

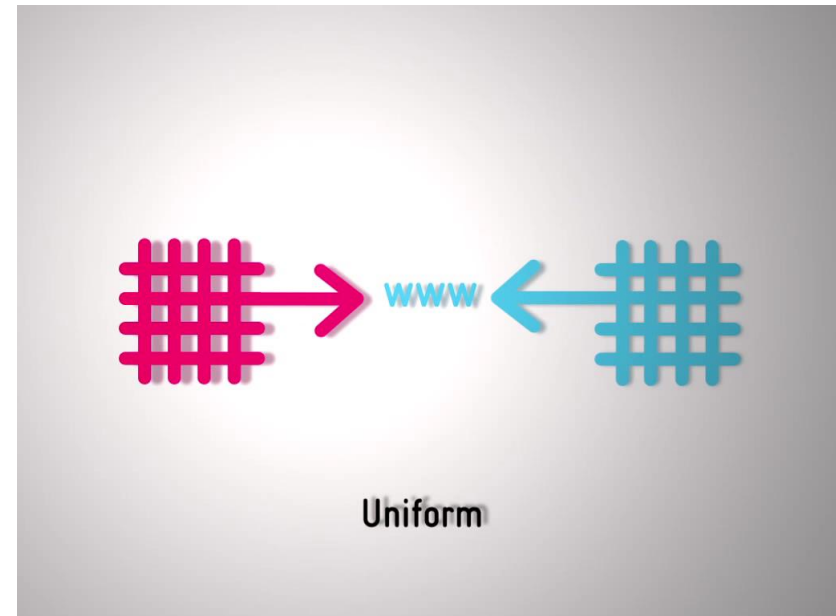


JDev – Atelier Datalift

RDF, Linked (Open) Data & Datalift

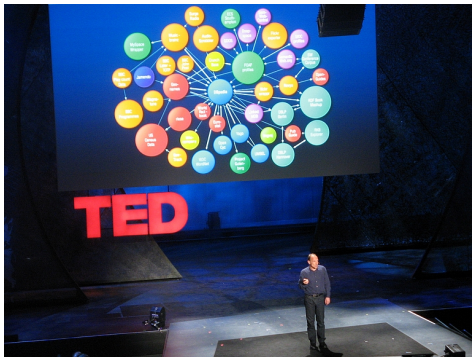
Le Linked Data

- ▶ Publier des **données structurées** sur le Web en les **reliant** entre elles pour constituer un réseau **global** d'informations.
- ▶ En finir avec les silos de données isolés.
- ▶ S'appuyer sur les **standards** du Web.
- ▶ **Partager** l'information entre **machines**.
- ▶ Faciliter la navigation par les êtres **humains**.



<http://damfoundation.org/2012/02/european-as-introduction-to-linked-open-data/>

Le Linked Data ou Web des données

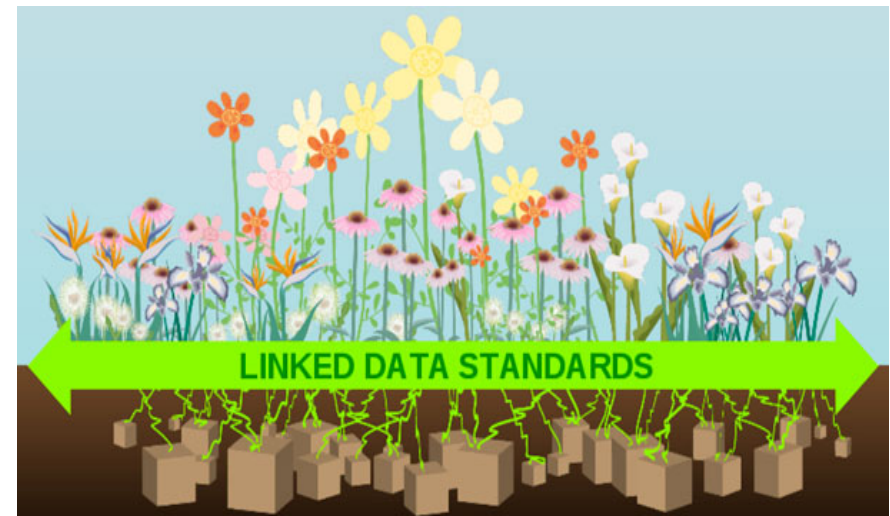


Tim Berners-Lee, « On the next Web »,
(conférence TED, février 2009)

« Raw Data now! »

Les 4 principes du *Linked Data* :

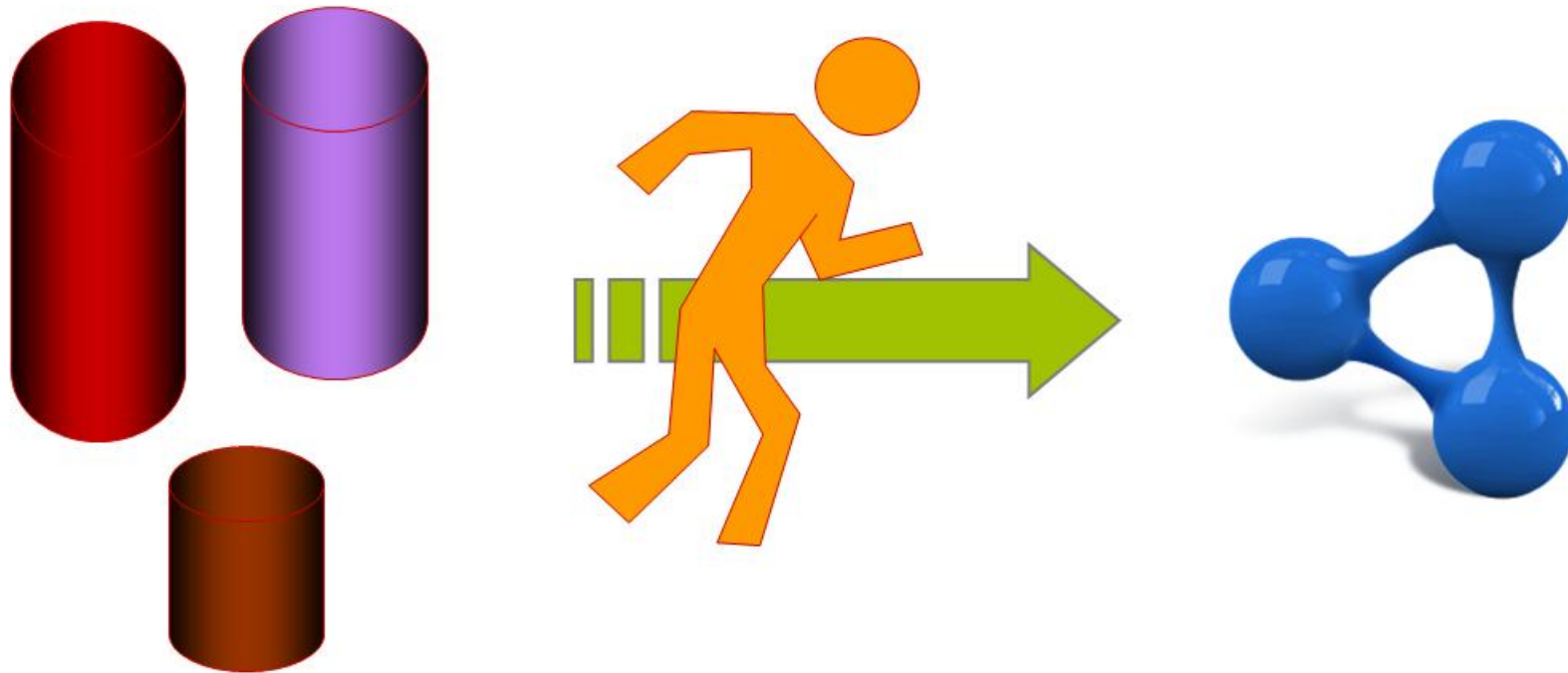
- ❖ Utiliser des URI
- ❖ Utiliser des URI accessibles via HTTP
- ❖ Donner accès aux données en utilisant les standards RDF et SPARQL
- ❖ Exprimer les URI des objets liés



Source : <http://www.w3.org/2009/Talks/0204-ted-tbl/#%281%29>

Open Data → Linked Open Data (LOD)

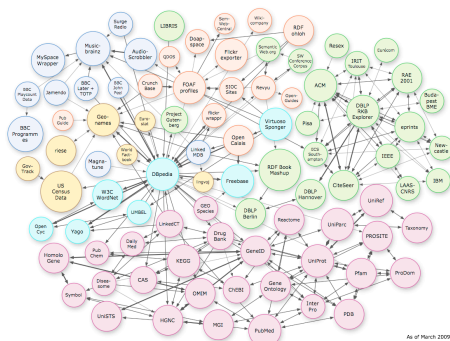
C'est un nouveau monde de données pour les données ouvertes



Formats propriétaires / hétérogènes

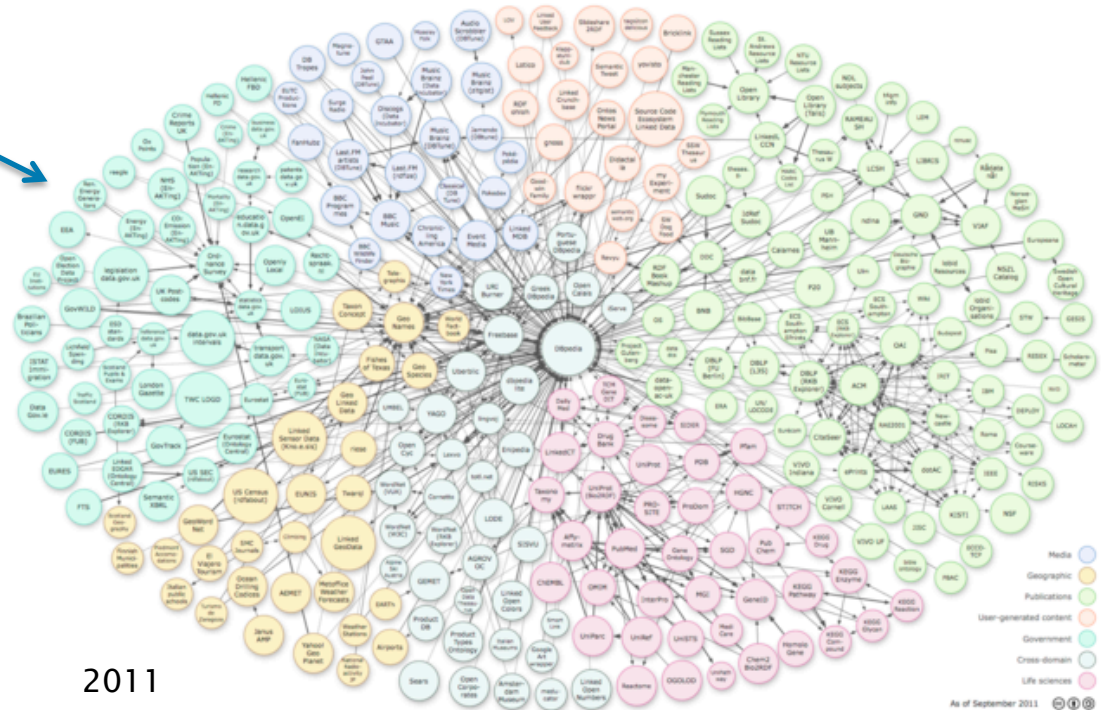
RDF/ SPARQL

The LOD Cloud



2009

As of March 2009



2011

As of September 2011

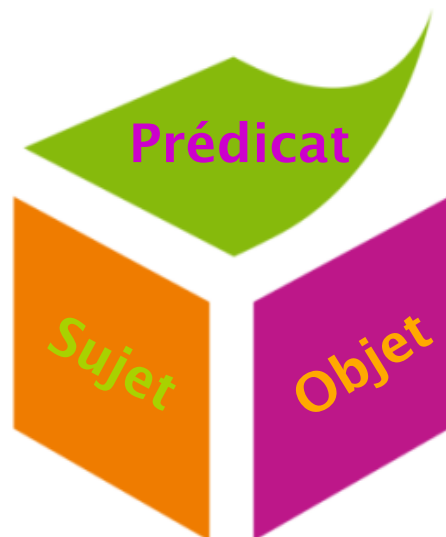
Source : <http://lod-cloud.net/>

JDev - 6 septembre 2013
Élévation de données avec Datalift



RDF - Le triplet, la molécule de base

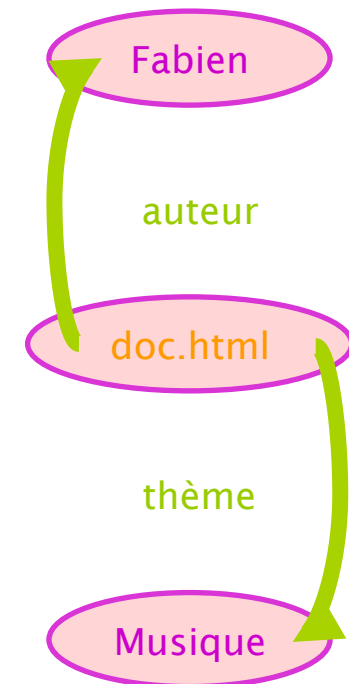
RDF Toute connaissance peut être ramenée à un ensemble équivalent de triplets (**sujet**, **prédicat**, **objet**).



doc.html a pour auteur Fabien et a pour thème la Musique



*doc.html a pour auteur Fabien
doc.html a pour thème Musique*



RDF – Un modèle abstrait

RDF signifie :

Resource *Pages, images, vidéos, ...
(tout ce qui peut avoir une URI)*

Description *Attributs, fonctions et relations
des ressources*

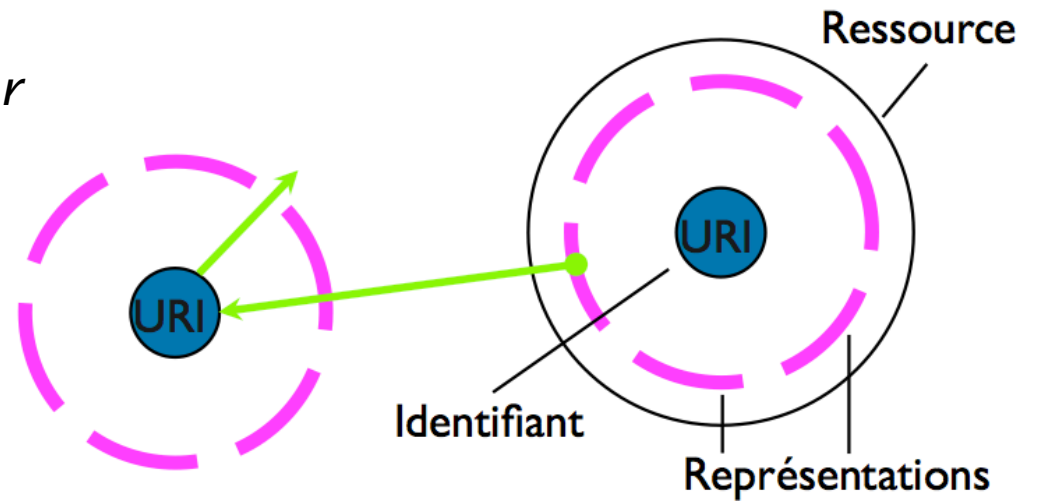
Framework *Modèle abstrait de ces
descriptions*

URI - Une des bases de l'architecture du Web

URI : *Universal Resource Identifier*

- Pérenne
- Maîtrisable
- Extensible

→ **URL** : *Universal Resource Locator*



```
foo://example.com:8042/over/there?name=ferret#nose
  \  /      \  /      / \  /      \  /      \  /
  |         |         |         |         |
scheme  authority  path      query      fragment
  |
  / \ /
urn:example:animal:ferret:nose
```

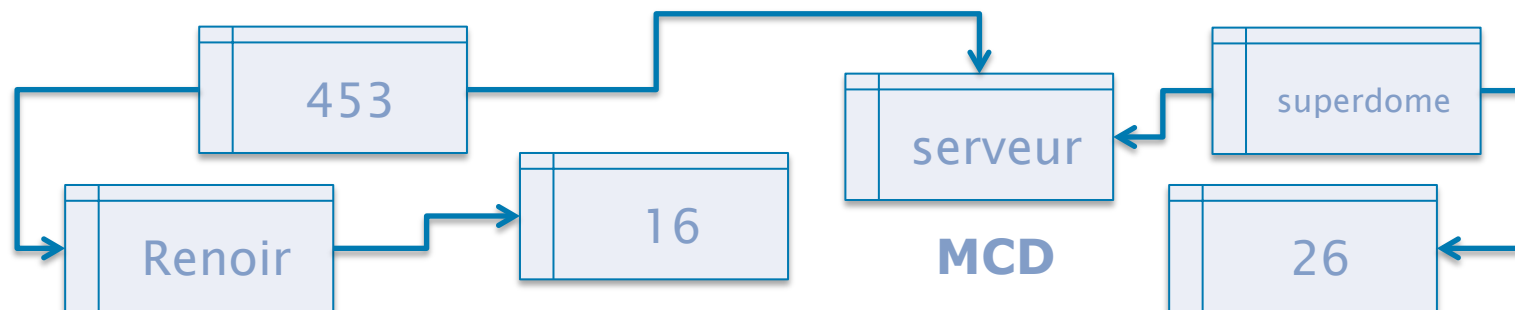
RDF - Qu'est-ce que l'élévation de données ?

► **Exemple d'élévation de données dans un datacenter... Les faits**

- Les faits : Le superdome 453 qui est dans de la salle Renoir dont la température est de 16°C et la climatisation en marche, est un serveur HP dont le constructeur garantit le bon fonctionnement jusqu'à 26°C.

► **Leur réécriture**

- Le superdome 453 est dans la salle Renoir
- La salle Renoir est à la température de 16°C
- La climatisation de la salle Renoir est en marche
- Le superdome 453 est un serveur HP
- HP garantit le bon fonctionnement des superdomes jusqu'à 26°C

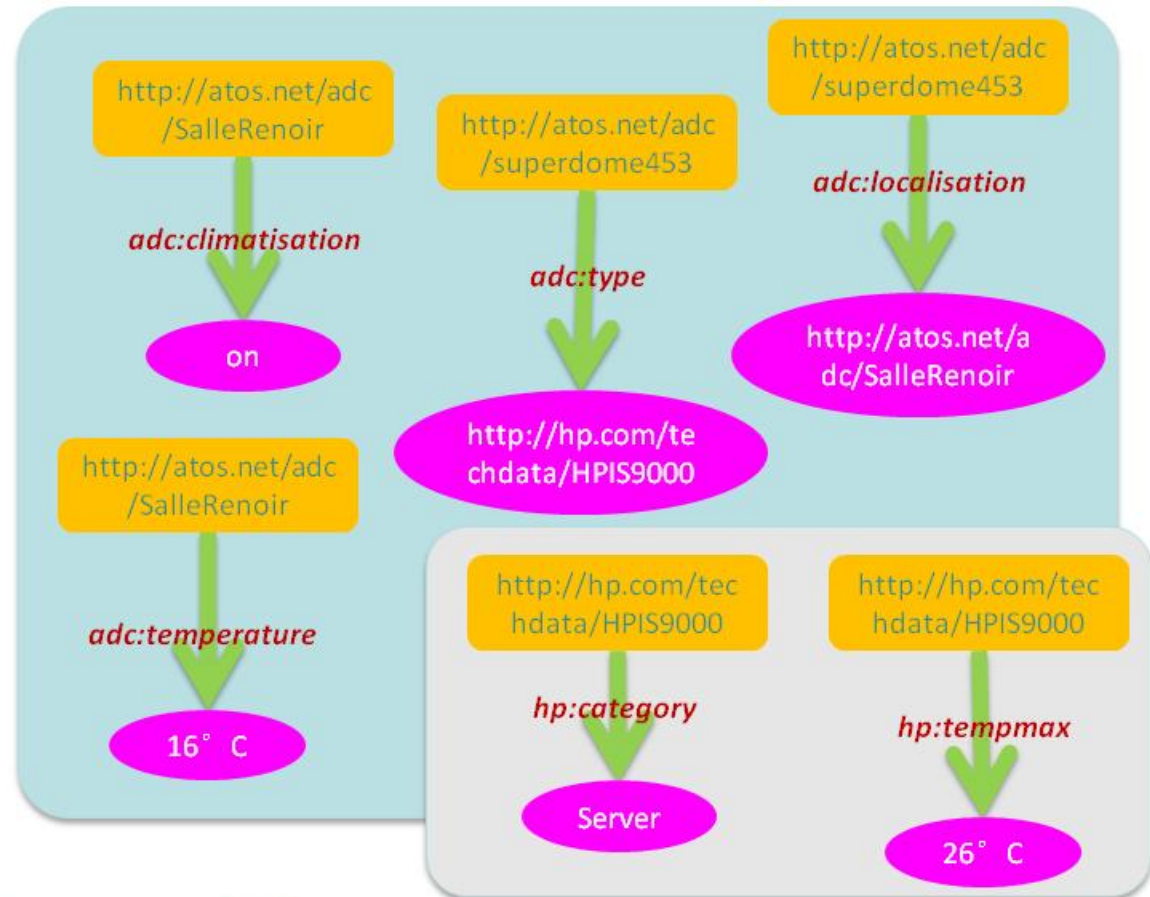


RDF - Qu'est-ce que l'élévation de données ?

URI : identifier
 Triplet : exprimer
 Graphe : relier

Vocabulaire = Ontologie

- ▶ Atos Data Center : **adc**
 - Type
 - Localisation
 - Température
 - Climatisation
 - Etc.
- ▶ HP : **hp**
 - Tempmax
 - Category
 - Etc.



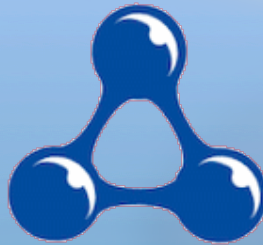
Sujet	Prédicat	Objet
(Ressource)	(Ressource)	(Ressource / Littéral)

RDF - Qu'est-ce que l'interconnexion ?

Latitude
Paris
48°51' 24" N



Longitude
Paris
2°21' 07" E

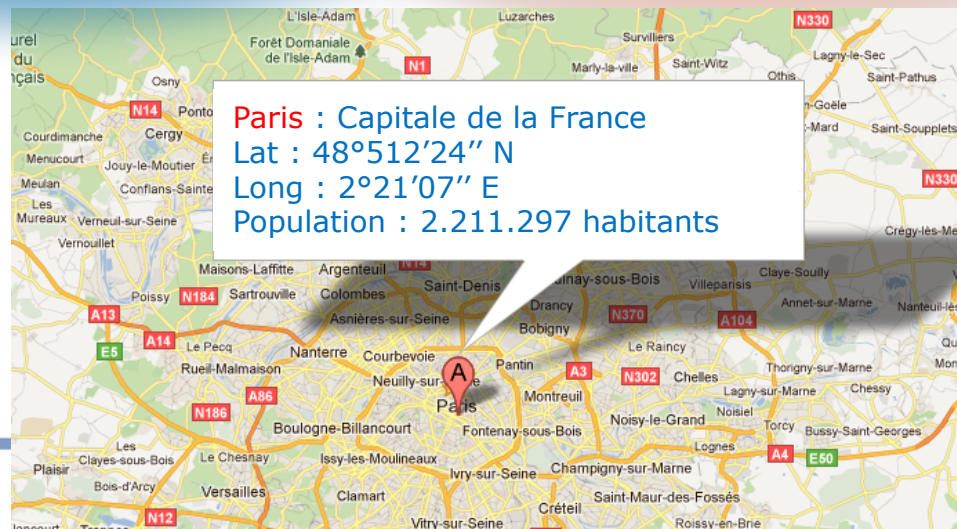


Données géographiques

Données statistiques
Population
Paris
2.211.297



Les données s'enrichissent mutuellement au bénéfice du lecteur.



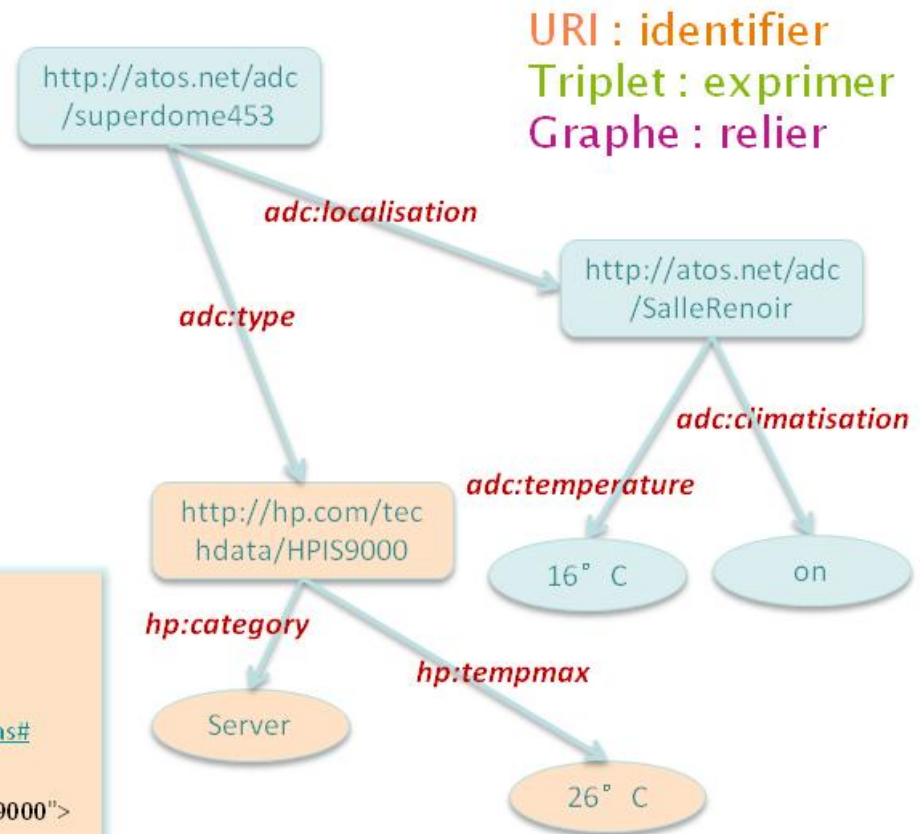
RDF - Qu'est-ce que l'interconnexion ?

RDF/XML – Données chez AtoS

- ▶ <?xml version="1.0"?>
- ▶ <rdf:RDF
- ▶ xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:adc="http://atos.net/metadata#" xmlns:hp="http://hp.com/techdata#">
- ▶ <Description about="http://atos.net/adc/superdome453">
- ▶ <adc:localisation>SalleRenoir</adc:localisation>
- ▶ <adc:type>http://hp.com/techdata/HPIS9000</adc:type>
- ▶ </Description>
- ▶ <Description about="http://atos.net/adc/SalleRenoir">
- ▶ <adc:temperature>16° C</adc:temperature>
- ▶ <adc:climatisation>on</adc:climatisation>
- ▶ </Description>
- ▶ </rdf:RDF>

RDF/XML – Données chez HP

- ▶ <?xml version="1.0"?>
- ▶ <rdf:RDF
- ▶ xmlns=http://www.w3.org/1999/02/22-rdf-syntax-ns# xmlns:hp="http://hp.com/techdata#">
- ▶ <Description about="http://hp.com/techdata/HPIS9000">
- ▶ <hp:category>server</hp:category>
- ▶ <hp:tempmax>26° C</hp:tempmax>
- ▶ </Description>
- ▶ </rdf:RDF>



URI : identifier
Triplet : exprimer
Graphe : relier

Sujet	Prédictat	Objet
(Ressource)	(Ressource)	(Ressource / Littéral)

RDF – Les triple stores

- ▶ **Bases de données spécialisées RDF**
 - Pas de schéma de données prédéfini ou imposé
 - Partitionnement des données en « graphes nommés » (entités administratives)
 - Support de l'inférence de données
- ▶ **Natives**
 - Sesame, OWLIM, *Virtuoso*, AllegroGraph, BigData, Jena (Fuseki / TDB)...
- ▶ **Ou basées sur des bases de données relationnelles**
 - Sesame, Oracle, Jena (SQL DB)...
- ▶ **Scalables**
 - Plusieurs milliards de triplets par nœud
- ▶ **Langage de requêtage & manipulation évolué : SPARQL**

SPARQL – Un langage et un protocole

« SPARQL Protocol And RDF Query Language »

- ▶ Langage défini par le W3C
 - SPARQL 1.0 (2008) : lecture seule
 - SPARQL 1.1 (2013) : modification des données
- ▶ Points d'accès (endpoints) normalisés
 - HTTP (REST)
 - SOAP
 - Fédération
 - Graph Access Protocol
- ▶ Extensions
 - GeoSPARQL

The screenshot shows a web-based SPARQL query editor. At the top, it says "Editeur de requête SPARQL". Below that, there are tabs for "Données publiées" and "Format de la réponse" with options: HTML, RDF/XML, N3/Turtle, N-Triples, TriG, TriX, and CSV. The "Requête" field contains the following SPARQL query:

```
SELECT ?auteur ?serie ?resume WHERE
{
  ?s1 glenat:serie ?serie ;           # Titre de la série dans le catalogue Glénat
    glenat:Auteur ?auteur .         # Auteur de la série
  ?s2 rdfs:label ?nom ;             # Titre dans DBPedia
    dbp:abstract ?resume .         # Résumé dans DBPedia
  FILTER(STR(?serie) = STR(?nom))   # Séries de même titre
  FILTER(LANGMATCHES(LANG(?nom), "fr")) # Titres en français uniquement
  FILTER(LANGMATCHES(LANG(?resume), "fr")) # Résumés en français uniquement
}
```

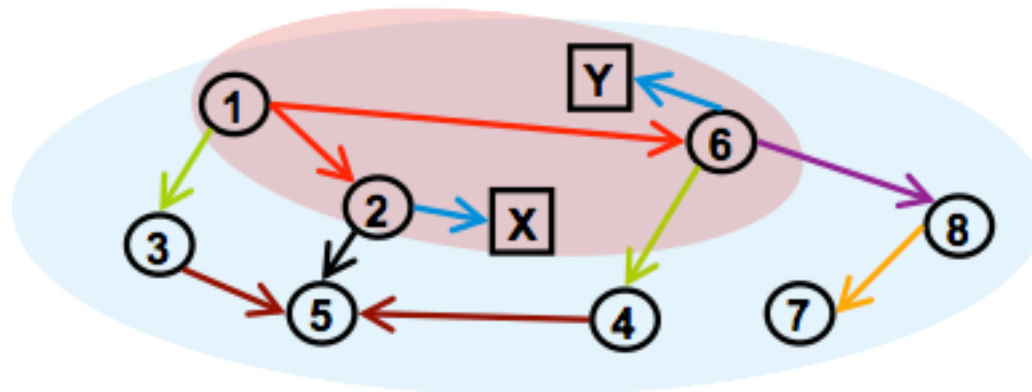
Below the query, there are buttons for "Requêtes prédéfinies" (Toutes les entrées, Sujets, Types, Contextes, Exemple 1) and a "Nombre maximum de résultats" field set to 5000. An "Exécuter la requête" button is also present.

The results are displayed in a table with columns "s", "p", and "o". The table contains 10 rows of data, including URIs and literals.

s	p	o
http://localhost:9091/paris/kiosques/1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://localhost:9091/paris/kiosques#kiosques-ouverts-a-
http://localhost:9091/paris/kiosques/1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2006/vcard/ns#VCard
http://localhost:9091/paris/kiosques/1	http://localhost:9091/paris/kiosques#adresse	"PCE DU CHATELET"
http://localhost:9091/paris/kiosques/1	http://localhost:9091/paris/kiosques#arrdt	"75001"
http://localhost:9091/paris/kiosques/1	http://localhost:9091/paris/kiosques#ouv	"08:30:00.0000^^"
http://localhost:9091/paris/kiosques/1	http://localhost:9091/paris/kiosques#ferm	"20:00:00.0000^^"
http://localhost:9091/paris/kiosques/1	http://localhost:9091/paris/kiosques#ouv-dim	true
http://localhost:9091/paris/kiosques/1	http://www.w3.org/2006/vcard/ns#adr	_:node17e90b9uox1
http://localhost:9091/paris/kiosques/1	http://www.w3.org/2006/vcard/ns#fn	"Kiosque PCE DU CHATELET"
http://localhost:9091/paris/kiosques/1	http://www.w3.org/2006/vcard/ns#geo	_:node17e90b9uox1

SPARQL – Requêtage sur graphes

Soit le graphe suivant enregistré dans une base de données RDF (triple store) :



SPARQL permet d'extraire un sous-ensemble de ce graphe par expression de contraintes sous la forme d'équations

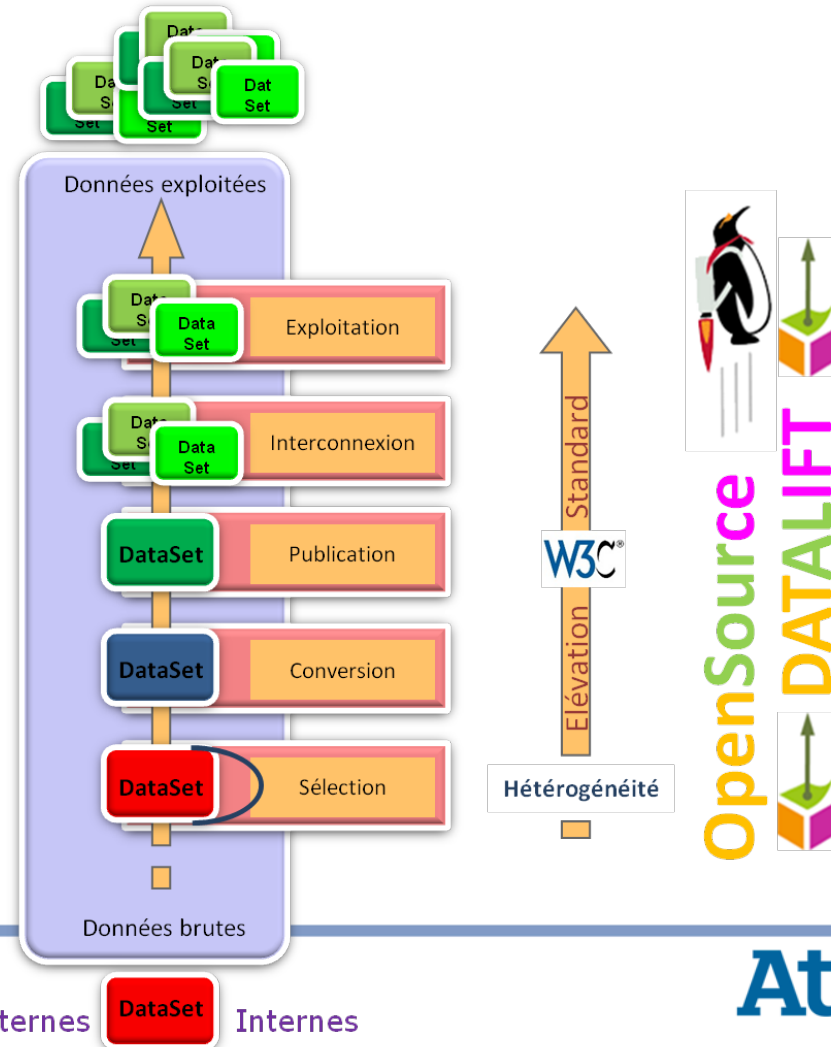
Exemple :

Je cherche les ressources liées à 1 par prédicat « rouge » et les chaînes de caractères liées à ces ressources par le prédicat « bleu » :

1. Les ressources liées à 1 par le prédicat « rouge » : `<1> <rouge> ?resources`
2. Les chaînes liées à ces ressources par le prédicat « bleu » : `?resources <bleu> ?string`

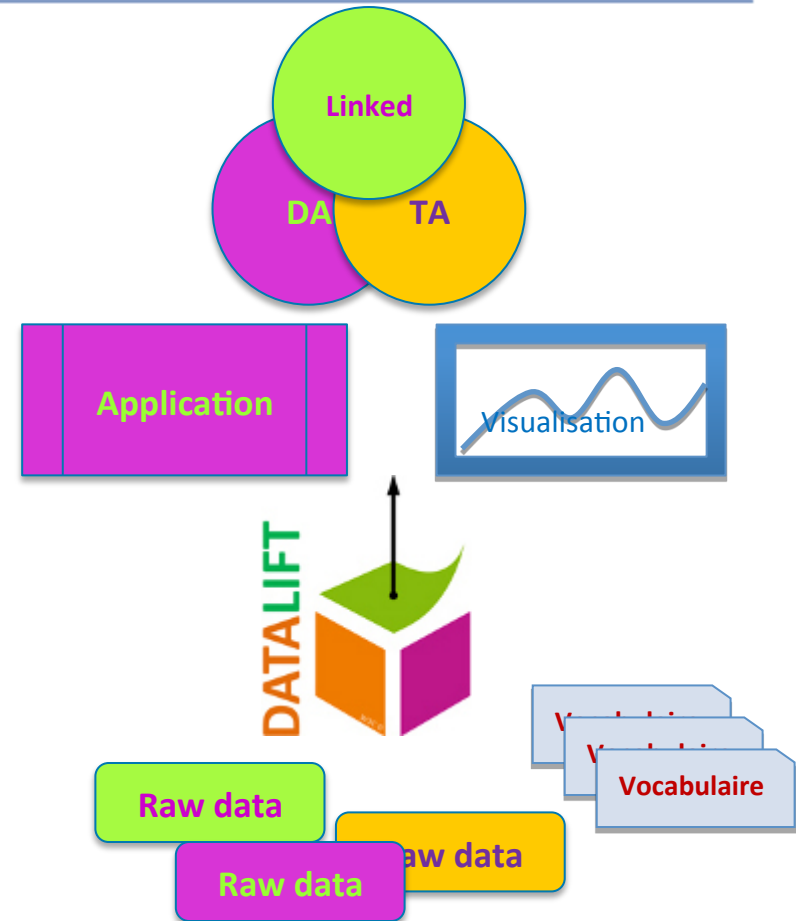
Datalift - L'élévateur de données

Elever la donnée
=
Passer
de la donnée brute
à la donnée sémantique
interconnectée



Datalift – La technologie en quelques mots

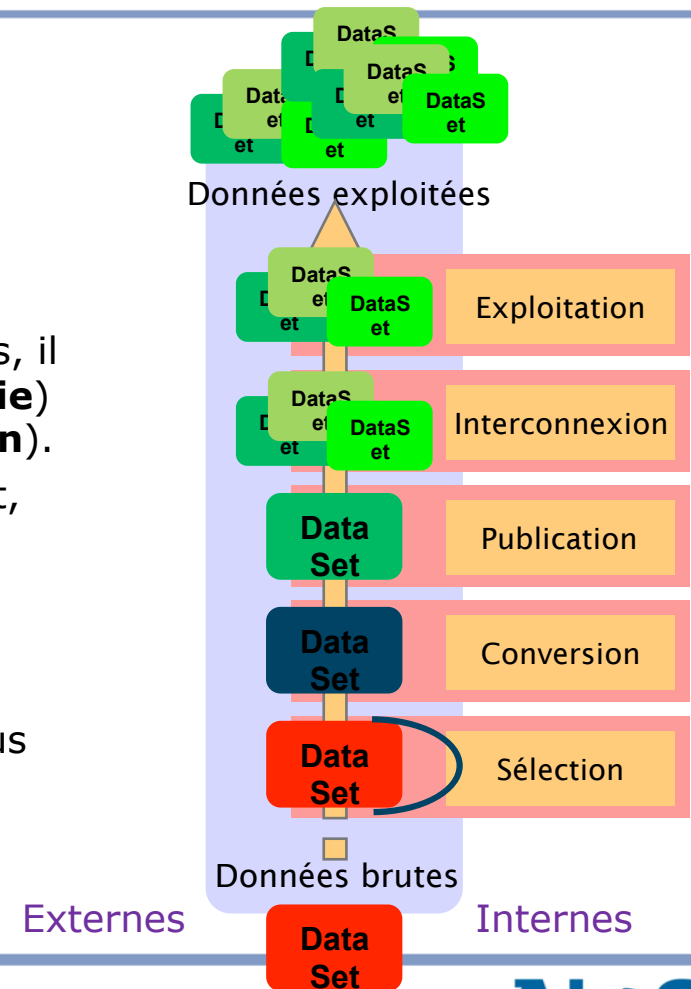
- ▶ Datalift est une plateforme **open source**. Elle intègre de nombreux modules et est conçue pour le passage à l'échelle (virtualisation, cloud).
- ▶ L'utilisateur gère des **processus d'élévation** et les exécute. Un processus prend en entrée des jeux de données brutes et hétérogènes. Des métadonnées, ajoutées à l'aide de vocabulaires, convertissent les sources. Puis les liens entre les concepts des vocabulaires sollicités permettent d'enrichir les données.
- ▶ **L'exploitation des données** résultantes peut prendre la forme
 - d'applications utilisant des **web services** d'accès aux données,
 - de production de **nouveaux jeux de données interconnectés et sémantisés**



Datalift - Cinématique

L'élévation et l'interconnexion de jeux de données

- Le dispositif prend en compte des données brutes de toutes provenances, formats et structures.
- Ces données sont dans des blocs statiques, des datasets ou jeux de données.
- Pour les transformer en données du web des données, il faut déterminer le vocabulaire (**sélection d'ontologie**) qui va permettre leur expression en clair (**conversion**).
- Les données converties ont la forme de triplets (sujet, prédicat, objet), ce sont désormais des données **publiables** dans le web des données.
- Les données des blocs convertis sont ensuite **interconnectées** avec les données déjà publiées.
- Cet enrichissement est la dernière phase du processus d'élévation de données.
- La valeur des données résultantes est augmentée et leur **exploitation** pourra en profiter.



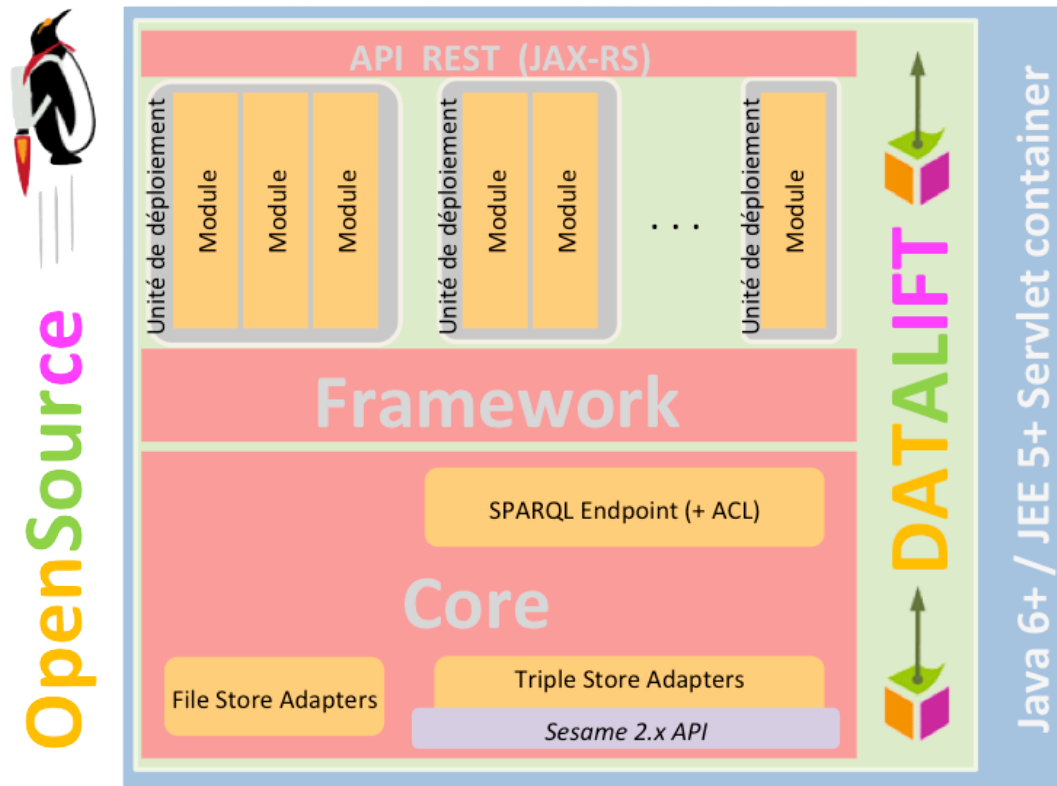
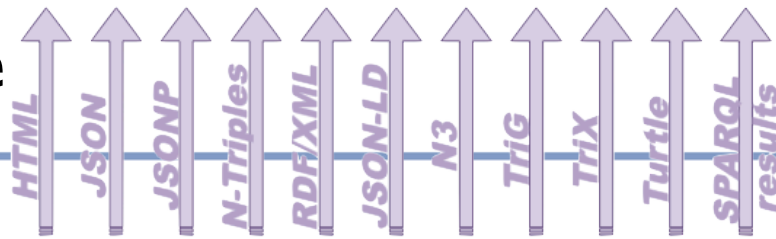
Objectif - Des données 5 étoiles

- ▶ 2010 : Lors de la « Gov 2.0 Expo », Tim Berners Lee a présenté une échelle d'ouverture des données devenue célèbre et qui fait référence
- ▶ ★ Mettre vos données sur le web (dans n'importe quel format) ... et penser à expliciter la licence
- ▶ ★★ Mettre vos données dans un format structuré (ex. une feuille Excel plutôt qu'une image scannée d'un tableau)
- ▶ ★★★ Préférer un format non-propriétaire (ex. CSV plutôt que Excel)
- ▶ ★★★★ Utiliser des URL pour identifier les choses pour que d'autres personnes puissent pointer vers vos données
- ▶ ★★★★★ Lier vos données avec d'autres données pour mieux contextualiser vos données



<http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>

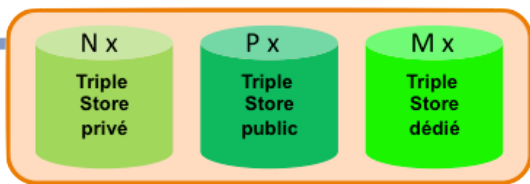
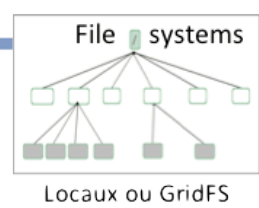
Datalift - Architecture



OpenSource

DATALIFT

Java 6+ / JEE 5+ Servlet container



- Entrepôts RDF :
- Sesame 2.x,
 - OWLIM,
 - Virtuoso,
 - AllegroGraph,
 - BigData

JDev - 6 septembre 2013
Élévation de données avec Datalift

Datalift - Installation

- ▶ **Serveur d'applications JEE (Tomcat, JBoss...)**
 - Fichier WAR (Web ARchive) + environnement de travail
 - Triple stores locaux ou distants
 - Multi-utilisateur (contrôle d'accès)

- ▶ **Wrapper**
 - Script de lancement : Jetty + Datalift + Sesame (+ *quelques modules*)
 - Mono-utilisateur
 - Un environnement par utilisateur système :
 - Windows : `%APPDATA%\Datalift`
 - Mac OS X : `~/Library/Application Support/Datalift` (sauf caches & logs)
 - Linux : `~/.datalift`

Datalift – Environnement de travail (Wrapper)

▼ Datalift	
▼ conf	
datlift-application.properties	← Configuration Datalift
datlift-log4j.xml	← Configuration des traces
licenses.ttl	
predefined-queries.ttl	
s4ac-policies.ttl	
▼ lib	← Bibliothèques partagées (drivers)
postgresql-9.2-1002.jdbc4.jar	
▼ modules	← Modules propres à l'utilisateur
donuts.jar	
▼ repositories	
▶ OpenRDF Sesame	← Bases de données RDF Sesame
▼ storage	← Stockage fichiers interne,
▼ public	← Public, et
▶ project	← Fichiers sources des projets
▼ webapps	← Applications web propres à
openrdf-workbench.war	l'utilisateur (e.g. admin. Sesame)

Datalift – Le consortium



Ateliers & cas d'utilisation

▶ **Élévation de données avec Datalift**

- Kiosques de la ville de Paris : conversion, transformation et consommation

▶ **Croisement de données en SPARQL**

- Catalogues d'étoiles (Datalift)
- Intégration de données hétérogènes (EEA)

▶ **Linked Open Data**

- INSEE : données ouvertes
- IGN : publication de données « semi-ouvertes »

Atelier 1 – Élévation de données avec Datalift

- ▶ Données Open Data classiques : Kiosques de la ville de Paris
 - Format : CSV
- ▶ Conversion en RDF
 - Transformation brute (« *direct mapping* »)
- ▶ Transformation vers un vocabulaire standard
 - Ontologie vCard (W3C)
- ▶ Enrichissement de données (démonstration)
 - Données GPS (Google)
- ▶ Publication
- ▶ Consommation
 - Requêtage SPARQL
 - Fusion Google Maps (démonstration)

Atelier 1 – Elévation de données avec Datalift

▶ Etape 1 : Récupération des données

<http://opendata.paris.fr/opendata/jsp/site/Portal.jsp>

The screenshot shows the PARISDATA website interface. At the top, there's a navigation bar with links: 'Les données', 'La licence', 'La démarche', 'Le forum', and 'XP V2 Bêta'. Below this is a search bar with the text 'kiosque' and a 'Valider' button. The search results are displayed in a table with four entries, each with a title and a URL. A blue circle with the number '1' is overlaid on the bottom left of the screenshot.

Titre	Date
Liste des kiosques presse, théâtre et Loto	26/09/12
Liste des kiosques presse, théâtre et Loto	26/09/12
Liste des kiosques presse à Paris	26/09/12
Liste des kiosques presse à Paris	26/09/12

The screenshot shows the dataset page for 'Liste des kiosques presse à Paris'. It features a yellow 'Culture' tag. The description states: 'Liste des kiosques Parisiens en activité avec les horaires d'ouverture et fermeture. Liste établie par Médiakiosk.' Under 'Exemples de réutilisations:', it lists '- Kiosques à Paris, www.kiosque-paris.fr'. A blue circle with the number '2' is overlaid on the right side of the page.

This screenshot shows the details and download options for the dataset. It includes the following information:

- Thème : Culture
- Date de publication : 06/04/2011
- Fréquence de mise à jour : annuelle
- Période couverte par le jeu de données : Le fichier courant contient les dernières mises à jour
- Download options: CSV, RESTful Webservice
- Rating: 5 stars (Votez: ★★★★★)
- Stats: téléchargement(s) : 2812 | vue(s) : 4039 | vote(s) : 59
- Warning: En cochant cette case j'accepte les CGU et la licence pour télécharger le fichier
- Link: [Consulter les C.G.U. et la Licence]

A blue circle with the number '3' is overlaid on the right side of the page.

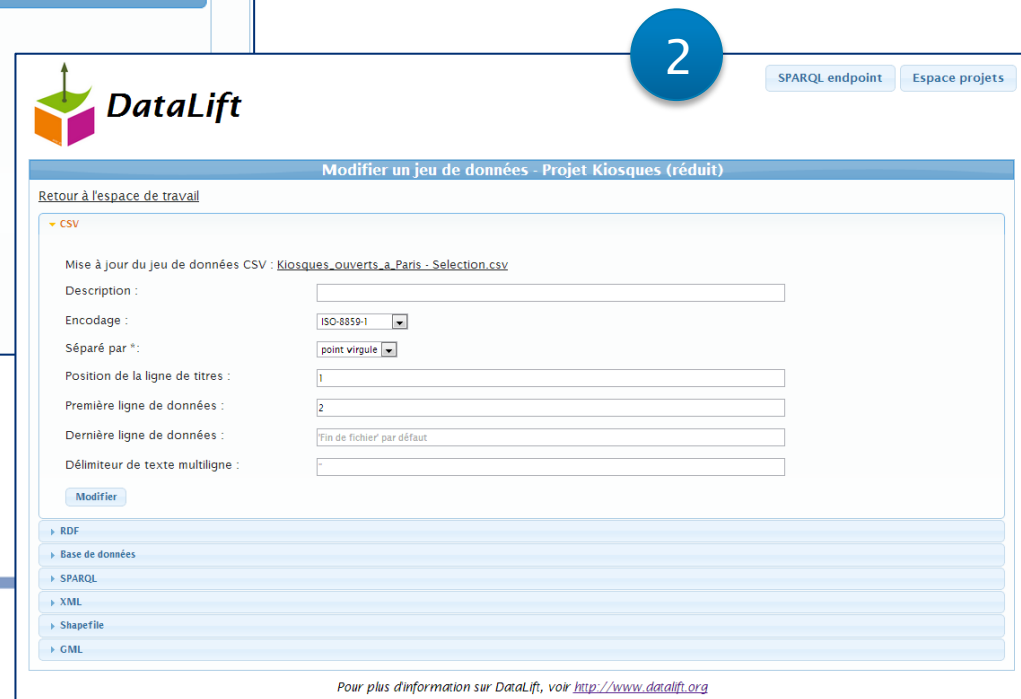
<http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/source/kiosques-ouverts-a-paris-csv>

Atelier 1 – Elévation de données avec Datalift

- ▶ Etape 2 : Import des données CSV



The screenshot shows the Datalift interface for a project named 'Projet Kiosques de Paris'. The page has a blue header with the Datalift logo and a 'SPARQL endpoint' button. Below the header, there are tabs for 'Description', 'Sources', and 'Ontologies'. The 'Description' tab is active, showing details about the project's license (Attribution 3.0 non transposé) and creation/modification dates. A 'Nouveau projet' button is visible in the bottom left corner, marked with a blue circle containing the number '1'.



The screenshot shows the 'Modifier un jeu de données' page for the 'Projet Kiosques (réduit)'. The page has a blue header with the Datalift logo and buttons for 'SPARQL endpoint' and 'Espace projets'. A blue circle with the number '2' is positioned above the page. The main content area is titled 'Retour à l'espace de travail' and contains a 'CSV' section with a dropdown menu. Below this, there are several form fields for editing the CSV file: 'Mise à jour du jeu de données CSV', 'Description', 'Encodage' (set to ISO-8859-1), 'Séparé par *' (set to point virgule), 'Position de la ligne de titres' (set to 1), 'Première ligne de données' (set to 2), 'Dernière ligne de données' (set to 'Fin de fichier' par défaut), and 'Délimiteur de texte multiligne'. A 'Modifier' button is located below the form fields. At the bottom, there is a list of other data formats: RDF, Base de données, SPARQL, XML, Shapefile, and GML. A footer note at the bottom right says 'Pour plus d'information sur Datalift, voir <http://www.datalift.org>'.

Atelier 1 – Elévation de données avec Datalift

► Etape 3 : Conversion directe de CSV vers RDF

1 **Projet Donuts**

Description :
Licence : Attribution 3.0 non transposé (CC BY 3.0)
Dernière modification le 21 févr. 2013 11:56:49
Créé le 21 févr. 2013 11:35:48 par fr21954

Modules de transformation applicables

1. Transformation directe CSV vers RDF

2 **Transformation directe CSV vers RDF - Projet Donuts**

Source à convertir : Kiosques_ouverts_a_Paris.csv

Nom de la source à créer : Kiosques_ouverts_a_Paris.csv - (RDF #1)

URI du graphe nommé associé : http://localhost:9091/datalift/project/donuts/source/kiosques-ouverts-a-paris-csv-rdf-1

URI de base : http://localhost:9091/datalift/project/donuts/source/kiosques-ouverts-a-paris-csv-rdf-1

Type RDF cible : myClass

Identifiant d'objet : Automatique (numéro de ligne)

▼ **Typage des données**

Colonne "ADRESSE" : Chaîne de caractères

Colonne "ARRDT" : Chaîne de caractères

Colonne "OUV." : Date ou heure

Colonne "FERM." : Date ou heure

Colonne "OUV. DIM" : Valeur booléenne

Valeurs booléennes vraies : oui_mat

Format des dates : HH'imm

3 **Contenu du graphe nommé <http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/source/kiosques-ouverts-a-paris-selection-csv-rdf-1> (5 000 premiers résultats)**

Sujet	Prédicat	Objet
http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-rec
http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	"PCE DU CHATELET"
http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	"75001"
http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	"08:30:00.000"^^<http://www.w3.org/2001/XMLSchema#time>
http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	"20:00:00.000"^^<http://www.w3.org/2001/XMLSchema#time>
http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	http://datalift.si.fr.atosorigin.com/datalift/project/kiosques-reduit/	"true"^^<http://www.w3.org/2001/XMLSchema#boolean>

Atelier 1 – Elévation de données avec Datalift

- ▶ Etape 3 : Sélection de l'ontologie (LOV)

The screenshot shows the Datalift web interface. At the top left is the Datalift logo. At the top right is a 'SPARQL endpoint' button. The main content area is titled 'Projet Kiosques (réduit)' and has three tabs: 'Description', 'Sources', and 'Ontologies'. The 'Ontologies' tab is active, displaying a table with the following data:

Nom	Lien	Date d'ajout	Opérateur
vcard	http://www.w3.org/2006/vcard/ns#	23 mai 2013 13:21:52	datalift

At the bottom of the interface, there is a link: 'Pour plus d'information sur Datalift, voir <http://www.datalift.org>'.

Atelier 1 – Elévation de données avec Datalift

► Etape 4 : Conversion vers l'ontologie vCard

DataLift SPARQL endpoint

Projet Donuts

Projets
Donuts

Description :
Licence : Attribution 3.0 non transférée (CC BY 3.0)
Dernière modification le 4 mars 2013 13:22:54
Créé le 4 mars 2013 13:15:40 par fr21954

Modules de transformation applicables

1. Transformation directe CSV vers RDF
2. RDF to RDF transformation (CONSTRUCT)
3. Renommage des URI RDF
4. Conversion vers l'ontologie VCard
5. Conversion RDF à base d'ontologies OWL
6. Publication vers le RDF store public
7. Export des données vers fichier RDF
8. Export des données vers fichier CSV

Nouveau projet

Modifier Supprimer

DataLift SPARQL endpoint Espace projets

Ontology Mapper

Source à convertir : Kiosques_Paris.csv - (RDF #1)

Créer une nouvelle source pour les données converties

Nom de la nouvelle source : Kiosques localisés

URI du graphe nommé associé : http://localhost:9091/datalift/project/kiosques/source/kiosques-localises

Ontologie : vcard

Type RDF cible : VCard - Resources that are vCard **Valider**

DataLift SPARQL endpoint Espace projets Connecté en tant que Ibihanic

Ontology Mapper

Conversion vers adr: Address

Propriétés

Typage

Nouvelle propriété : country-name - The country o

Valeur: "France"

Ajouter

- VCard **Supprimer**
- adr: Address (Work, Pref) **Supprimer**
- country-name = "France" **Supprimer**
- locality = "Paris" **Supprimer**
- postal-code = arrdt **Supprimer**
- street-address = adresse **Supprimer**
- fn = "Kisoque " + adresse **Supprimer**

Afficher la requête SPARQL générée

► Convertir les données

Atelier 1 – Elévation de données avec Datalift

► Etape 4 : Conversion vers l'ontologie vCard (suite)

Prévisualisation de la requête SPARQL

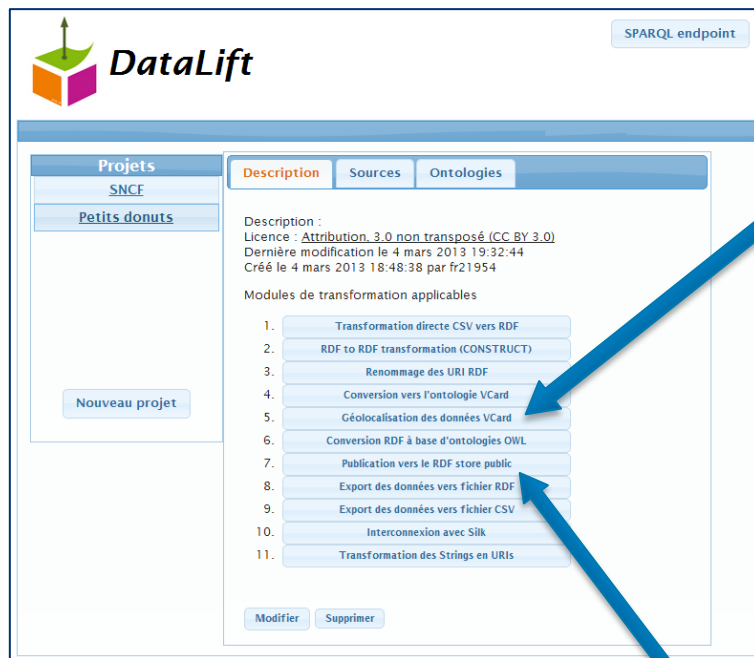
```
PREFIX p1: <http://localhost:9091/datalift/project/kiosques/source/kiosques-paris-csv-rdf-1#>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>

CONSTRUCT {
    ?v1    a vcard:VCard ;
          vcard:adr _:b3 ;
          vcard:fn ?v2 .
    _:b3   a vcard:Address ,
          vcard:Work ,
          vcard:Pref ;
          vcard:country-name "France" ;
          vcard:locality "Paris" ;
          vcard:postal-code ?arrdt ;
          vcard:street-address ?adresse .
}
WHERE {
    GRAPH <http://localhost:9091/datalift/project/kiosques/source/kiosques-paris-csv-rdf-1#> {
        ?v1    p1:adresse ?adresse ;
              p1:arrdt ?arrdt .
    }
    BIND(concat("Kisoque ", ?adresse) AS ?v2)
}
}
```

4

Atelier 1 – Elévation de données avec Datalift

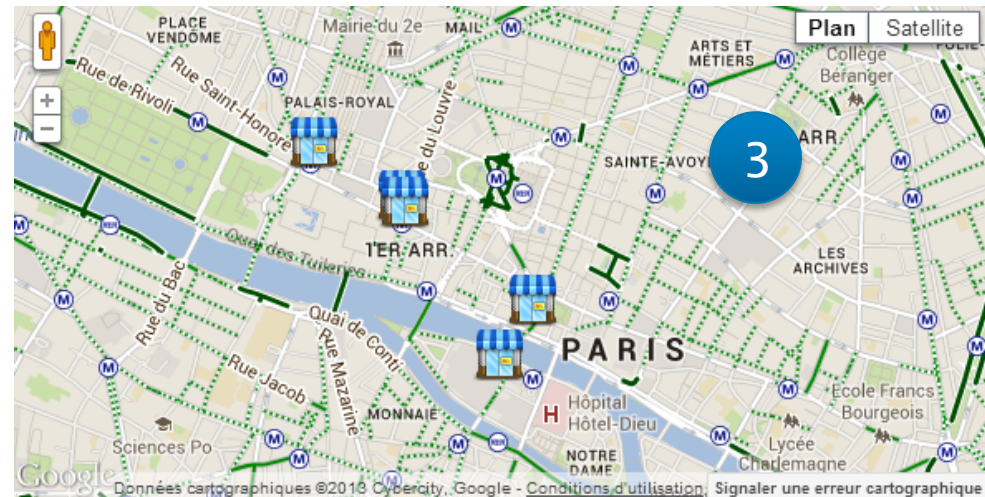
▶ Etape 5 : Publication & Consommation



The screenshot shows the Datalift web interface. On the left, there is a sidebar with 'Projets' containing 'SNCF' and 'Petits donuts'. The main area has tabs for 'Description', 'Sources', and 'Ontologies'. Under 'Description', it shows project details and a list of 11 applicable transformation modules. Two blue arrows point from the interface to the map: one from the 'Conversion vers l'ontologie VCard' module (labeled '1') and another from the 'Export des données vers fichier RDF' module (labeled '2').

Modules de transformation applicables
1. Transformation directe CSV vers RDF
2. RDF to RDF transformation (CONSTRUCT)
3. Renommage des URI RDF
4. Conversion vers l'ontologie VCard
5. Géolocalisation des données VCard
6. Conversion RDF à base d'ontologies OWL
7. Publication vers le RDF store public
8. Export des données vers fichier RDF
9. Export des données vers fichier CSV
10. Interconnexion avec Silk
11. Transformation des Strings en URIs

Démo : fusion SPARQL - Google Maps



<http://datalift.si.fr.atosorigin.com/datalift/donutsapp>

Atelier 2 – Croisement de données

- ▶ Données Open Data classiques : 2 catalogues d'étoiles
 - Format : CSV
- ▶ Conversion en RDF
 - Transformation brute (« *direct mapping* »)
- ▶ Requête SPARQL de rapprochement des données
 - Trouver les étoiles présentes dans les 2 catalogues

Atelier 2 – Croisement de données

► Etape 1 : Récupération des données

<http://vizier.u-strasbg.fr/viz-bin/VizieR>

► Catalogues :

– 2mass : <http://datalift.si.fr.atosorigin.com/datalift/project/etoiles/source/2mass-csv>

– usno500 : <http://datalift.si.fr.atosorigin.com/datalift/project/etoiles/source/usno500-csv>

UDS Portal Simbad VizieR Aladin X-Match Other Help

VizieR Service

new access to the [CFHTLS-T0007 catalogs](#) (Wide and Deep fields)

Find catalogs among 10670 available

Clear Find...

Expand search

Wavelength Mission Astronomy

Radio	AKARI	AGN
IR	ANS	Abundances
optical	ASCA	Ages
UV	BeppoSAX	Associations
EUV	CGRO	Atomic_Data
X-ray	COBE	BL_Lac_objects
Gamma-ray	Chandra	Binaries:cataclysmic

Search for catalogs by column descriptions (UCD) ?

Search by Position across 11125 tables

Target Name (resolved by [Sesame](#)) or Position: J2000 2 arcmin Go!

Target dimension: Radius Box size

[More about VizieR](#) Find Catalogs

Browsing modes: [Designation](#), [Acronyms](#), [Favorites](#), [Dates](#), [Image spectra](#), [Kohonen](#)
Or list [the large surveys](#)

→ Thanks for acknowledging the VizieR Service © UDS/CNRS Contact

Atelier 2 – Croisement de données

▶ Etape 2 : Import des données CSV

DataLift

Ajouter un jeu de données - Proj

[Retour à l'espace de travail](#)

▼ CSV

Fichier CSV à télécharger *: Aucun fichier choisi

ou URL de téléchargement *:

Nom de la source *:

Description :

Encodage :

Séparé par *: **1**

Position de la ligne de titres :

Première ligne de données :

Dernière ligne de données :

Délimiteur de texte multiligne :

- ▶ RDF
- ▶ Base de données
- ▶ SPARQL
- ▶ XML
- ▶ Shapefile
- ▶ GML

DataLift

Ajouter un jeu de données - Projet etoiles

[Retour à l'espace de travail](#)

▼ CSV

Fichier CSV à télécharger *: Aucun fichier choisi

ou URL de téléchargement *:

Nom de la source *:

Description :

Encodage :

Séparé par *: **2**

Position de la ligne de titres :

Première ligne de données :

Dernière ligne de données :

Délimiteur de texte multiligne :

- ▶ RDF
- ▶ Base de données
- ▶ SPARQL
- ▶ XML
- ▶ Shapefile
- ▶ GML

Atelier 2 – Croisement de données

- ▶ Etape 3 : Conversion directe de CSV vers RDF



The screenshot displays the DataLift web interface. At the top left is the DataLift logo, which consists of a stylized cube with an upward-pointing arrow. To the right of the logo is the text "DataLift". In the top right corner of the interface, there is a blue circular badge containing the text "x2". Below the logo, the page title "Projet etoiles" is visible. The main content area is divided into a left sidebar and a main panel. The sidebar has a "Projets" header and a list item "etoiles". Below the sidebar is a "Nouveau projet" button. The main panel has three tabs: "Description" (selected), "Sources", and "Ontologies". Under the "Description" tab, the text reads: "Description :", "Licence : Attribution 3.0 non transposé (CC BY 3.0)", "Dernière modification le 26 nov. 2012 13:38:19", and "Créé le 26 nov. 2012 13:34:44 par datalift". Below this is the section "Modules de transformation applicables" with a numbered list: "1. Transformation directe CSV vers RDF" and "2. Interconnection". At the bottom of the main panel are "Modifier" and "Supprimer" buttons.

Atelier 2 – Croisement de données

▶ Etape 4 : Requête SPARQL

DataLift Espace projets

Editeur de requête SPARQL

Dépôt: RDF store interne | Format de la réponse: RDF/XML | Exécuter la requête

```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX u:<http://datalift.si.fr.atosorigin.com/datalift/project/etoiles/source/usno500-csv-rdf-1#>
3 PREFIX m:<http://datalift.si.fr.atosorigin.com/datalift/project/etoiles/source/2mass-csv-rdf-1#>
4
5 SELECT * WHERE {
6   ?s1 u:dej2000 ?Decl1 ;
7   u:raj2000 ?AD1 .
8   ?s2 m:dej2000 ?Decl2 ;
9   m:raj2000 ?AD2 .
10  BIND (abs(xsd:decimal(?AD1)-xsd:decimal(?AD2)) as ?deltaAD)
11  BIND (abs(xsd:decimal(?Decl1)-xsd:decimal(?Decl2)) as ?deltaDecl)
12  FILTER (?deltaAD < 0.00001)
13  FILTER (?deltaDecl < 0.00001)
14 } LIMIT 20
  
```

Line: 3; Position: 85; Query is valid

Résultats de la requête

s1	Decl1	AD1	s2	Decl2	AD2	deltaAD	deltaDecl
http://.../project/etoiles/source/usno500-csv-rdf-1/228	+36.553489	282.838067	http://.../project/etoiles/source/2mass-csv-rdf-1/91	+36.553493	282.838059	0.000008	0.000004

s	Decl	AD
http://.../project/etoiles/source/usno500-csv-rdf-1/228	+36.553489	282.838067
http://.../project/etoiles/source/2mass-csv-rdf-1/91	+36.553493	282.838059

Atelier 2 – Croisement de données

► Etape 4 : Requête SPARQL (détail)

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX u: <http://.../project/etoiles/source/usno500-csv-rdf-1#>
PREFIX m: <http://.../project/etoiles/source/2mass-csv-rdf-1#>
```

```
SELECT * WHERE {
    ?s1      u:dej2000 ?Decl1 ;
            u:raj2000 ?AD1 .
    ?s2      m:dej2000 ?Decl2 ;
            m:raj2000 ?AD2 .

    BIND (abs(xsd:decimal(?AD1) - xsd:decimal(?AD2)) as ?deltaAD)
    BIND (abs(xsd:decimal(?Decl1) - xsd:decimal(?Decl2)) as ?deltaDecl)

    FILTER (?deltaAD < 0.00001)
    FILTER (?deltaDecl < 0.00001)
}
LIMIT 20
```

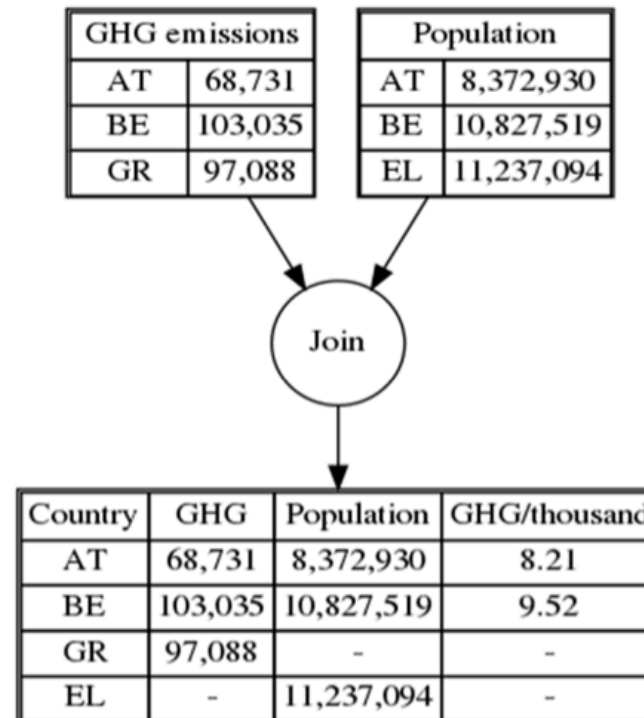
Cas d'utilisation 1 - Agence Européenne de l'Environnement

Avant

Formats propriétaires

locality	Field2	Field3	Field4
[tsieb010] - BIP pro Kopf in KK3			
BIP pro Kopf in Kaufkraftstanda			
geo\time	1995		1996
EU (27 Länder)	100		100
EU (25 Länder)	105		105
EU (15 Länder)	116		116
Euroraum (16 Länder)	114		114
Euroraum (15 Länder)	116		115
Belgien	129		126
Bulgarien	32		28
Tschechische Republik	73		75
Dänemark	132		133
Deutschland	129		127
Estland	36	(b)	38
Irland	103		108
Griechenland	84		84
Spanien	92		92
Frankreich	116		115

Référentiels hétérogènes



Cas d'utilisation 1 – Agence Européenne de l'Environnement

Demands

- ▶ Données au seul format RDF
 - Autres formats refusés : CSV, Excel...
 - Utilisation de vocabulaires et référentiels (libres)
- ▶ Mise à disposition de fichiers (HTTP)
 - Accès par web services refusé
- ▶ Mise à disposition d'un catalogue des jeux de données disponibles
 - Format imposé VOID (RDF)

Cas d'utilisation 1 – Agence Européenne de l'Environnement

Mise en œuvre

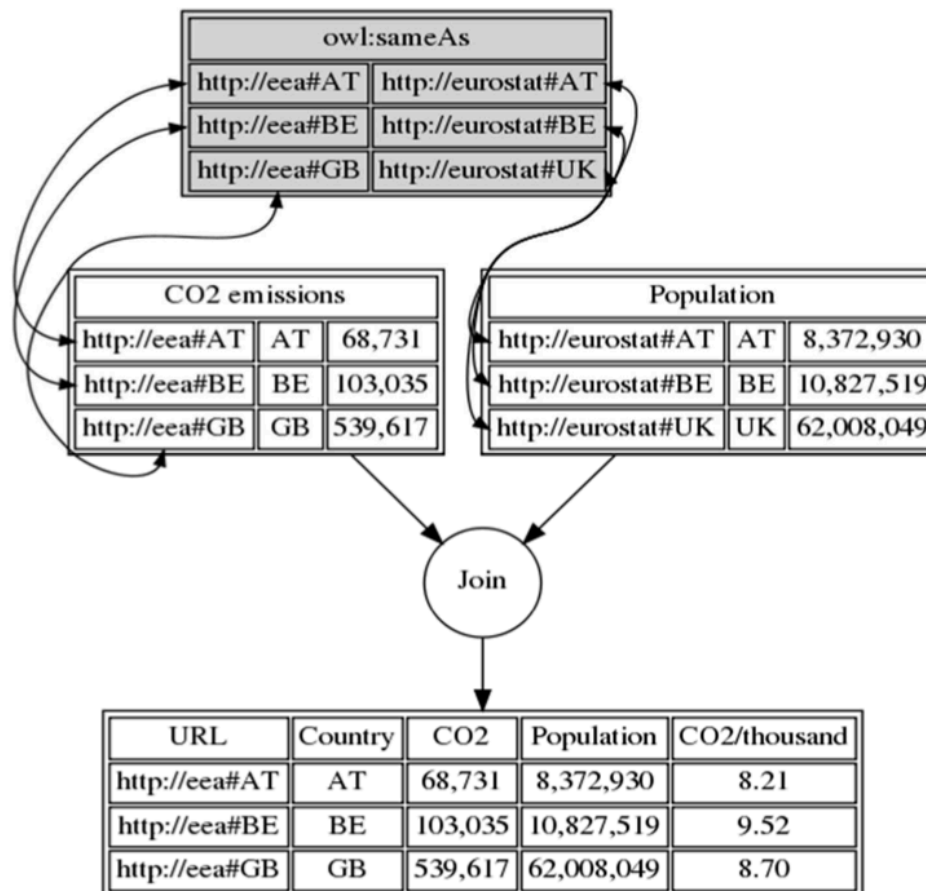
- ▶ Mise en correspondance des référentiels des états membres
 - Via une ontologie :

<http://eurostat.europa.eu/countries#UK> = <http://eea.europa.eu/countries.rdf#GB>

- ▶ Utilisation d'un triple store RDF supportant l'inférence
 - Prise en compte transparente des correspondances
- ▶ Chargement de tous les jeux de données dans ce triple store
 - Interrogation SPARQL

Cas d'utilisation 1 - Agence Européenne de l'Environnement

Après



Cas d'utilisation 1 – Agence Européenne de l'Environnement

Exemple : Avenir des population de loutres

PREFIX art17: <http://rdfdata.eionet.europa.eu/art17/ontology/>
PREFIX eea: <http://rdfdata.eionet.europa.eu/eea/ontology/>

```
SELECT ?country ?region ?future WHERE {  
[] art17:forSpecies <http://eunis.eea.europa.eu/species/1435>;  
  art17:hasRegionalReport ?report.  
?report art17:conclusion_future ?future;  
  art17:forCountry ?curl;  
  art17:region ?bregion.  
?bregion eea:name ?region.  
?curl eea:name ?country  
} ORDER BY ?country ?region
```

country	region	future
Austria	Alpine	Inadequate (U1)
Austria	Continental	Inadequate (U1)
Belgium	Atlantic	Bad (U2)
Belgium	Continental	Bad but improving (U2+)
Czech Republic	Continental	Favourable (FV)
Czech Republic	Pannonian	Favourable (FV)
Estonia	Boreal	Favourable (FV)

Cas d'utilisation 2 : data.insee.fr

► Publication de données avec Datalift

- [Données géographiques](#) : données du Code officiel géographique (COG) incluant régions, départements, arrondissements, cantons et communes
- [Codes et nomenclatures](#) : nomenclature d'activités française (NAF), nomenclatures des professions et catégories professionnelles (PCS) et des catégories juridiques (CJ).
- [Données de population](#) : populations légales issues du Recensement.

► Bénéfices

- Format de publication unique : RDF
- Interconnexion IGN / partage de référentiel
- Politique d'URI séparant ressource informationnelles et non informationnelles
- Interface unifiée pour pages HTML et web services (*en cours*)
- Représentations multiples (négociation de contenu)
- Données requêtables en ligne (temps d'exécution contrôlé)

Cas d'utilisation 2 : data.insee.fr



Requête SPARQL

Format de la réponse : **HTML** RDF/XML N3/Turtle NTriples TriG TriX CSV

Requête ([Aide](#)) :

```
# Remplacer '92002' par le code de la commune cherchée
PREFIX idemo:<http://rdf.insee.fr/def/demo#>
PREFIX igeo:<http://rdf.insee.fr/def/geo#>

SELECT ?commune ?code ?popTotale WHERE {
  ?commune igeo:codeCommune ?code .
  ?commune igeo:nom "Montesson" .
  ?commune idemo:population ?popLeg .
  ?popLeg idemo:populationTotale ?popTotale .
}
```

Requêtes prédéfinies : **Région par son nom** Groupe NAF par son code Population d'une commune

Nombre maximum de résultats :

Rechercher dans

commune	code	popTotale
http://id.insee.fr/geo/commune/78418	78418	15412

Cas d'utilisation 3 : data.ign.fr

► Publication de données avec Datalift

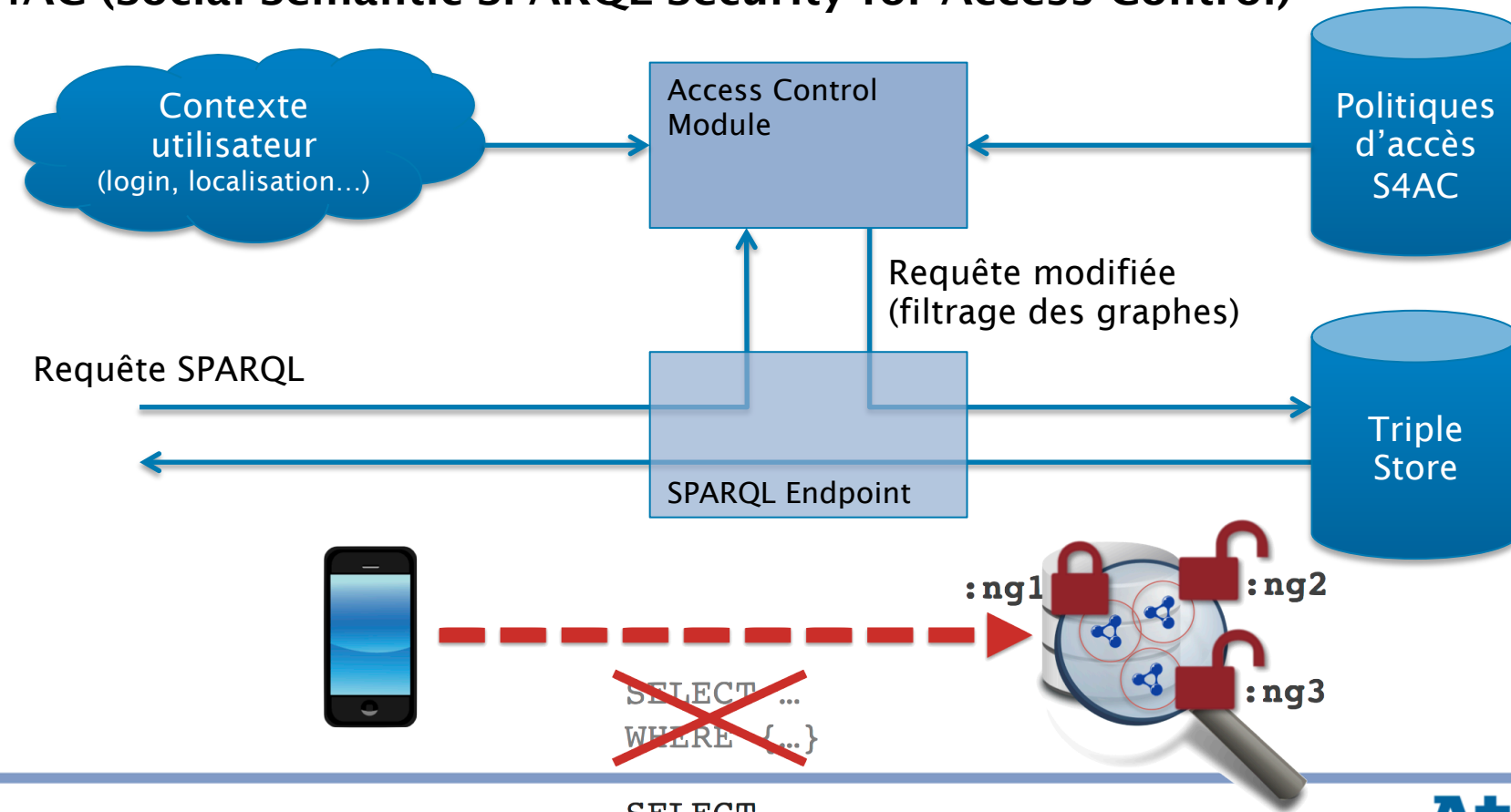
- Site expérimental : DB Topo uniquement
- Lien avec le Géoportail (INSPIRE) *en cours*

► Bénéfices

- Interconnexion INSEE / partage de référentiel
- Représentations multiples (négociation de contenu)
 - GML, WKT, cartes Géoportail...
- Contrôle d'accès au données
 - Données semi-ouvertes (administrations / secteur privé)
 - Composant Datalift S4AC (Social Semantic SPARQL Security for Access Control)
- Gestion dynamique des licences
 - Calcul des termes de licence applicables aux données extraites
 - Composant Datalift SHILD (*en cours*)

Cas d'utilisation 3 : data.ign.fr

S4AC (Social Semantic SPARQL Security for Access Control)



Cas d'utilisation 3 : data.ign.fr

The screenshot displays the DATA.IGN.FR website interface. At the top, the title "DATA.IGN.FR" is prominently displayed in white on a dark blue background. Below the title, a navigation menu includes "ACCUEIL", "DONNÉES", "SPARQL ENDPOINT", and "CONTACTS". The main content area is divided into two sections. On the left, the "SPARQL ENDPOINT" section features a "Format de la réponse" dropdown menu with options: HTML (selected), RDF/XML, N3/Turtle, NTriples, and TriG. Below this is a "Requête" input field containing the SPARQL query:

```
SELECT * WHERE {  
  ?s ?p ?o .  
}
```

 and a "Nombre maximum de résultats" input field set to "5000". A blue button labeled "Exécuter la requête" is positioned below the input fields. On the right, a "LINKS" section lists "Datalift", "Linked data", "W3C", and "data.gouv.fr". At the bottom of the sidebar, there are logos for "DATALIFT" and "IGN INSTITUT NATIONAL DE L'INFORMATION GEOGRAPHIQUE ET FORESTIERE".

Thanks

For more information, please contact:

Laurent BIHANIC
laurent.bihanic@atos.net

Atos France
River Ouest
80, quai Voltaire
95877 Bezons Cedex

atos.net

Atos, the Atos logo, Atos Consulting, Atos Worldline, Atos Sphere, Atos Cloud and Atos WorldGrid are registered trademarks of Atos SA. July 2011

© 2011 Atos Consulting. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.

Your business technologists. **Powering progress**

