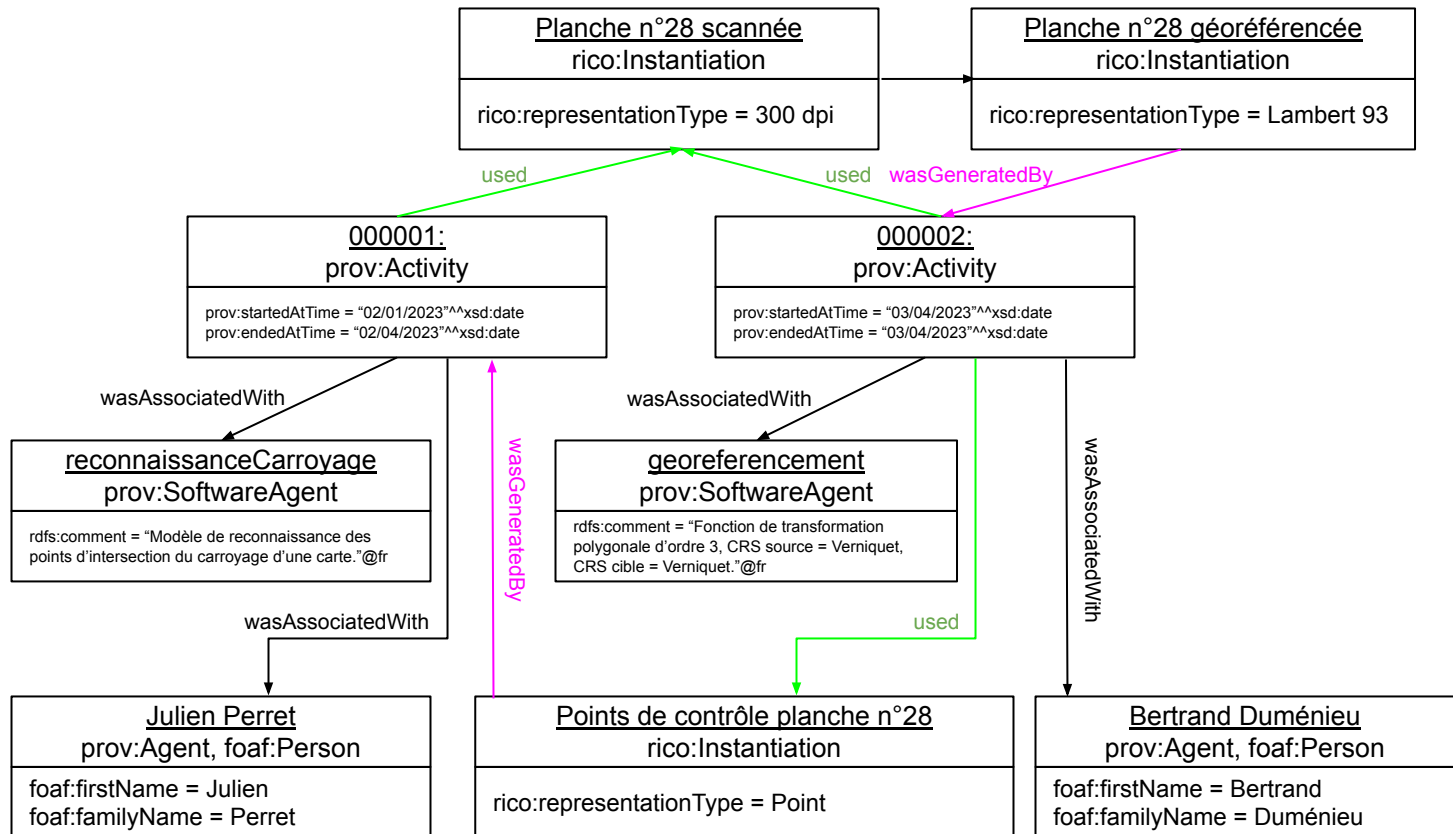
Modélisation suivant RiC - CM



Proposition d'alignement ISO 19115 / RiC-CM



Décrire les relations généalogiques entre ressources avec Prov-O



Liens entre RiC-CM et IFLA-LRM

