



La plateforme **OPEN DATA** de l'information touristique

Blaise de CARNÉ-CARNAVALET – Serwan CRAVIC

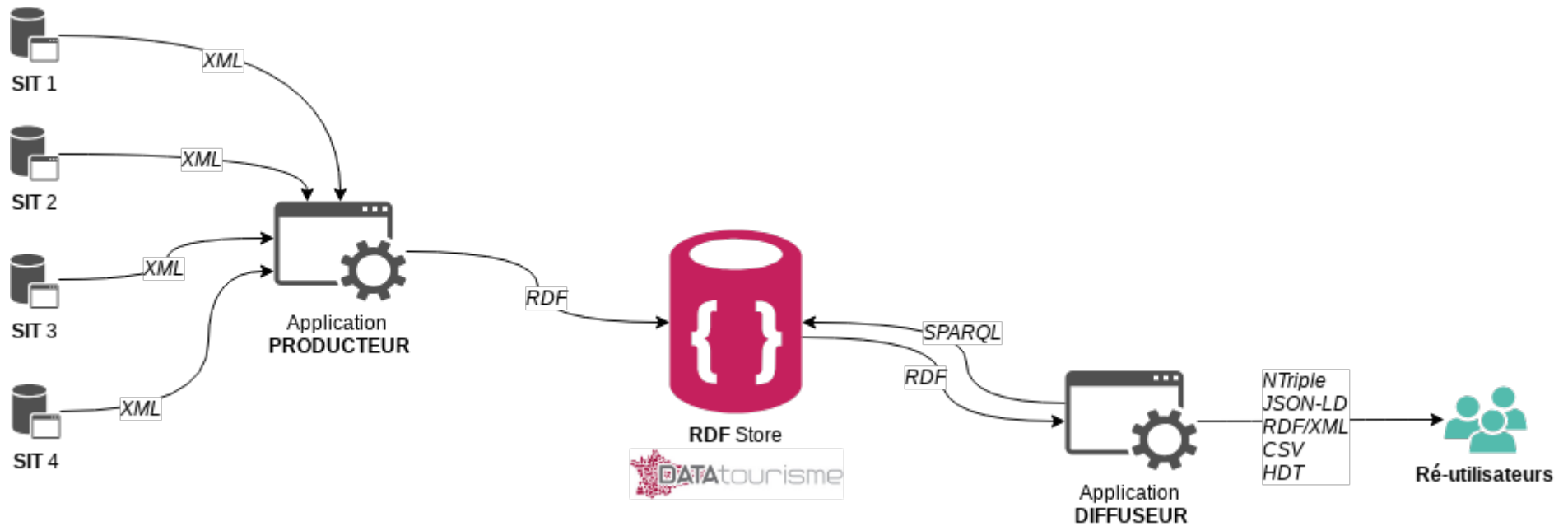
DATAtourisme

- Projet français de mise à disposition des données touristiques en **Linked Open Data**
- Initiative initialement portée par Tourisme & Territoires
- Pilotage par la DGE dans le cadre du **Programme d'investissement d'avenir**
- Lancé depuis le **12 décembre 2017**

DATAtourisme

- **Agrégation** des données
depuis l'ensemble des SIT
- **Transformation** dans un format commun
grâce à l'ontologie nationale
- **Stockage**
dans un entrepôt sémantique
- **Diffusion** en open-data
au travers de l'application diffuseur

DATAtourisme



Quelques chiffres (octobre 2018)

- **45** structures productrices
- **174** flux producteurs
- **244 087** points d'intérêts
- **96,99 %** de taux publication
- **15 289 319** triplets
- **9 133 242** réutilisations de POI
- **2 410** flux diffuseurs

Enjeux de la conception

- **Abstraire** les technologies du web sémantique
- Rendre **ontology driven** les interfaces
- **Optimiser** de multiple processus pour une volumétrie importante
- Imaginer une architecture **résiliente** aux erreurs

Interface d'alignement

- Permet aux producteurs d'aligner la structure **XML vers le format DATAtourisme**
- Interface **visuelle** s'appuyant sur l'ontologie DATAtourisme
- **Ontology driven** : l'interface s'adapte aux mises à jour de l'ontologie

Interface d'alignement : démo

The screenshot shows the DATAtourisme interface with the following elements:

- Header:** Includes logos for the French Republic and the Ministry of Economy and Finance, along with navigation links for 'Tableau de bord', 'Thésaurus', 'Flux', 'Réutilisations', and 'Aide'.
- Left Sidebar:** Contains 'Point d'intérêt' (617), 'Itinéraire touristique' (1), and 'Aperçu des POI'.
- Main Content Area:**
 - Point d'intérêt:** A list of properties with status icons (checkmarks or exclamation marks).
 - Identifiant (xsd:string):** Identifiant de la ressource. (Green checkmark)
 - Nom (rdf:langString):** Etiquette courte décrivant la ressource. (Green checkmark)
 - Type (Classe):** Le type de la ressource. (Green checkmark)
 - est localisé à (Lieu):** La localisation du POI, et donc le lieu où il peut être potentiellement consommé. (Blue checkmark)
 - a pour description courte ou longue (Description):** Description textuelle courte ou longue du POI pouvant être associée à une audience. (Green checkmark)
 - Date de dernière mise à jour (aaaa-mm-jj) (xsd:date):** (Yellow exclamation mark)
 - a été publié par (Agent):** L'agent qui a publié ce POI dans le système d'information. (Yellow exclamation mark)
 - a pour contact (Agent):** L'agent à contacter pour affaires générales relatives à ce POI. (Yellow exclamation mark)
 - a été créé par (Agent):** L'agent qui a créé ce POI dans le système d'information. (Yellow exclamation mark)
 - a comme conditions de vente (Prestation):** Moyens de paiement et politique tarifaire (A la journée, Par personne, Tarifs..). (Yellow exclamation mark)
 - Lieu:** A list of properties with status icons (exclamation marks or question marks).
 - a pour coordonnées géographiques (Coordonnées géographiques, Forme géographique):** Les coordonnées géographiques de la ressource. (Yellow exclamation mark)
 - a pour adresse postale (Adresse postale):** L'adresse postale du lieu concerné. (Yellow exclamation mark)
 - a pour horaires d'ouverture (schema:OpeningHoursSpecification):** Les heures d'ouverture d'un lieu/service/produit. (Yellow exclamation mark)
 - Code INSEE fictif (xsd:string):** Cette propriété permet de stocker des codes INSEE fictifs, c'est à dire des codes INSEE qui auraient été créés par des systèmes d'information. Ex: faire d'une station de ski une ville. (Yellow exclamation mark)
 - Accès internet (xsd:boolean):** Vrai si le produit propose un accès à internet. (Yellow exclamation mark)
 - Air conditionné (xsd:boolean):** Vrai si cette surface propose l'air conditionné. (Question mark)
 - Non fumeur (xsd:boolean):** Vrai si cette surface est non fumeur. (Question mark)
 - Animaux acceptés (xsd:boolean):** (Question mark)

Éditeur de requête

- Permet aux diffuseur de concevoir un flux sémantique **sans connaître SPARQL**
- Interface **visuelle** s'appuyant sur l'ontologie DATAtourisme
- **Ontology driven** : l'interface s'adapte aux mises à jour de l'ontologie

Éditeur de requête : démo

Types de point d'intérêt

Tous les types

Filtres

Type

Meublés de tourisme

POI > Type

fait partie du département

Aveyron

POI > est localisé à > a pour adresse postale > a comme l...

+ Ajouter un filtre

Liste Carte SPARQL

1 670 résultats

La prévisualisation ne contient que les **10 premiers résultats** de votre requête et ignore le paramétrage des langues
Vous retrouverez l'ensemble des résultats dans le fichier mis à disposition par le flux, accessible par téléchargement manuel ou par webservice

Christian et Josiane Cayla - Gîte La Gineste 12320 Conques-en-Rouergue

Gîte de caractère avec piscine privée, sur un terrain de 2000 m² en partie arboré, à l'écart de routes fréquentées et à proximité de nombreux sentiers de randonnées.

Gîte La Petite Ecole 12260 Balaguier-d'Olt

Gîte adapté aux locataires avec chien (s)

Andrée VILLENEUVE - AYG8023 12380 Saint-Sernin-sur-Rance

Gîte plain pied, dans maison comprenant 2 autres gîtes (entrée commune), à proximité du propriétaire et 4 chambres d'hôtes, dans un bourg. 1 chambre (1 lit 140), salle de bains, kitchenette, chauffage complet électrique, salon de jardin, barbecue, draps et linge de maison fournis. Commun : lave et sèche-linge, jardin en terrasses (250 m²) non clos, portique, piscine clôturée (6 x 12).

Les Terrasses de Labade (Gîte) 12190 Coubisou

Située dans un hameau, cette ancienne grange restaurée vous offre calme et détente dans un cadre verdoyant.

ROBERT Frédéric 12160 Baraqueville

Le gîte est situé à 15 minutes de Rodez, dans un jolie hameau au calme.

DEVOILLE Pascal et SEGONDS Jeannine 12470 Prades-d'Aubrac

En Aveyron, sur les montagnes d'Aubrac, en pleine nature, cette ancienne ferme a été restaurée avec des matériaux écologiques. Jeanine et Pascal, artisans en constructions et rénovations écologiques, ont aménagé ce gîte ouvert toute l'année, ainsi qu'une cabane perchée dans les châtaigniers. Ils animent pratiquent une agriculture vivrière en bio, et leur grande maison accueille des enfants pour des séjours à thème.

Cassanodres 12440 Salvagnac-Cajarc

L' Hacienda du bonheur 12620 Saint-Beauzély

Gîte de plain pied en rez-de-chaussée labellisé tourisme et handicap (moteur, visuel, mental et auditif). Dans le Parc Naturel Régional des Grands Causses, près du Viaduc de Millau et des gorges du Tarn, ce gîte est situé au pied de Saint-Beauzély. Depuis la terrasse, bel espace de détente (vue sur le village), on perçoit le clapotis de La Muzette qui longe le jardin. Un intérieur plaisant : murs aux teintes chaudes et colorées, de la luminosité, de l'espace ... un tout très agréable.

Utilisation des données

- Les données sémantiques peuvent être chargées dans un triple store
- Une [API PHP GraphQL](#) pour s'interfacer avec le triple store
- Un [environnement docker](#) incluant :
 - Blazegraph (triple store)
 - API GraphQL en serveur
 - Interface de requête GraphiQL

API GraphQL

- GraphQL : langage de requêtes développé par Facebook
- Permet d'**interroger** un triplestore grâce à **GraphQL**
- **Arguments** et **Filtres** pour des requêtes avancées
- **Ontology driven** : le schéma GraphQL s'adapte aux mises à jour de l'ontologie
- Documentation : <https://datatourisme.frama.io/api/>

API GraphQL : démo

GraphiQL ▶ ◀ Docs

```
1 # Welcome to GraphiQL
2 #
3 # GraphiQL is an in-browser IDE for writing, validating, and
4 # testing GraphQL queries.
5 #
6 # Type queries into this side of the screen, and you will
7 # see intelligent typeaheads aware of the current GraphQL type schema and
8 # live syntax and validation errors highlighted within the text.
9 #
10 # To bring up the auto-complete at any point, just press Ctrl-Space.
11 #
12 # Press the run button above, or Cmd-Enter to execute the query, and the result
13 # will appear in the pane to the right.
14
15 query { poi {
16   total
17   results {
18     _uri
19     rdfs_label {
20       value
21       lang
22     }
23     isLocatedAt {
24       schema_address {
25         _uri
26         rdf_type
27         schema_addressLocality
28         schema_postalCode
29         schema_streetAddress
30         schema_postOfficeBoxNumber
31         cedex
32       }
33     }
34   }
35 }}
```

```
{
  "data": {
    "poi": {
      "total": 41835,
      "results": [
        {
          "_uri": "https://data.datatourisme.gouv.fr/10/000283ba-f94b-3bce-8f57-e5c10bec6cd4",
          "rdfs_label": [
            {
              "value": "VILLA GALLO ROMAINE DU GROSSWALD",
              "lang": "fr"
            }
          ],
          "isLocatedAt": [
            {
              "schema_address": [
                {
                  "_uri": "https://data.datatourisme.gouv.fr/00b41f8c-4bda-3474-afc3-2f32325687a1",
                  "rdf_type": [
                    "http://schema.org/PostalAddress",
                    "http://www.datatourisme.fr/ontology/core/1.0#PostalAddress"
                  ],
                  "schema_addressLocality": [
                    "Sarreinsming"
                  ],
                  "schema_postalCode": [
                    "57905"
                  ],
                  "schema_streetAddress": [
                    "Forêt du Grosswald"
                  ],
                  "schema_postOfficeBoxNumber": null,
                  "cedex": null
                }
              ]
            }
          ]
        }
      ]
    },
    {
      "_uri": "https://data.datatourisme.gouv.fr/10/000ce9fa-eaca-3d78-a395-8798b5965cbc",
      "rdfs_label": [
```

QUERY VARIABLES

Optimisation des processus

- Contrainte temporelle forte due aux données chaudes → 24h pour tout traiter.
- **Décomposition** des requêtes de mises à jour du triplestore, notamment pour l'évolution de ressources existantes.
- Utilisation massive des **query hints** et de l'**index géospatial** de Blazegraph

Optimisation des processus

- Gestion des volumes de données sérialisées **en stream**, avec pré-chargement dans un key-value store
- Utilisation des **jena rules** pour :
 - Valider les Ressources
 - Pré-calculer l'inférence métier
 - Pré-calculer l'inférence d'instance

Merci !