



« ET SI ON INDEXAIT LES DONNÉES ? »

Tentative de déconstruction du paradigme métadatique.

Guillaume SUEUR
Rencontres DecryptaGeo 2015

ETUDE DE CAS

Françoise a très envie de réaliser une carte précise d'un site qu'elle aime beaucoup en Cerdagne, les gorges de la Carança, près de Thuès-entre-Valls...



ETUDE DE CAS

Françoise repère sur GoogleMaps quelques sites à faire figurer sur sa carte, et notamment la rivière « La Carença ».



ETUDE DE CAS

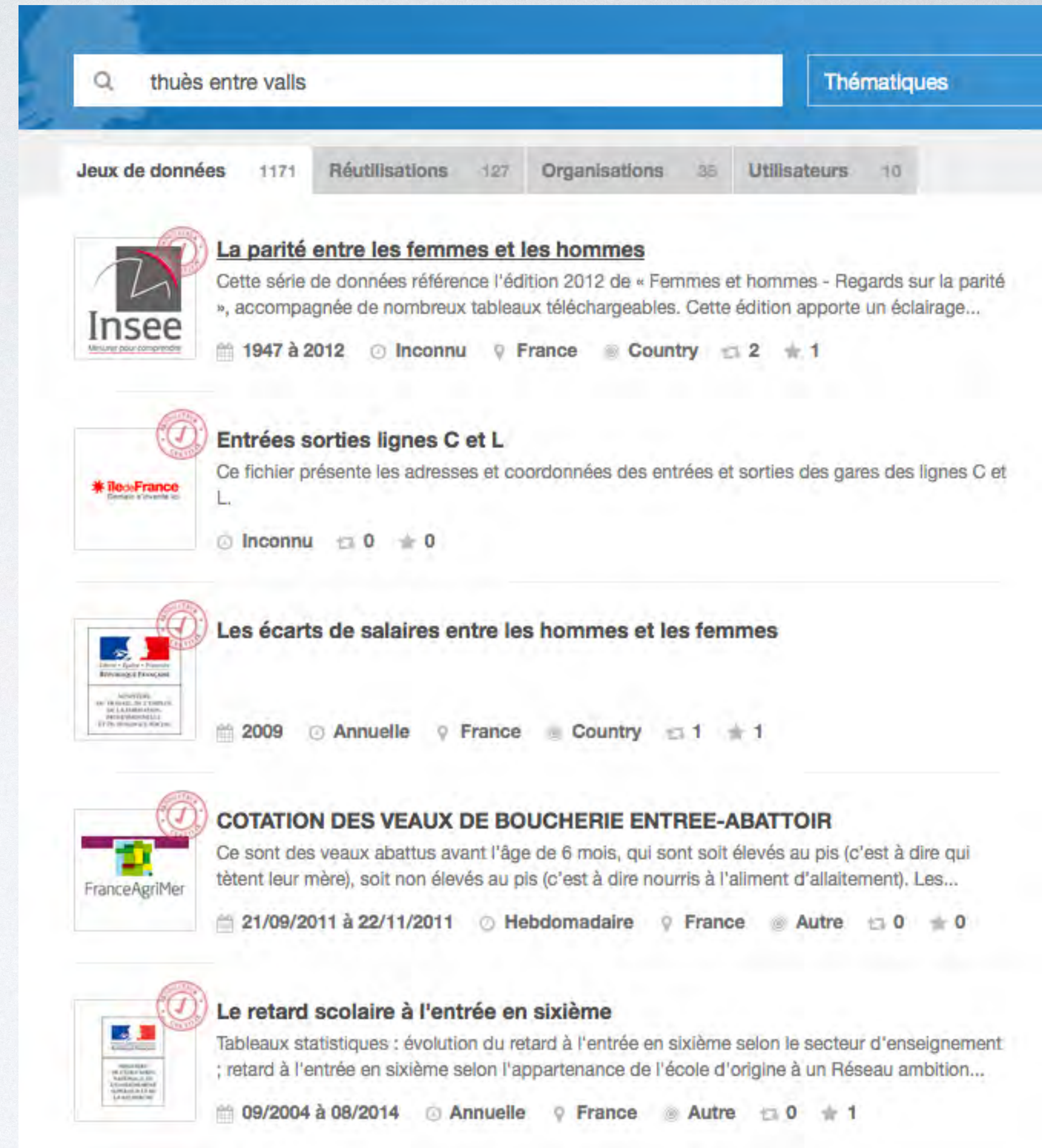
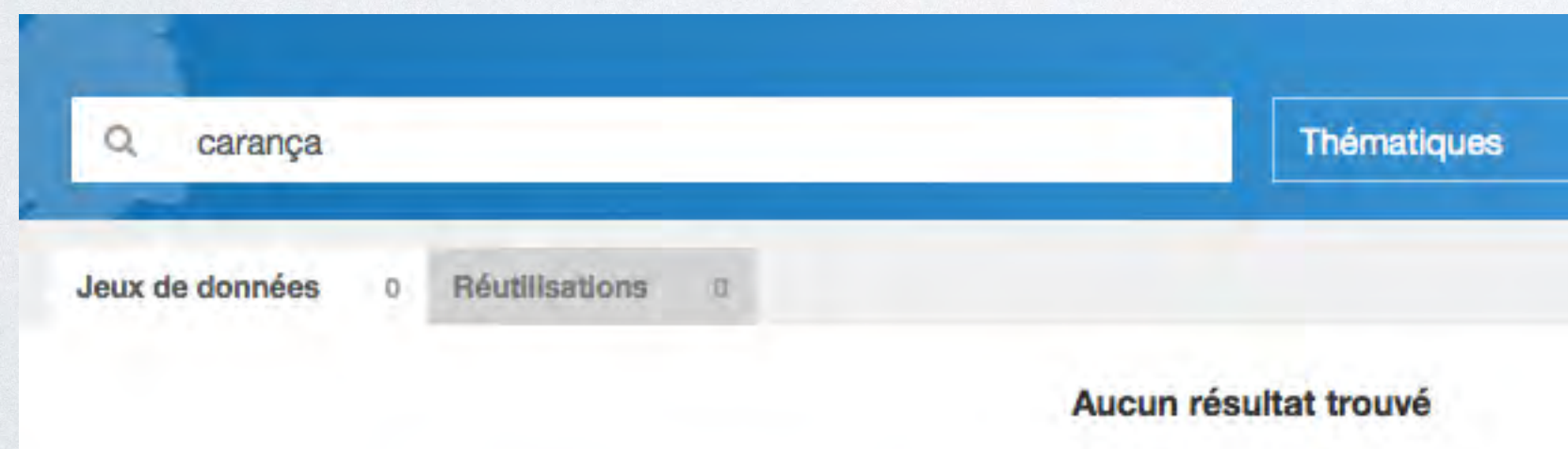
Françoise entend parler depuis plusieurs années de sites web qui permettaient de trouver des données géographiques. Elle espère bien y trouver de quoi faire sa carte...



The screenshot shows the homepage of data.gouv.fr. At the top left is the French Republic logo with the motto 'Liberté • Égalité • Fraternité' and 'RÉPUBLIQUE FRANÇAISE'. Next to it is the 'data.gouv.fr' logo. On the right, it says 'Plateforme ouverte des données publiques françaises'. Below this is a navigation bar with links: 'Comment ça marche ?', 'Données', 'Tableau de bord', 'Dataconnexions', 'Etalab', and 'CADA'. On the far right of the navigation bar are flags and a link for 'Connexion / Inscription'. The main content area has a search bar labeled 'Recherche'. Below the search bar is a vertical list of categories: 'Agriculture et alimentation', 'Culture', 'Économie et Emploi', 'Éducation et Recherche', 'International et Europe', 'Logement, Développement Durable et Énergie', 'Santé et Social', 'Société', and 'Territoires et Transports'. To the right of this list is a large blue box with the text 'Partagez, améliorez et réutilisez les données publiques' and a button that says '+ CONTRIBUEZ !'.

ETUDE DE CAS

Mais sur data.gouv, pas de Carança, et beaucoup de résultats non pertinents pour la commune de Thuès-entre-Valls



ETUDE DE CAS



Dans le GeoCatalogue,
aucun résultat dans les deux cas...

Françoise est bien triste, elle ne va pas
pouvoir faire sa carte...

QUOI OÙ

PRÉCISEZ VOTRE RECHERCHE
Aucun affinage possible

Aucun résultat

Trier par pertinence  5 résultats / page 



Votre recherche n'a donné aucun résultat. Merci de faire une nouvelle recherche en précisant vos critères.

Précédent 1 Suivant

QUOI OÙ

PRÉCISEZ VOTRE RECHERCHE
Aucun affinage possible

Aucun résultat

Trier par pertinence  5 résultats / page 

Votre recherche n'a donné aucun résultat. Merci de faire une nouvelle recherche en précisant vos critères.

Précédent 1 Suivant

ETUDE DS CAS

Françoise appelle alors un **ami**...



ETUDE DE CAS

Il lui explique que les données existent bien. Elles sont même classées selon 34 thèmes pour mieux s'y **retrouver**.

QUOI communes OÙ

PRÉCISEZ VOTRE RECHERCHE

Résultats 1 à 5 sur 29262

Trier par pertinence 5 résultats / page

Type de ressource

- Données (28499)
- Visualisables (2795)
- Accessibles via WFS (1912)
- Téléchargeables (1819)
- Services (601)
- Consultation (307)
- WFS (256)
- Téléchargement (27)
- Autres (11)

★ **Communes françaises (version SIH)**
Communes françaises (Métropole + DOM)

★ **Communes limites**
Limites de communes unités administratives: les communes

QUOI hydrographie OÙ

PRÉCISEZ VOTRE RECHERCHE

Résultats 1 à 5 sur 879

Trier par pertinence 5 résultats / page

Type de ressource

- Données (857)
- Visualisables (394)
- Accessibles via WFS (152)
- Téléchargeables (142)
- Cartes (12)
- Interactives (12)
- Services (10)
- Consultation (7)
- WFS (2)
- Autres (1)

Thèmes INSPIRE

- Hydrographie (535)
- Réseaux de transport (78)
- Zones de gestion, de restriction ou de réglementation et unités de déclaration (20)
- Installations agricoles et aquacoles (19)
- Unités administratives (13)
- Occupation des terres (13)
- Altitude (10)
- Ortho-imagerie (8)
- Bâtiments (6)
- Régions biogéographiques (5)

Type de représentation

- Vecteur (629)
- Raster (74)
- Table texte (24)
- T.I.N. (4)

★ **hydrographie lagunaire 1772 (Cassini)**
numérisation de l'hydrographie des lagunes du complexe palavasiens, de la peyrade, de thau, de l'or et de la petite Camargue

★ **hydrographie lagunaire 1965**
numérisation de l'hydrographie des lagunes du complexe palavasiens, de la peyrade et de l'or

★ **Cadastre - Hydrographie - Grand Dijon**
Hydrographie sur le territoire du Grand Dijon

★ **BD TOPO® Hydrographie : tronçons des cours d'eau**
Hydrographie : tronçons des cours d'eau

★ **BD TOPO® Hydrographie : surfaces en eau**
Hydrographie : surfaces en eau

Précédent 1 | 2 | 3 | 4 | 5 | 6 ... Suivant

ETUDE DE CAS

On a donc bien des **données** mises à disposition, mais que les moteurs de recherche n'arrivent à **trouver** que si on passe par leurs **métadonnées**...

Donc à l'heure où **INSPIRE** et **OPENDATA** promeuvent de concert **l'ouverture** des données, on ne peut toujours les trouver qu'en passant par un concept que peu de personnes maîtrisent et qui **restreint** considérablement la capacité à réellement trouver ce que l'on cherche.

Comment en est-on arrivé là ?

BREVE HISTOIRE DES MÉTADONNÉES

- Les **métadonnées** (données décrivant les données géographiques) sont devenues nécessaires quand la quantité de données géographiques produites, détenues ou manipulées par un organisme était devenue telle que même le meilleur géomaticien ne pouvait tout savoir de leur usage.
- Elles ont été popularisées par la mise en oeuvre de la directive **INSPIRE** qui oblige les **producteurs de données** à renseigner des **fiches** dans un **catalogue**.
- Il s'agissait surtout alors de pouvoir **contacter** le producteur afin de lui commander/demander gentiment les données.
- Les métadonnées sont faites **par** des géomaticiens **pour** des géomaticiens.

BREVE HISTOIRE DES MÉTADONNÉES

- Ces mécanismes n'ont jamais été prévus pour **accéder à la donnée directement**. Même les services de téléchargement supposent des modalités de recherche sur la **description** du jeu de données et non son **contenu**.
- La dimension **attributaire** des objets est **négligée**.
- Même (surtout ?) en WFS/FE, une recherche attributaire est **fastidieuse**.
- L'Opendata, apparu plus récemment, a adopté le même principe **d'indexation** de **mots-clés** et **catégories** définis manuellement par le producteur.

BREVE HISTOIRE DES MÉTADONNÉES

- En bref, c'est un peu comme **Yahoo!** aux débuts de l'Internet !
- Il serait temps de passer à **Google** !

LE CLOUD A LA RESCOUSSE

- Les **capacités** du Cloud permettent désormais de s'affranchir des problèmes de stockage qui ont longtemps pénalisé la publication de données géographiques
- La **scalabilité** du Cloud permet de s'adapter au volume à traiter, à publier et à la demande.
- Les outils du BigData sont parfaitement adaptés à la **manipulation de données** géographiques.

LE CLOUD A LA RESCOUSSE

- Les **contraintes techniques** qui ont façonné nos modes de référencement de la donnée géographique ont donc vécu.
- Des outils beaucoup plus **puissants** permettent désormais de s'affranchir de mécanismes intermédiaires pour non plus **chercher** mais **trouver** de la donnée.

ONEGEO, ISSU DE LA R&D NEOGEO

- ◆ Neogeo Technologies développe un puissant moteur de **recherche** et **d'indexation** des données géographiques, **OneGeo**.
- ◆ **OneGeo** permet de **moissonner, stocker, indexer** des flux **WFS**, des fichiers **Opendata**, des données **OpenStreetMap**.
- ◆ **OneGeo** permet d'affiner les **requêtes** et de publier les **résultats**.

ONEGEO, ISSU DE LA R&D NEOGEO

- ◆ **OneGeo** est développé en **node.js** pour l'API et **angular.js** pour la partie cliente.
- ◆ **OneGeo** s'appuie sur une base de données NoSQL distribuée **ElasticSearch** pour **stocker et indexer** l'ensemble des objets géographiques qui lui sont soumis.
- ◆ **OneGeo** dispose de **profils** spécifiques pour les différents flux, lui permettant d'utiliser des **alias**, d'indexer des **champs particuliers** ou encore de les **moissonner** à intervalles réguliers.

ONEGEO

◆ **OneGeo** trouve des résultats **concrets** à vos recherches et permet de les **télécharger**.

The screenshot shows the OneGeo website interface. At the top, the logo 'OneGeo' is displayed in white on a red background. Below the logo is a search bar containing the text 'carança' and a 'Chercher' button. To the right of the search bar is a button labeled 'Afficher la carte'. Below the search bar, there is a section titled 'Filtrer les résultats' with two tabs: 'Service' and 'Couche de données'. Below this, it indicates '5 résultats (0.012s)'. The search results are listed as follows:

- Cours d'eau métropole**
Office International de l'eau / SANDRE
- Cours d'eau métropole**
Office International de l'eau / SANDRE
- Cours d'eau métropole**
Office International de l'eau / SANDRE
- Cours d'eau métropole**
Office International de l'eau / SANDRE
- Cours d'eau métropole**
Office International de l'eau / SANDRE

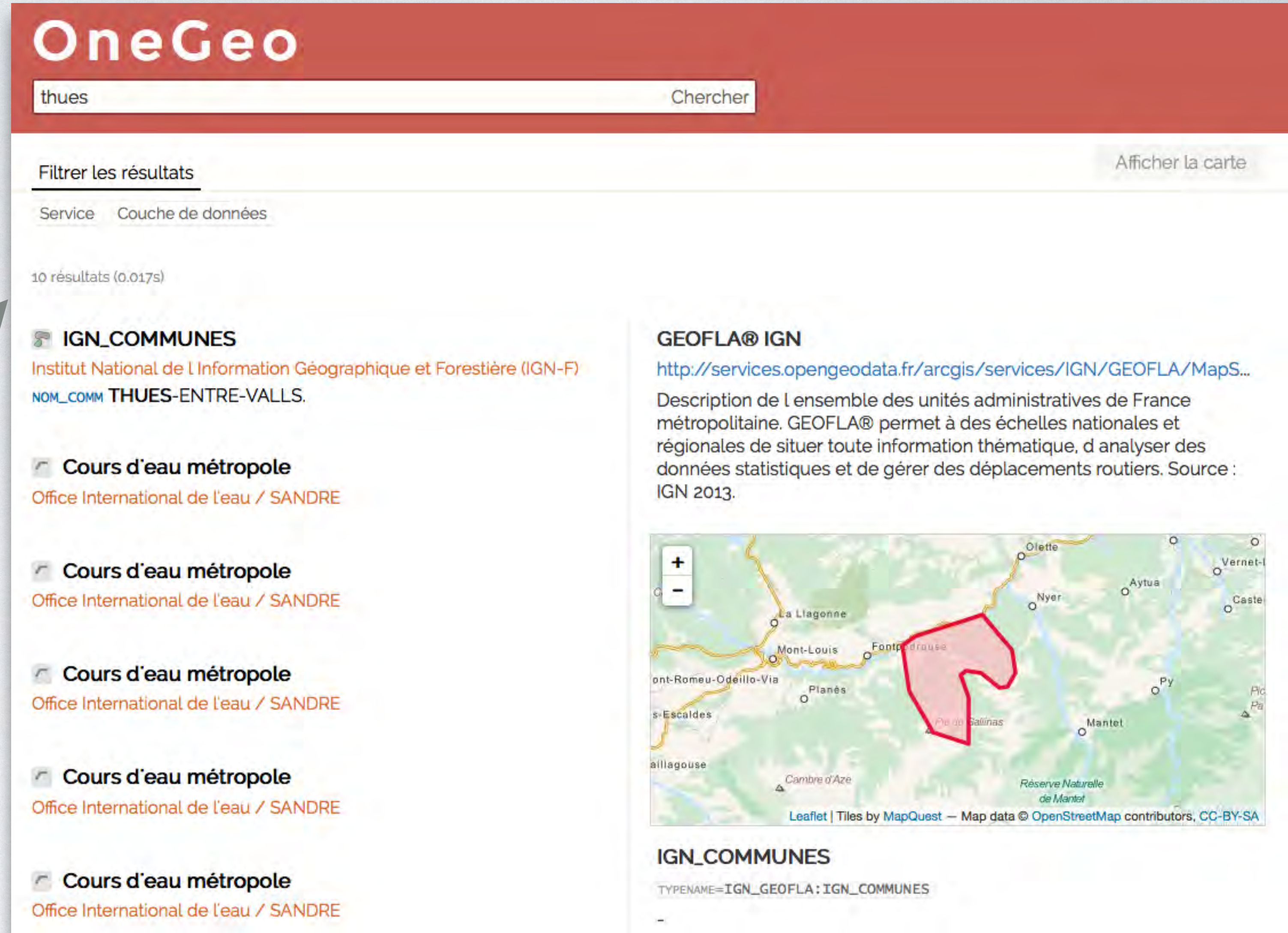
At the bottom of the results list, there is a pagination control with buttons for 'Premier', 'Précédent', '1' (selected), 'Suivant', and 'Dernier'. To the right of the search results, there is a detailed view for the first result, titled 'Référentiel hydrographique Métropole'. It includes a URL: http://services.sandre.eaufrance.fr/geo/eth_FXX? and a description: 'Le Sandre (Service d'Administration Nationale des Données et Référentiels sur l'Eau) met à disposition le référentiel hydrographique.' Below this is a map showing the location of the 'Cours d'eau métropole' in the Pyrenees region, with a red line indicating the water course. The map includes labels for various locations like La Liageonne, Mont-Louis, Fontpédroux, and Pic du Galinas. Below the map, the title 'Cours d'eau métropole' is repeated, followed by the text 'TYPENAME=CoursEau'. At the bottom, there is a table of metadata for the selected result:

gid:	70201
cdEntiteHy:	Y0410640
NomEntiteH:	La Carança
DateMajEnt:	2013-10-28
DateCreati:	2006-01-01
Milieu:	0
Classe:	4
Longueur:	15.327843620054001

ONEGEO

◆ OneGeo répond vite !

10 résultats (0.017s)



The screenshot shows the OneGeo search interface. At the top, the search bar contains the text 'thues' and a 'Chercher' button. Below the search bar, there is a 'Filtrer les résultats' section with tabs for 'Service' and 'Couche de données'. The search results are displayed as a list of 10 items, all of which are 'Cours d'eau métropole' from the 'Office International de l'eau / SANDRE'. The first result is 'IGN_COMMUNES' from the 'Institut National de l'Information Géographique et Forestière (IGN-F)', with the layer name 'NOM_COMM THUES-ENTRE-VALLS'. To the right of the search results, there is a 'GEOFLA® IGN' section with a description and a map. The map shows a geographical area with a red outline, and the text below it reads 'IGN_COMMUNES' and 'TYPENAME=IGN_GEOFLA:IGN_COMMUNES'.

OneGeo

thues

[Afficher la carte](#)

Filtrer les résultats

Service Couche de données

10 résultats (0.017s)

IGN_COMMUNES
Institut National de l'Information Géographique et Forestière (IGN-F)
NOM_COMM THUES-ENTRE-VALLS.

Cours d'eau métropole
Office International de l'eau / SANDRE

Cours d'eau métropole
Office International de l'eau / SANDRE

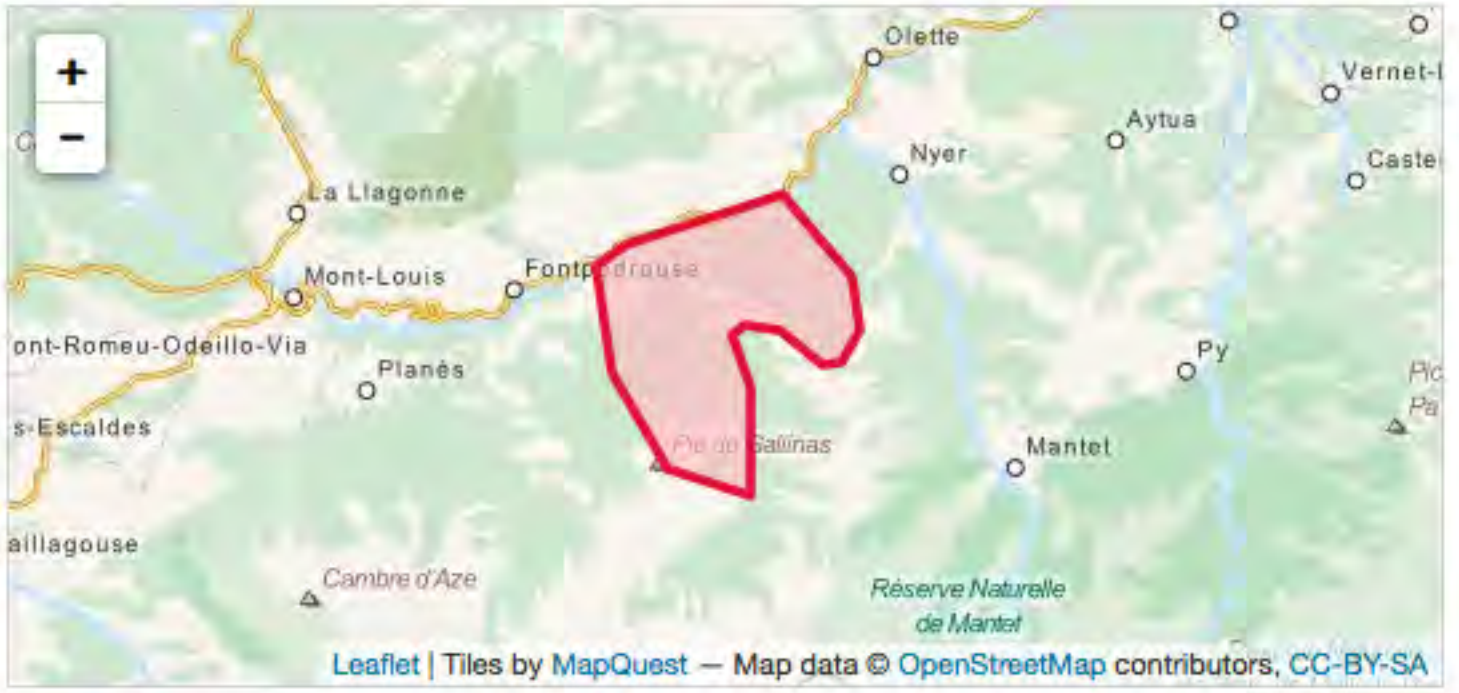
Cours d'eau métropole
Office International de l'eau / SANDRE

Cours d'eau métropole
Office International de l'eau / SANDRE

Cours d'eau métropole
Office International de l'eau / SANDRE

Cours d'eau métropole
Office International de l'eau / SANDRE

GEOFLA® IGN
<http://services.opengeodata.fr/arcgis/services/IGN/GEOFLA/MapS...>
Description de l'ensemble des unités administratives de France métropolitaine. GEOFLA® permet à des échelles nationales et régionales de situer toute information thématique, d'analyser des données statistiques et de gérer des déplacements routiers. Source : IGN 2013.



IGN_COMMUNES
TYPENAME=IGN_GEOFLA:IGN_COMMUNES

ONEGEO

◆ **OneGeo** fait des recherches floues

OneGeo

[Afficher la carte](#)

Filtrer les résultats

Couche de données Service

9 résultats (0.158s)

Station Autopartage
Grand Lyon - Direction des systèmes d'informations et de télécommunications
adresse 30, quai **Perrache**.

Station Autopartage
Grand Lyon - Direction des systèmes d'informations et de télécommunications
adresse Centre d'échange de **Perrache**.


Circuits de déneigement
Grand Lyon - Direction des systèmes d'informations et de télécommunications

Chantier perturbant
Grand Lyon - Direction des systèmes d'informations et de télécommunications
precisionlocalisation Entre la rue Smith et le quai **Perrache**.

Point de canevas topographique du Grand Lyon
Grand Lyon - Direction des systèmes d'informations et de télécommunications
adresse Carrefour Quai de **Perrache** et Cours Suchet.

Serveur Grand Lyon

<https://download.data.grandlyon.com/wfs/grandlyon?>
Base de la plateforme d'intégration de données géographiques du Grand Lyon.
Données sous Licence Ouverte (Etalab).



Leaflet | Tiles by MapQuest — Map data © OpenStreetMap contributors, CC-BY-SA

Station Autopartage

TYPENAME=pvo_patrimoine_voirie.pvostationautopartage
Station Autopartage

nom: Centre d'échange de Perrache
identifiantstation: -
adresse: Centre d'échange de Perrache
commune: Lyon 2^{ème}

ONEGEO

- ◆ **OneGeo** permet de remonter au jeu de données et aux services initiaux

ons et de

Station Autopartage

Voir tous les objets de la couche

TYPENAME=pvo_patrimoine_voirie.pvostationautopartage

Station Autopartage

Référentiel hydrographique Métropole

URL du service OGC:WFS http://services.sandre.eaufrance.fr/geo/eth_FXX?

Le Sandre (Service d'Administration Nationale des Données et Référentiels sur l'Eau) met à disposition le référentiel hydrographique.



Leaflet | Tiles by MapQuest - Map data © OpenStreetMap contributors, CC-BY-SA

Cours d'eau métropole

TYPENAME=CoursEau

ONEGEO

- ◆ **OneGeo** dispose de filtres à facette et de pagination des résultats.

The screenshot displays the OneGeo search interface. At the top, the search bar contains the text 'perrache'. Below the search bar, there are two tabs: 'Couche de données' (selected) and 'Service'. A dropdown menu is open under 'Couche de données', listing four items with checkmarks and counts: 'pvo_patrimoine_voirie.pvostationautopartage (2)', 'gid_deneigement.gidcircuitdeneigement (1)', 'pvo_patrimoine_voirie.pvochantierperturbant (1)', and 'top_plan_canevas.topcanevas (1)'. Below the dropdown, the text 'télécommunications' is visible in orange, followed by 'adresse 30, quai Perrache.' in blue. A second result is shown below, titled 'Station Autopartage' with a bus icon, followed by 'Grand Lyon - Direction des systèmes d'informations et de télécommunications' in orange and 'adresse Centre d'échange de Perrache.' in blue. At the bottom, a pagination bar shows 'Premier', 'Précédent', '1' (highlighted in blue), 'Suivant', and 'Dernier'.

UN NOUVEAU PARADIGME ?

- ◆ Nous sommes désormais à l'heure du **Cloud**, qui peut tout **ingurgiter**, tout **avaler** et tout **recracher** s'il est pourvu des **outils** adéquats.
- ◆ Pourquoi rester **enfermés** dans une vision du référencement **manuel** et **métier** des jeux de données cartographiques ?
- ◆ Pourquoi **imposer** aux gestionnaires de données la tâche **fastidieuse** de saisir des fiches de métadonnées qui ne permettront pas forcément aux utilisateurs de **trouver** leur bonheur ?
- ◆ L'enjeu de l'OpenData et de l'**ouverture** des données est un enjeu **citoyen** qui ne doit pas s'adresser qu'à des experts. Travailler entre nous, c'est facile, mais il va falloir rendre nos données réellement **accessibles et utilisables**.

MERCI POUR VOTRE ATTENTION
ET PLACE AUX QUESTIONS !