

Presentation metadata

Open Data Support is funded by the European Commission under SMART 2012/0107 'Lot 2: Provision of services for the Publication, Access and Reuse of Open Public Data across the European Union, through existing open data portals' (Contract No. 30-CE-0530965/00-17).

© 2013 European Commission



OPEN DATA SUPPORT

Module de formation 1.2

Introduction aux données liées

Les objectifs de cette formation

À la fin de ce module de formation, vous devriez avoir une compréhension de:

- Que sont les données liées;
- Que sont les données ouvertes;
- Quelle est la différence entre les données liées et ouvertes;
- Comment publier des données liées (schéma 5 étoiles);
- Les aspects économiques et sociaux des données liées.

Contenu

Ce module contient ...

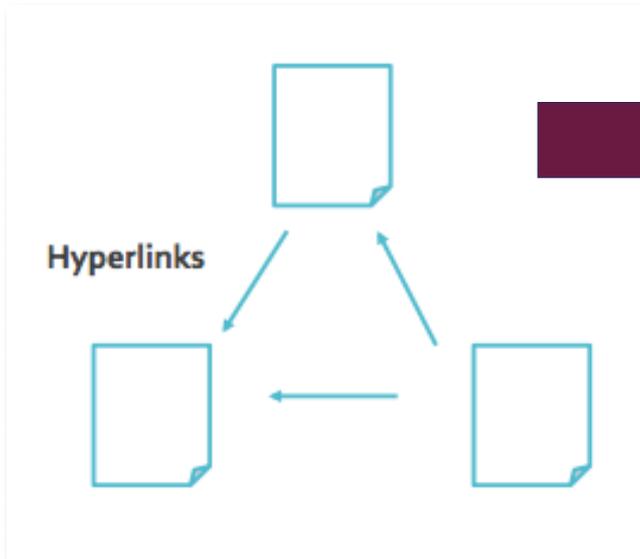
- Une introduction aux principes des données liées;
- Une introduction aux technologies des données liées;
- Un aperçu du système des 5 étoiles pour la publication des données liées;
- Un exemple de la façon dont les données tabulaires peuvent être publiées sous forme de données liées en utilisant “Open Refine”;
- Les bénéfices attendus des données liées pour les gouvernements;
- Un aperçu des initiatives pour les données liées en Europe.

Que sont les données liées?

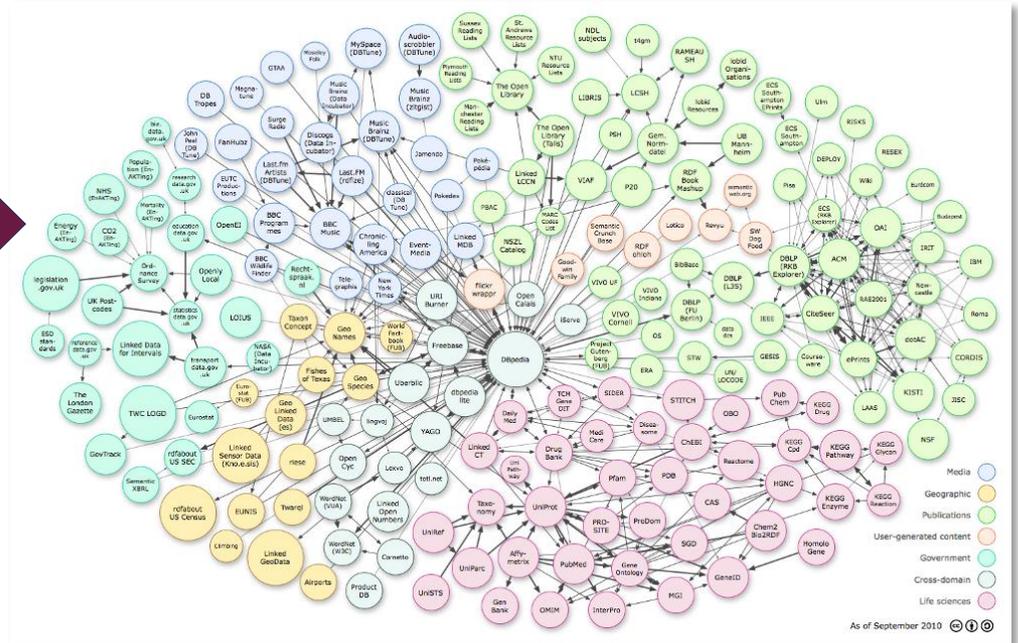
Evolution d'un Web basé sur les documents à un Web de données interconnectées.

Le Web évolue d'un «Web de documents liés» à un «Web de données liées» ... (1/2)

Web des documents ...



Web des données liées ...



Le Web évolue d'un «Web de documents liés» à un «Web de données liées » ... (2/2)

- Le Web a commencé comme une collection de documents publiés en ligne - accessibles à un endroit du Web identifié par une URL.
- Ces documents contiennent souvent des données sur les ressources du monde réel qui sont principalement écrites en langage humain et ne peuvent pas être comprises par les machines.
- Le Web des données vise à permettre l'accès à ces données, en les rendant disponibles dans des formats lisibles par machine et en les connectant à l'aide d'identificateurs de ressources uniformes (URI), permettant ainsi aux gens et aux machines de collecter ces données, et de les assembler afin de les utiliser pour toutes sortes de choses (pour autant que la licence le permette).

Les (méta)données lisibles par machine sont des données dans un format compréhensible par un ordinateur.

2 types de données lisibles par machine:

- données en langage humain qui sont annotées de façon à être comprises par les ordinateurs, RDFa;
- formats de données conçus principalement pour les ordinateurs, par ex. RDF, XML and JSON.

Voir aussi:

http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html

<http://linkeddatabook.com/editions/1.0/>

Définir ce que sont les données liées ...

“Les données liées sont un ensemble de principes de conception pour le partage de données lisibles par machine sur le Web pour une utilisation par les administrations publiques, les entreprises et les citoyens.”

EC ISA Case Study: How Linked Data is transforming eGovernment

Les **quatre principes de conception** des données liées (par Tim Berners Lee):

1. Utiliser des identificateur de ressources uniformes(URI) pour les noms des choses.
2. Utiliser des URIs HTTPde sorte que les gens puissent consulter ces adresses.
3. Quand quelqu'un consulte une URI, fournir des informations utiles, en utilisant les standards (RDF *, SPARQL).
4. Inclure des liens vers d'autres URIs afin qu'ils puissent découvrir plus de choses.

Voir aussi:

http://www.youtube.com/watch?v=4x_xzT5eF5Q
<http://www.w3.org/DesignIssues/LinkedData.html>
<http://www.youtube.com/watch?v=uju4wT9uBIA>

Données publiques liées (ouvertes) – proposition de valeur

- **Integration flexible des données:** LOGD facilite l'intégration de données et permet l'interconnexion de bases de données publiques auparavant disparates.
- **Augmentation de la qualité des données:** L'augmentation de la (ré)utilisation des LOGD déclenche une demande croissante pour améliorer la qualité des données. Grâce au crowd-sourcing et aux mécanismes de self-service, les erreurs sont progressivement corrigées.
- **Nouveaux services:** La disponibilité des LOGD donne naissance – de nouveaux services offerts par le secteur public et/ou privé.
- **Réduction des coûts:** La réutilisation des LOGD dans les applications d'e-gouvernement conduit à des baisses de coûts considérables.

Voir aussi:

Etude ISA sur les modèles d'affaires pour les LOGD

<https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Les quatre principes dans la pratique ... (1)

1. Utiliser des identificateur de ressources uniformes(URI) pour les noms des choses.
2. Utiliser des URIs HTTPde sorte que les gens puissent consulter ces adresses.

*Exemple: pour une organisation: **UNICEF***

- <http://publications.europa.eu/resource/authority/corporate-body/UNICEF>



Les quatre principes dans la pratique... (2)

1. Quand quelqu'un consulte une URI, fournir des informations utiles, en utilisant les standards (RDF *, SPARQL).
2. Inclure des liens vers d'autres URIs afin qu'ils puissent découvrir plus de choses.

```
<skos:Concept rdf:about="http://publications.europa.eu/resource/authority/corporate-body/UNICEF"  
  at:deprecated="false">  
  <skos:inScheme rdf:resource="http://publications.europa.eu/resource/authority/corporate-body"/>  
  <skos:broader rdf:resource="http://publications.europa.eu/resource/authority/corporate-body/UNO"/>  
  <at:authority-code>UNICEF</at:authority-code>  
  <at:op-code>UNICEF</at:op-code>  
  <atold:op-code>UNICEF</atold:op-code>  
  <dc:identifier>UNICEF</dc:identifier>  
  <at:start.use>1951-01-01</at:start.use>  
  <skos:prefLabel xml:lang="bg">Унищф - Детски фонд на ООН</skos:prefLabel>  
  <skos:prefLabel xml:lang="cs">UNICEF - Dětský fond Organizace spojených národů</skos:prefLabel>  
  <skos:prefLabel xml:lang="da">UNICEF - De Forenede Nationers Børnefond</skos:prefLabel>  
  <skos:prefLabel xml:lang="de">Unicef - Kinderhilfswerk der Vereinten Nationen</skos:prefLabel>  
  <skos:prefLabel xml:lang="el">Unicef - Ταμείο των Ηνωμένων Εθνών για τα Παιδιά</skos:prefLabel>  
  <skos:prefLabel xml:lang="en">Unicef - United Nations Children's Fund</skos:prefLabel>
```

Données liées >< données ouvertes

« Les données ouvertes sont des données qui peuvent être librement utilisées, réutilisées et redistribuées par tout un chacun – à la seule condition, au plus, de devoir en attribuer la source de les partager sous les mêmes conditions. »
OpenDefinition.org

Données ouvertes

Les données peuvent être publiées et rendues publiques sous une licence ouverte sans les lier à d'autres sources.



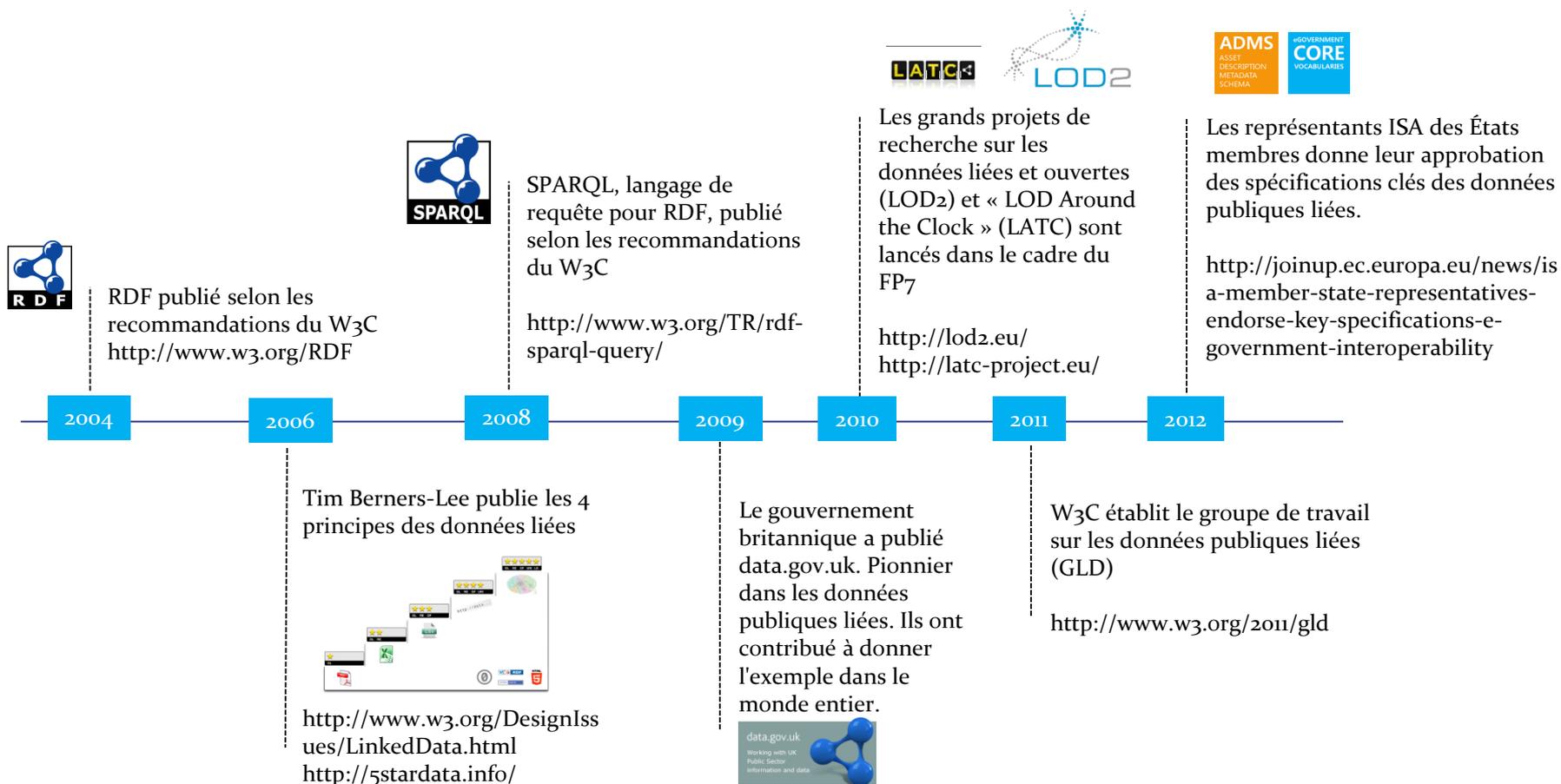
Données liées

Les données peuvent être liées aux URIs d'autres sources de données, en utilisant des standards ouverts tels que RDF, sans être disponibles publiquement sous une licence

Voir aussi:

Cobden et al., A research agenda for Linked Closed Data
http://ceur-ws.org/Vol-782/CobdenEtAl_COLD2011.pdf

Les principales étapes pour les données liées publiques



Les fondements des données liées

Des URIs pour nommer les choses, RDF pour décrire des données et SPARQL pour les interroger.

Identificateur de ressource uniforme(URI)

“Un identifiant uniforme de ressource (URI) est une séquence restreinte de caractères qui identifie une ressource abstraite ou physique.”

– ISA’s 10 Rules for Persistent URIs

Un pays, par ex. la Belgique

- <http://dbpedia.org/resource/Belgium>



Une organisation, par ex. Office des publications de l’Union européenne

- <http://publications.europa.eu/resource/authority/corporate-body/PUBL>



Un ensemble de données, par ex. liste des indicateurs de pays

- <http://publications.europa.eu/resource/authority/country/>



Voir aussi:

<http://www.slideshare.net/OpenDataSupport/design-and-manage-persitent-uris>

RDF & SPARQL

Le **cadre de description des ressources** (RDF) est une syntaxe pour représenter les données et les ressources du Web.

RDF décompose chaque élément d'information en **triplets**:

- **Sujet** - une ressource qui peut être identifiée par un URI.
- **Prédicat** - une spécification réutilisée et identifiée par URI de la relation.
- **Objet** - une ressource ou constante à laquelle le sujet est lié.

<http://dbpedia.org/resource/Brussels> is the capital of “Belgium”.

OU

<http://dbpedia.org/resource/Brussels> is the capital of <http://dbpedia.org/resource/Belgium>.

Sujet

Prédicat

Objet

SPARQL est un langage standardisé pour l'interrogation de données RDF.

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/introduction-to-rdf-sparql>

Comment publier des données liées?

Ouvrir la voie à des données liées 5 étoiles

Le schéma 5 étoiles des données liées (ouvertes)

- ★ Rendre vos données disponibles sur le Web (peu importe le format) sous une licence ouverte optionnel
- ★★ Mettre vos données à disposition sous forme de données structurées (par exemple, Excel au lieu d'une image de numérisation d'un tableau)
- ★★★ Utiliser des formats non-propriétaires (par exemple, CSV au lieu d'Excel)
- ★★★★ Utiliser des URIs pour désigner les choses, de sorte que les gens puissent rediriger vers vos choses
- ★★★★★ Lier vos données à d'autres données pour fournir du contexte

★ *Rendre vos données disponibles sur le Web sous une licence ouverte*



Sustainable development targets for 2011-12

Our business plan for 2011–2015 sets out our strategic objectives for the next four years and our specific business priorities for 2011–12. Our aim of 'a more sustainable Kew' sets out actions which will deliver significant sustainability benefits.

1. Reduce carbon emissions at Kew by 6% from 2010–11 levels, balancing record preservation and environmental conditions

Period	Electricity (KWh)	Gas (KWh)	Carbon (tonnes)	Change on 2010–11 (%)
April	762,625	354,062	479	-8.6
May	757,291	348,324	475	-6.5
June	846,364	388,369	530	-13.0
July	908,864	338,278	555	-17.1
August	928,827	384,925	574	-6.1
September	868,526	463,960	556	-2.7
October	810,768	376,137	509	-11.0
November	697,957	439,482	459	-17.1
December	536,080	472,718	378	-24.9

Performance to the end of December 2011 is -11.6%, well ahead of target. Our long-term commitment, which we are on track to meet or exceed, is to reduce emissions from buildings and business-related travel by 25% from 2009–10 levels by April 2015.

Licensed under **OGI** Open Government Licence.

Avantages et inconvénients des données ouvertes★

En tant que consommateur ...	En tant qu'éditeur ...
✓Vous pouvez consulter les données	✓Simplicité de publication.
✓Vous pouvez les stocker localement.	✓Vous n'avez pas à expliquer à plusieurs reprises aux autres qu'ils peuvent utiliser vos données.
✓Vous pouvez entrer les données dans un autre système.	
✓Vous pouvez modifier les données.	
✓Vous pouvez partager les données avec n'importe qui.	

★ ★ Mettre à disposition sous forme de données structurées

Table DA2301 (SST2.10): Security and fire safety - dwellings, 2010

all dwellings

	smoke alarm*	burglar alarm	door viewer	external lighting	secure windows and doors	all dwellings in group (000s)	sample size (unweighted)
	<i>percentage of dwellings within group</i>						
tenure							
owner occupied	-	36,9	51,9	63,3	77,3	14.860	8.791
private rented	-	20,0	48,5	53,2	66,0	3.706	3.096
local authority	-	11,9	67,4	60,9	76,7	1.801	2.276
housing association	-	11,9	75,3	68,0	78,7	2.018	2.507
all private	-	33,6	51,2	61,3	75,1	18.567	11.887
all social	-	11,9	71,6	64,7	77,8	3.819	4.783
dwelling age							
pre-1919	-	25,4	44,3	41,9	58,4	4.865	3.249
1919-44	-	33,1	51,1	54,9	72,3	3.751	2.684
1945-64	-	27,2	54,3	60,2	79,6	4.397	3.609
1965-80	-	26,0	56,6	67,8	81,8	4.602	3.593
1981-90	-	31,2	57,9	77,5	78,6	1.880	1.429
post 1990	-	42,6	72,2	87,5	90,3	2.892	2.196
dwelling type							
end terrace	-	28,6	51,9	51,3	75,3	2.251	
mid terrace	-	24,4	49,6	40,4	72,3	4.105	
small terraced house	-	22,1	49,5	42,1	71,7	2.171	
medium/large terraced house	-	27,8	51,0	45,4	74,3	4.185	



Avantages et inconvénients des données ouvertes ★ ★

Tous les avantages des données ouvertes ★, ainsi que

En tant que consommateur ...	En tant qu'éditeur ...
✓ Vous pouvez traiter les données directement avec des logiciels propriétaires pour agréger les données, effectuer des calculs, visualiser, etc	✓ Toujours facile à publier.
✓ Vous pouvez exporter vos données dans un autre format (structuré et/ou non propriétaire).	

★ ★ ★ Utiliser des formats non-propriétaires

- Propriétaire: Excel, Word, PDF...
- Non- propriétaire : XML, CSV, RDF, JSON, ODF...

Sécurité routière –
Accidents 2006:

```
Acc_Index,Vehicle_Reference,Casualty_Reference,Casualty_Class,Sex_of_Ca
ge_Band_of_Casualty,Casualty_Severity,Pedestrian_Location,Pedestrian_Mo
ar_Passenger,Bus_or_Coach_Passenger,Pedestrian_Road_Maintenance_worker,
_Type,Casualty_Home_Area_Type
200601BS70001,1,1,1,1,6,3,0,0,0,0,-1,4,1
200601BS70002,1,1,1,1,7,2,0,0,0,0,-1,3,1
200601BS70002,1,2,3,1,6,3,1,1,0,0,-1,0,1
200601BS70003,2,1,1,1,9,3,0,0,0,0,-1,9,1
200601BS70005,1,1,3,2,5,3,5,3,0,0,-1,0,1
200601BS70006,2,1,1,1,7,3,0,0,0,0,-1,3,1
200601BS70007,1,1,3,2,10,3,5,9,0,0,-1,0,3
200601BS70009,1,1,3,1,11,3,5,1,0,0,-1,0,-1
200601BS70010,2,1,1,1,8,2,0,0,0,0,-1,1,1
200601BS70013,1,1,1,1,7,3,0,0,0,0,-1,9,1
200601BS70015,1,1,3,2,10,2,1,1,0,0,-1,0,1
200601BS70017,1,1,1,1,7,3,0,0,0,0,-1,5,1
200601BS70018,1,1,1,2,6,3,0,0,0,0,-1,9,1
200601BS70019,2,1,1,2,5,3,0,0,0,0,-1,1,1
200601BS70020,2,1,1,1,6,2,0,0,0,0,-1,3,1
200601BS70021,1,1,1,1,8,3,0,0,0,0,-1,3,1
200601BS70021,1,2,2,1,2,3,0,0,0,0,-1,3,-1
200601BS70022,1,1,1,1,5,3,0,0,0,0,-1,5,1
200601BS70023,2,1,1,1,6,3,0,0,0,0,-1,5,1
200601BS70024,1,1,3,2,5,3,1,3,0,0,-1,0,1
200601BS70025,1,1,1,2,9,3,0,0,0,0,-1,9,1
200601BS70027,1,1,3,1,6,2,1,1,0,0,-1,0,-1
200601BS70028,1,1,1,2,7,3,0,0,0,0,-1,9,1
200601BS70029,2,1,1,1,5,2,0,0,0,0,-1,3,-1
200601BS70030,1,1,3,2,6,3,4,1,0,0,-1,0,1
200601BS70031,2,1,1,1,6,2,0,0,0,0,-1,5,1
200601BS70033,2,1,1,1,6,3,0,0,0,0,-1,3,-1
200601BS70034,1,1,3,1,5,3,4,3,0,0,-1,0,1
200601BS70035,1,1,1,1,6,3,0,0,0,0,-1,1,1
200601BS70036,1,1,1,2,6,2,0,0,0,0,-1,9,1
200601BS70037,1,1,1,1,8,3,0,0,0,0,-1,9,1
200601BS70037,1,2,2,1,-1,3,0,0,1,0,-1,9,-1
200601BS70038,1,1,1,1,6,3,0,0,0,0,-1,5,1
200601BS70039,1,1,1,1,7,3,0,0,0,0,-1,9,1
200601BS70040,1,1,3,2,11,2,4,1,0,0,-1,0,1
200601BS70041,1,1,1,2,6,3,0,0,0,0,-1,9,1
200601BS70042,2,1,1,1,7,2,0,0,0,0,-1,1,1
200601BS70043,1,1,1,1,5,3,0,0,0,0,-1,2,1
200601BS70044,2,1,1,2,9,3,0,0,0,0,-1,9,1
200601BS70045,1,1,3,2,2,3,1,1,0,0,-1,0,1
200601BS70046,2,1,1,2,6,3,0,0,0,0,-1,3,1
200601BS70047,1,1,3,2,3,2,9,5,0,0,-1,0,1
200601BS70048,1,1,3,1,2,3,5,2,0,0,-1,0,1
200601BS70050,1,1,1,1,8,3,0,0,0,0,-1,9,1
200601BS70051,1,1,1,1,10,3,0,0,0,0,-1,9,-1
200601BS70052,1,1,1,1,5,3,0,0,0,0,-1,2,1
200601BS70053,2,1,1,1,8,3,0,0,0,0,-1,9,1
```



Avantages et inconvénients des données ouvertes ★ ★ ★

Tous les avantages des données ouvertes ★ ★ , ainsi que

En tant que consommateur ...	En tant qu'éditeur ...
✓ Vous pouvez manipuler les données comme bon vous semble, sans être limité par les capacités d'un logiciel particulier.	✓ Toujours facile à publier.
	- Mais, vous aurez pas besoin de convertisseurs ou plug-ins pour exporter les données à partir du format propriétaire.

★ ★ ★ ★ *Utiliser des URIs pour designer les choses*

Par exemple, la création d'un URI pour l'une des unités du ministère grec de la Réforme administrative et de l'e-gouvernance.



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Υπουργείο Διοικητικής Μεταρρύθμισης
και Ηλεκτρονικής Διακυβέρνησης

<http://data.ydmed.gov.gr/doc/organization/16180>

URI <http://data.ydmed.gov.gr/id/organization/16180>

Type Organization

Raw data [HTML](#) | [RDF/XML](#) | [Turtle](#)

Category ΥΠΟΥΡΓΕΙΑ

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/design-and-manage-persitent-uris>

Avantages et inconvénients des données ouvertes ★ ★ ★



Tous les avantages des données ouvertes ★ ★ ★ , ainsi que

En tant que consommateur ...	En tant qu'éditeur ...
✓ Vous pouvez y faire référence à partir de n'importe où.	✓ D'autres éditeurs de données peuvent maintenant lier des données au vôtres, en les élevant à 5 étoiles.
✓ Vous pouvez mettre en signet.	✓ Vous pourrez réutiliser les vocabulaires, les (méta)données, et des modèles de conception d'URI au lieu de les créer à partir de zéro.
✓ Vous pouvez accéder à l'information à propos d'une ressource spécifique à travers son URI, sans avoir à télécharger l'ensemble de donnée au complet.	- Mais vous devrez sans doute investir du temps à restructurer et travailler vos données.
✓ Vous pourriez réutiliser des outils et bibliothèques existants	- Vous devez investir dans une réglementation, gestion et infrastructure stable pour les URI persistants.
✓ Vous pouvez combiner les données en toute sécurité avec d'autres données.	- Mais la compréhension de la technologie exige un effort et peut avoir une courbe d'apprentissage abrupte.

★ ★ ★ ★ ★ *Lier vos données à d'autres données pour fournir du contexte*

 **About: Office of the Deputy Minister for Administrative Reform and e-governance**
An Entity of Type : Office,

References | **Referenced By**

type

preferred label

hasUnit

- <http://org.testproject.eu/mareg/def/orgunit/Office>
- Office of the Deputy Minister for Administrative Reform and e-governance
- Γραφείο Υφυπουργού Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης
- Office of the Secretary General for Administrative Reform and e-governance
- Managing Authority of the Operational Programme "Administrative Reform 2007-2013"

 **About: Office of the Secretary General for Administrative Reform and e-governance**
An Entity of Type : Office,

References | **Referenced By**

type

preferred label

hasUnit

- <http://org.testproject.eu/mareg/def/orgunit/Office>
- Office of the Secretary General for Administrative Reform and e-governance
- Γραφείο Γενικού Γραμματέα Διοικητικής Μεταρρύθμισης και Ηλεκτρονικής Διακυβέρνησης
- Directorate General of Financial and Administrative Services
- Directorate General of Administrative Reform and e-Governance
- Directorate General of Human Resources Management



Avantages et inconvénients des données ouvertes ★ ★ ★ ★ ★

Tous les avantages des données ouvertes ★ ★ ★ ★ , ainsi que

En tant que consommateur ...	En tant qu'éditeur ...
✓ Vous pouvez découvrir plus des données (apparentées) tout en consommant les données.	✓ Vous rendez vos données trouvables.
✓ Vous pouvez directement apprendre le schéma des données.	✓ Vous augmentez le contexte, l'expressivité, la qualité et la valeur de vos données (et par conséquent vous donner de la visibilité à votre organisation).
✓ Vous pouvez combiner des données de différentes sources, être novateur, acquérir de nouvelles connaissances, entreprendre ...	- Cela nécessite un investissement en temps, en argent, dans les technologies et les compétences / aptitudes.
- Mais, vous avez maintenant à faire face à des liens morts vers des données. Tous les éditeurs / sources de données ne seront pas fiables.	

Exemple

Utiliser « Open Refine » pour RDF afin de publier des données tabulaires sous forme de données liées.

Qu'est-ce que l'extension "Open Refine RDF"



L'extension "Open Refine RDF" vous permet d'importer facilement des données dans différents formats tels que:

- CSV;
- Excel(.xls and .xlsx);
- JSON;
- XML; and
- RDF/XML.

Et puis déterminer la structure désirée d'un ensemble de données RDF, en dessinant un modèle de graphe.

Voir aussi:
LOD 2 Webinar – Open Refine
<http://www.youtube.com/watch?v=4Ve93C238gI>

Etude de cas: Rapprochement des données sur les produits phytopharmaceutiques

Nous allons montrer comment un ensemble de données du ministère grec du Développement rural et de l'Alimentation a été décrit à l'aide d'une ontologie développée par la DG Santé et consommateurs et a ensuite été publié en tant que données liées.

L'ensemble de données était au format CSV.

Linking data about applications and decisions for authorisation of plant protection products

PPP semantic asset

- [PPP Ontology](#)
- [PPP Taxonomies](#)

Sample queries

- [Find the country where the product is authorised](#)
- [Find a product made with a given substance](#)
- [Find products made by a company](#)
- [Find the product to use on a given pest](#)

Find out more about Linked Data

- [Understanding Linked Data by example](#)
- [Case study on how Linked Data is transforming eGovernment](#)
- [Describe organizations in RDF with Core Business Vocabulary and ORG Ontology](#)
- [10 Rules for Persistent URIs](#)

Type a keyword:

SPARQL Query:

```
Acanto
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ppp: <http://ec.europa.eu/open-data/planthealth/ppp/>
SELECT DISTINCT ?c ?p ?o
FROM <http://health.testproject.eu/ppp/>
WHERE {
  ?c a ppp:Product;
  ?p ?o.
  FILTER (regex(?o, 'Acanto', 'i')).
}
LIMIT 100
```

Search using the [Faceted Browser](#)

This work is supported by [Action 1.1](#) of the [interoperability Solutions for European Public Administrations \(ISA\)](#), Programme of the European Commission.

[Copyrights for the available datasets.](#)

Linked Data pilots: [Core Location pilot](#) | [Core Public Service pilot](#) | [Organisation Ontology pilot](#) | [Plant Protection Products pilot](#) | [Maritime Surveillance pilot](#)

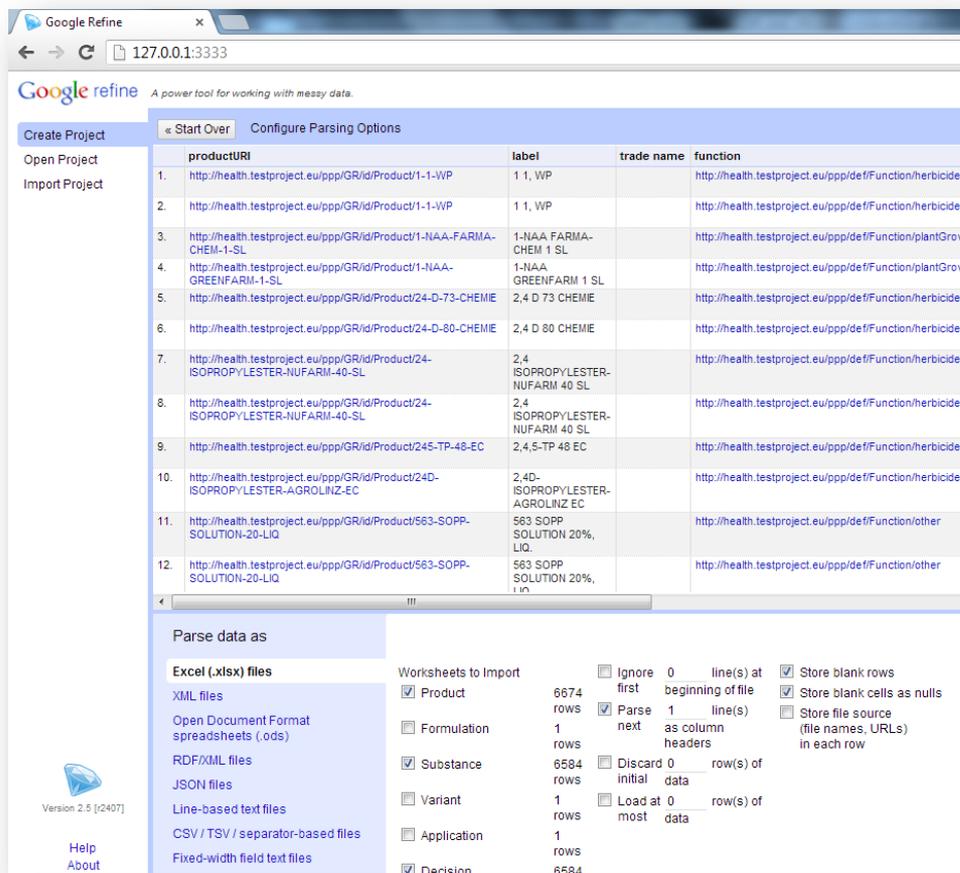
<http://health.testproject.eu/PPP/>

Voir aussi:

http://joinup.ec.europa.eu/asset/core_business/document/linking-data-about-applications-and-decisions-authorisation-ppp

Création du projet dans «Open Refine»

- Assurez-vous que « Open Refine » et l'extension RDF sont installés sur votre machine.
- Lancez « Open Refine ».
- Uploadez la feuille de calcul vers « Open Refine » et choisissez les feuilles que vous souhaitez traiter.
- Confirmez la création du projet.



The screenshot shows the Google Refine web interface in a browser window. The address bar displays the URL 127.0.0.1:3333. The page title is "Google Refine" with the subtitle "A power tool for working with messy data." The main content area is titled "Parse data as" and shows a table of 12 rows of data. The table has columns for "productURI", "label", "trade name", and "function". Below the table, there are several configuration options for parsing the data, including "Worksheets to Import", "Ignore first line(s)", "Parse next as column headers", "Discard initial data", "Load at most row(s)", "Store blank rows", "Store blank cells as nulls", and "Store file source (file names, URLs) in each row".

	productURI	label	trade name	function
1.	http://health.testproject.eu/ppp/GR/id/Product/1-1-WP	1 1, WP		http://health.testproject.eu/ppp/def/Function/herbicide
2.	http://health.testproject.eu/ppp/GR/id/Product/1-1-WP	1 1, WP		http://health.testproject.eu/ppp/def/Function/herbicide
3.	http://health.testproject.eu/ppp/GR/id/Product/1-NAA-FARMA-CHEM-1-SL	1-NAA FARMA-CHEM 1 SL		http://health.testproject.eu/ppp/def/Function/plantGro
4.	http://health.testproject.eu/ppp/GR/id/Product/1-NAA-GREENFARM-1-SL	1-NAA GREENFARM 1 SL		http://health.testproject.eu/ppp/def/Function/plantGro
5.	http://health.testproject.eu/ppp/GR/id/Product/24-D-73-CHEMIE	2,4 D 73 CHEMIE		http://health.testproject.eu/ppp/def/Function/herbicide
6.	http://health.testproject.eu/ppp/GR/id/Product/24-D-80-CHEMIE	2,4 D 80 CHEMIE		http://health.testproject.eu/ppp/def/Function/herbicide
7.	http://health.testproject.eu/ppp/GR/id/Product/24-ISOPROPYLESTER-NUFARM-40-SL	2,4 ISOPROPYLESTER-NUFARM 40 SL		http://health.testproject.eu/ppp/def/Function/herbicide
8.	http://health.testproject.eu/ppp/GR/id/Product/24-ISOPROPYLESTER-NUFARM-40-SL	2,4 ISOPROPYLESTER-NUFARM 40 SL		http://health.testproject.eu/ppp/def/Function/herbicide
9.	http://health.testproject.eu/ppp/GR/id/Product/245-TP-48-EC	2,4,5-TP 48 EC		http://health.testproject.eu/ppp/def/Function/herbicide
10.	http://health.testproject.eu/ppp/GR/id/Product/24D-ISOPROPYLESTER-AGROLINZ-EC	2,4D-ISOPROPYLESTER-AGROLINZ EC		http://health.testproject.eu/ppp/def/Function/herbicide
11.	http://health.testproject.eu/ppp/GR/id/Product/563-SOPP-SOLUTION-20-LIQ	563 SOPP SOLUTION 20%, LIQ.		http://health.testproject.eu/ppp/def/Function/other
12.	http://health.testproject.eu/ppp/GR/id/Product/563-SOPP-SOLUTION-20-LIQ	563 SOPP SOLUTION 20%, LIQ.		http://health.testproject.eu/ppp/def/Function/other

Parse data as

Excel (.xlsx) files

XML files

Open Document Format spreadsheets (.ods)

RDF/XML files

JSON files

Line-based text files

CSV / TSV / separator-based files

Fixed-width field text files

Worksheets to Import

Product 6674 rows

Formulation 1 rows

Substance 6584 rows

Variant 1 rows

Application 1 rows

Decision 6584 rows

Ignore 0 line(s) at first beginning of file

Parse 1 line(s) next as column headers

Discard 0 row(s) of Initial data

Load at 0 row(s) of most data

Store blank rows

Store blank cells as nulls

Store file source (file names, URLs) in each row

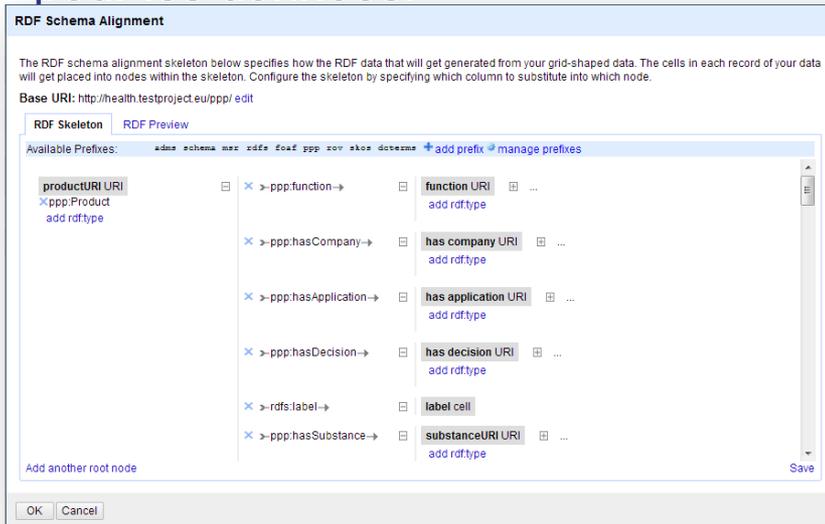
Version 2.5 [2407]

Help About

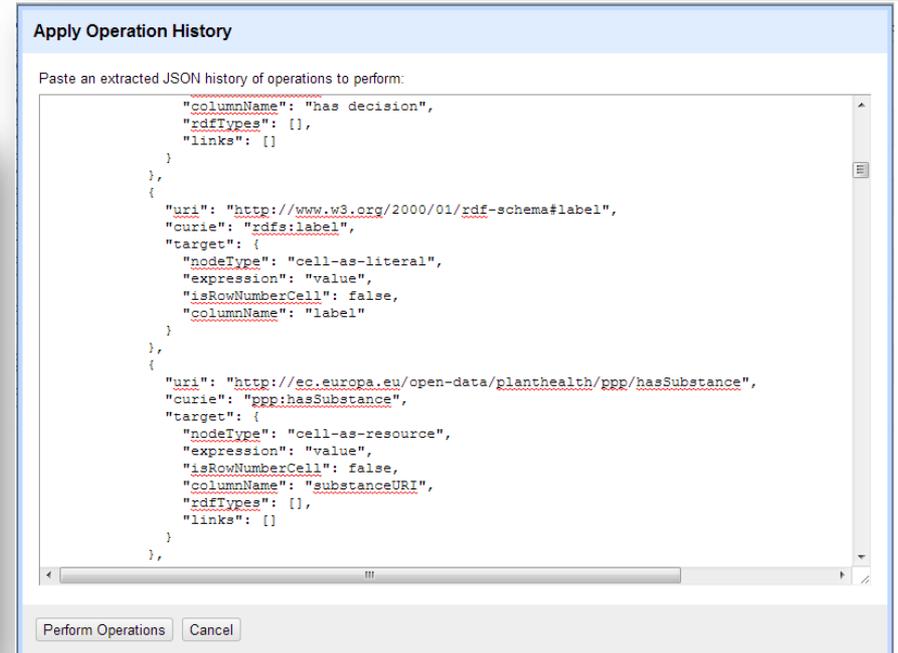
Faire la correspondance entre les données brutes et l'ontologie

Vous pouvez faire la correspondance entre les données et l'ontologie en utilisant une interface graphique simple pour créer ou modifier un squelette RDF existant.

Vous pouvez définir l'URI de base pour les données.



Interface graphique pour éditer un squelette RDF



Interface graphique pour copier / coller un squelette RDF existant

Export des données en RDF – Données Liées

Vous pouvez maintenant exporter vos données en:

- RDF/XML; ou
- Turtle

```
GR_ppp_v06.xlsx - Google x 127.0.0.1:3333/command/ x
127.0.0.1:3333/command/core/export-rows/GR_ppp_v06.xlsx.ttl 200
@prefix msr: <http://www.telegraphis.net/ontology/measurement/> .
@prefix schema: <http://schema.org/> .
@prefix adms: <http://www.w3.org/ns/adms> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ppp: <http://ec.europa.eu/open-data/planthealth/ppp/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rov: <http://www.w3.org/TR/vocab-regorg/> .
@prefix dcterms: <http://purl.org/dc/terms/> .

<http://health.testproject.eu/ppp/GR/id/Product/1-1-WP> a ppp:Product ;
    ppp:function
    <http://health.testproject.eu/ppp/def/Function/herbicide> ;
    ppp:hasCompany
    <http://health.testproject.eu/ppp/GR/id/Company/SHELL-COMPANY-HELLAS-LTD->
    ;
    ppp:hasDecision
    <http://health.testproject.eu/ppp/GR/id/Decision/7103> ;
    rdfs:label "1 1, WP" ;
    ppp:hasProducer
    <http://health.testproject.eu/ppp/GR/id/Company/CIFA> .

<http://health.testproject.eu/ppp/GR/id/Product/1-NAA-FARMA-CHEM-1-SL> a
ppp:Product ;
    ppp:function
    <http://health.testproject.eu/ppp/def/Function/plantGrowthRegulator> ;
    ppp:hasDecision
    <http://health.testproject.eu/ppp/GR/id/Decision/8159> ;
    rdfs:label "1-NAA FARMA-CHEM 1 SL" ;
    ppp:hasProducer
    <http://health.testproject.eu/ppp/GR/id/Company/FARMA-CHEM-SA> .

<http://health.testproject.eu/ppp/GR/id/Product/1-NAA-GREENFARM-1-SL> a
ppp:Product ;
    ppp:function
    <http://health.testproject.eu/ppp/def/Function/plantGrowthRegulator> ;
    ppp:hasCompany
    <http://health.testproject.eu/ppp/GR/id/Company/GREENFARM-CHEMICAL-S-A-> ;
    ppp:hasDecision
    <http://health.testproject.eu/ppp/GR/id/Decision/8124> ;
    rdfs:label "1-NAA GREENFARM 1 SL" ;
```

Export des
données en
Turtle

Facilitateurs et obstacles aux LOGD

A partir de l'étude sur les modèles d'affaires pour les LOGD du programme ISA de la Commission Européenne.

Facilitateurs de LOGD

- Gains d'efficacité en intégration de données – l'effet réseau.
- Stratégies prospectives.
- Liaison accrue et des services intégrés.
- Facilité de mise à jour du modèle.
- Facilité de navigation.
- Licences ouvertes et libres d'accès.
- Enthousiasme des “champions”.
- Emergence de meilleures pratiques.

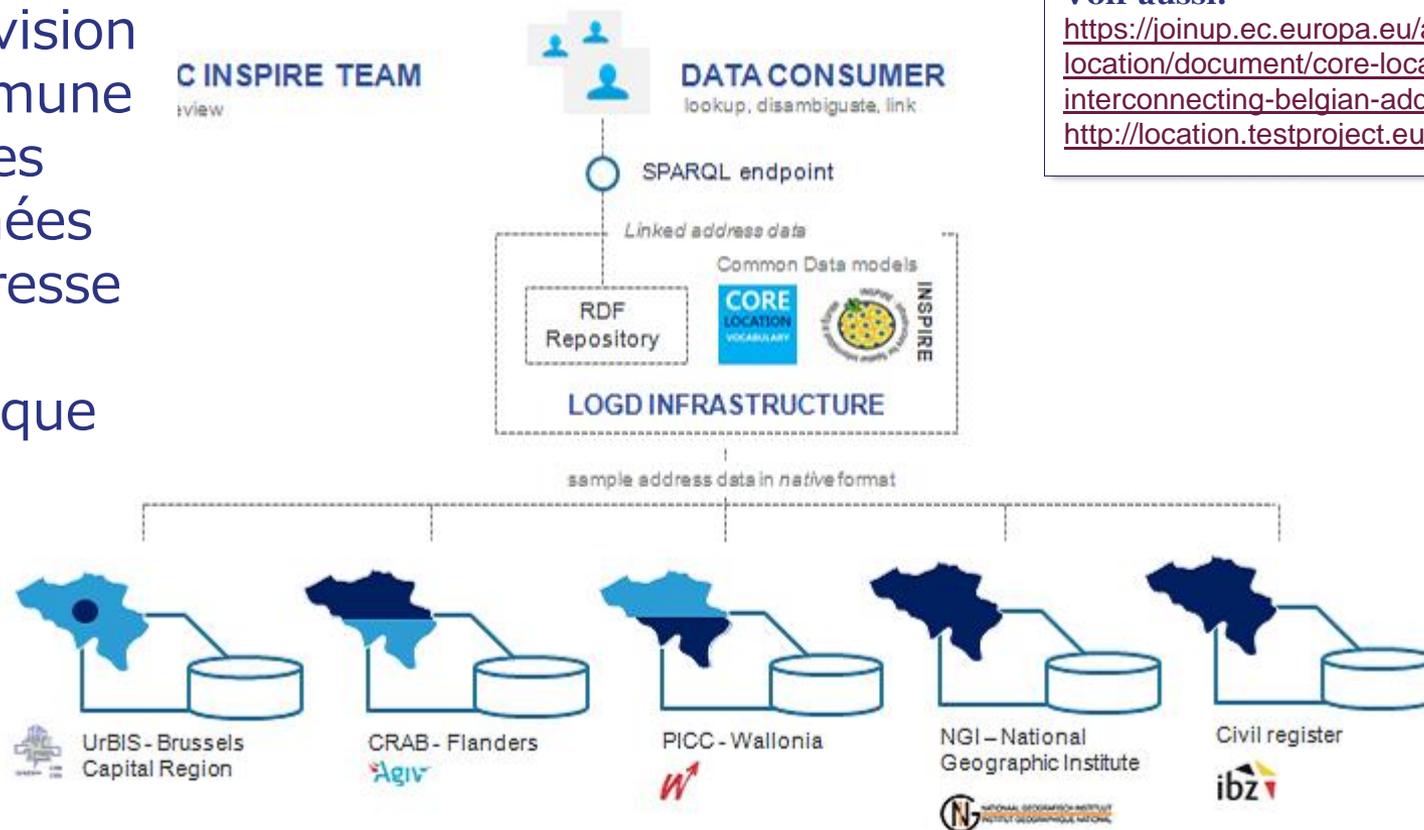
Voir aussi :

Etude ISA sur les modèles d'affaires LOGD

<https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Les données liées peuvent vous aider à publier des données structurées et intégrer des données de différentes sources

Construire une vision commune sur les données d'adresse en Belgique



Voir aussi:

<https://joinup.ec.europa.eu/asset/core-location/document/core-location-pilot-interconnecting-belgian-address-data>
<http://location.testproject.eu>

Obstacles aux LOGD

- Investissements nécessaires.
- Manque de compétences nécessaires.
- Manque d'outils.
- Absence de garanties de niveau de service.
- Licences manquantes, restrictives, ou incompatibles.
- Excès de vocabulaires normalisés.
- Inertie du statu quo - le changement se fait lentement.
- Missing, restrictive, or incompatible licences.

Voir aussi :

Etude ISA sur les modèles d'affaires LOGD

<https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd>

Les Initiatives de données liées en Europe

Quelques exemples supra-national, national, des initiatives régionales et privées dans le domaine des données liées (ouvertes) à travers l'Europe.

Initiatives des États membres - quelques exemples

DE – Bibliotheksverbund Bayern

Données liées provenant de 180 bibliothèques universitaires en Bavière, à Berlin et dans le Brandebourg.

IT – Agenzia per l'Italia digitale

Trois ensembles de données publiés sous forme de données liées: l'indice de l'administration publique, les contrats de la SPC pour les services Web et les systèmes de conduction et les classifications pour les données de l'administration publique.

NL – Building and address register

Le registre de base néerlandais des adresse et des bâtiments a été publié sous forme de données liées.

UK – Ordnance Survey

Trois produits open source de données ouvertes ont été publiés sous forme de données liées: 1:50 000 Scale Gazetteer, Code-Point Open et la géographie administrative en provenance de Boundary Line.

UK – Companies House

Publication sous forme de données liées de détails de base sur des sociétés à l'aide d'un simple URI pour chaque entreprise dans leur base de données.

Initiatives de données liées publiques et de métadonnées financées par la Commission Européenne



European Union Open Data Portal



DIRECTIVES

DIRECTIVE 2013/37/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 26 June 2013
amending Directive 2003/98/EC on the re-use of public sector information

Les pilotes de données liées publiques d'ISA

Linking data about applications and decisions for authorisation of plant protection products

PPP semantic asset

- PPP Ontology
- PPP Taxonomies

Sample queries

- Find the country where the product is authorised
- Find a product made with a given substance
- Find products made by a company
- Find the product to use on a given pest

Find out more about Linked Data

- Understanding Linked Data by example
- Case study on how Linked Data is transforming eGovernment
- Describe organizations in RDF with Core Business Vocabulary and ORG Ontology
- 10 Rules for Persistent URIs

Type a keyword:

SPARQL Query:

```

PREFIX sds: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX akso: <http://www.w3.org/2004/02/akso/core#>
PREFIX detsema: <http://purl.org/dc/terms/>
PREFIX sdsf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ppp: <http://ec.europa.eu/open-data/planthealth/ppp/>
SELECT DISTINCT ?s ?p ?o
FROM <http://health.testproject.eu/ppp/>
WHERE {
  ?s a ppp:Product;
  ?p ?o.
  FILTER(regex(?o,'Acanto','i')) .
}
LIMIT 100
    
```

Search using the [Faceted Browser](#)

This work is supported by [Action 1.1](#) of the [European Solutions for European Data Administrations \(ISA\)](#) Programme of the European Commission. [Click here to see the details.](#)

Linked Data pilots: [Core Business Vocabulary](#) / [Core Public Service Vocabulary](#) / [OpenAgriculture Ontology](#) / [E-Health Information Products](#) / [Healthcare Superlinks](#)

Linked maritime surveillance data

CISE semantic asset

- EU Fishing Fleet Ontology

Sample queries

- Find the fishing vessel from a given country
- Find the fishing vessel from a specific port
- Find all the AIS transmission for a fishing vessel
- Find incidents linked to a fishing vessel
- Find the trace for a fishing vessel

Find out more about Linked Data

- Understanding Linked Data by example
- Case Study on how Linked Data is transforming eGovernment
- Describe organizations in RDF with Core Business Vocabulary and ORG Ontology
- 10 Rules for Persistent URIs

Type a keyword:

SPARQL Query:

```

PREFIX sds: <http://maizline.testproject.eu/Fishing/ese/ontology/>
PREFIX sdsf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX detsema: <http://purl.org/dc/terms/>
SELECT ?(Fishing/ese/IRI) ?(Fishing/ese) ?(CountryCodeIRI) ?(Subsector(CountryCodeIRI)) ?(
  assem(CountryCodeIRI)) ?(
  (
    ?(Fishing/ese/IRI) a sds:Fishing/ese/
    sds:esse;
    detsema:spatial ?(CountryCodeIRI);
    detsema:title ?(Fishing/ese);
    FILTER(regex(?(Fishing/ese),'itanio','i')) .
  )
)
LIMIT 100
    
```

This work is supported by [Action 1.1](#) of the [European Solutions for European Data Administrations \(ISA\)](#) Programme of the European Commission. [Click here to see the details.](#)

Linked Data pilots: [Core Business Vocabulary](#) / [Core Public Service Vocabulary](#) / [OpenAgriculture Ontology](#) / [E-Health Information Products](#) / [Healthcare Superlinks](#)

<http://health.testproject.eu/PPP/>

Core Public Service Pilot: describe public services only once

Governments use local, regional and national access portals to give businesses, citizens, and public administrations basic information about their public services. Unfortunately, this basic information often duplicated, unstructured, and not machine-readable. This fragmentation makes it difficult for citizens, businesses, and public administrations to find information about the public service that they need. It also leads to situations where basically the same information about a public service is re-created in different applications or by different governments.

The [Core Public Service Vocabulary](#) allows public administrations to describe their service only once using a standard, extensible, and machine-readable vocabulary and make these descriptions re-used on many governmental access portals. This pilot implementation shows how this can be done using a Linked Data infrastructure for a small sample of public service descriptions originating from various European public administrations.

Sample public service descriptions

- From the [SPOCS large-scale pilot](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/ReaEstateAgencyLicence>
 - <http://cpsv.testproject.eu/dmbl/PublicService/CatenoEstablishmentLicence>
 - <http://cpsv.testproject.eu/dmbl/PublicService/ArchitectRegistration>
 - <http://cpsv.testproject.eu/dmbl/PublicService/TECRegistration>
 - <http://cpsv.testproject.eu/dmbl/PublicService/ArchitectRegistration>
- From the [Flemish Intergovernmental Product and Service Catalogue](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/24>
 - <http://cpsv.testproject.eu/dmbl/PublicService/117>
 - <http://cpsv.testproject.eu/dmbl/PublicService/179>
 - <http://cpsv.testproject.eu/dmbl/PublicService/279>
 - <http://cpsv.testproject.eu/dmbl/PublicService/291>
 - <http://cpsv.testproject.eu/dmbl/PublicService/201>
 - <http://cpsv.testproject.eu/dmbl/PublicService/246>
 - <http://cpsv.testproject.eu/dmbl/PublicService/347>
 - <http://cpsv.testproject.eu/dmbl/PublicService/415>
 - <http://cpsv.testproject.eu/dmbl/PublicService/588>
 - <http://cpsv.testproject.eu/dmbl/PublicService/932>

- From the [Irish Citizens Information portal](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/DivinaTest>
 - <http://cpsv.testproject.eu/dmbl/PublicService/EuropeanHealthInsuranceCard>
 - <http://cpsv.testproject.eu/dmbl/PublicService/PublicServiceCard>

- From the [e-CODEX large scale pilot](#):
- <http://cpsv.testproject.eu/dmbl/PublicService/SmallClaims>

<http://health.testproject.eu/CISE/>

<http://cpsv.testproject.eu/CPSV/> PORT

Les applications non-gouvernementales

The screenshot shows the BBC Music website. At the top, there are navigation links for Home, Showcase, Reviews, and Genres. The main content area is dedicated to 'The Beatles', with a large image of the band and the text 'Formed 1967 Disbanded 10 April 1970'. Below this, there are sections for 'Share This Page', 'BBC Music Showcase', and 'Latest Tracks Played On The BBC'. The tracks listed include 'Twist & Shout', 'Tomorrow Never Knows', and 'She Loves You'.

The screenshot shows the LinkedGeoData application interface. It features a search bar with the text 'berlin' and a map of Berlin with red markers indicating search results. A list of search results is displayed on the left, including 'Berlin, Deutschland, Europe' and 'Berlin, Stadt, Mitte, Berlin, Deutschland, Europe'. A detailed view of a specific result is shown on the right, with fields for Name, Description, Image, Source, and natural stone monument.

The screenshot shows the OpenCorporates website. The main headline reads 'We have information on 55,021,554 companies'. Below this, there is a search bar and a list of countries with the number of companies registered in each. The countries listed include Abu Dhabi (UAE), Alaska (US), Albania, Arizona (US), Aruba, Bahrain, Barbados, Bermuda, California (US), Canada, Connecticut (US), Croatia, District of Columbia (US), Dominican Republic, Dubai (UAE), Finland, Florida (US), Georgia (US), Gibraltar, Greenland, Iceland, Idaho (US), India, Iowa (US), Ireland, Isle of Man, Jamaica, Jersey, Liechtenstein, and Louisiana (US). The website also features a 'Recently updated corporate groupings' section and a 'NEW! Search officers/directors' section.

The screenshot shows the SIGMA Semantic Information Mashup website. The main headline reads 'SIGMA SEMANTIC INFORMATION MASHUP'. Below this, there is a search bar and a list of search results for 'Tim Berners-Lee'. The search results include 'Tim Berners-Lee', 'Tim Berners-Lee - Biography', 'Tim Berners-Lee facts', 'Tim Berners-Lee profile', 'Tim Berners-Lee 180 facts', 'Tim Berners-Lee 218 facts', 'Tim Berners-Lee 130 facts', 'Tim Berners-Lee 137 facts', and 'About: Timothy Berners-Lee'. The website also features a 'Sources' section and a 'Filter by jurisdiction' section.

Conclusions

- Les données liées sont un ensemble de principes de conception pour le partage de données lisibles par machine sur le Web.
- Les données liées et les données ouvertes ne sont pas la même chose.
- URI, RDF et SPARQL forment la couche de fondation pour les données liées.
- Les données liées offre un certain nombre d'avantages pour:
 - L'intégration des données avec peu d'impact sur les systèmes existants;
 - Catalyseur pour l'interopérabilité sémantique;
 - Permet la créativité et l'innovation à travers le contexte et la création de connaissances.

Question de groupe



<http://www.visualpharm.com>

Y a-t-il une offre et une demande pour les Données Publiques (Liées) Ouvertes dans votre pays?



<http://www.visualpharm.com>

Quels sont, d'après vous, les pièges et les bénéfices attendus des Données Liées?



<http://www.visualpharm.com>

Y a-t-il des initiatives de Données Liées (Ouvertes) dans votre pays? Si oui, combien d'étoiles leur donneriez-vous?

Passez le test en ligne [ici!](#)

Merci!
... et maintenant
vos questions?

This presentation has been created by Open Data Support

Disclaimers

1. The views expressed in this presentation are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission. The European Commission does not guarantee the accuracy of the information included in this presentation, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission. All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.
2. This presentation has been carefully compiled by PwC, but no representation is made or warranty given (either express or implied) as to the completeness or accuracy of the information it contains. PwC is not liable for the information in this presentation or any decision or consequence based on the use of it. PwC will not be liable for any damages arising from the use of the information contained in this presentation. The information contained in this presentation is of a general nature and is solely for guidance on matters of general interest. This presentation is not a substitute for professional advice on any particular matter. No reader should act on the basis of any matter contained in this publication without considering appropriate professional advice.

Authors:

Michiel De Keyzer, Nikolaos Loutas, Christophe Colas and Stijn Goedertier

Références

Slide 6:

- EUCLID. Course 1: Introduction and Application Scenarios. <http://www.euclid-project.eu/modules/course1>
- Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

Slides 8, 13, 36, 38:

- ISA Programme. Case study on how Linked Data is transforming eGovernment. <https://joinup.ec.europa.eu/community/semic/document/case-study-how-linked-data-transforming-egovernment>
- Tim Berners-Lee. Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>

Slide 9:

- <http://sig.ma/search?q=european+commission>

Slide 12:

- The Open Knowledge Foundation. Open Data – An Introduction. <http://okfn.org/opendata/>

Slides 18-28:

- 5 ★ Open Data. <http://5stardata.info/>

Slide 19:

- UK National Archives, Sustainable development targets 2011-12.

Slide 21:

- Data.gov.uk. Housing stock. <http://data.gov.uk/dataset/uk-housing-stock>

Slide 23:

- Data.gov.uk. Road Safety Data. <http://data.gov.uk/dataset/road-accidents-safety-data>

Slide 25&27:

- ISA Organization Ontology pilot - Linking public sector's organisational data, https://joinup.ec.europa.eu/asset/core_business/document/organization-ontology-pilot-linking-public-sectors-organisational-data
<http://data.ydmed.gov.gr/>

Slide 37:

- ISA Programme. Core Location Pilot - Interconnecting Belgian Address Data. https://joinup.ec.europa.eu/asset/core_location/document/core-location-pilot-interconnecting-belgian-address-data

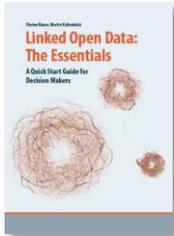
Slides 30-34:

- Open Refine: <https://github.com/OpenRefine>
- RDF Extension: <http://refine.deri.ie/>
- ISA Programme, Linking data about applications and decisions for authorisation of PPP, http://joinup.ec.europa.eu/asset/core_business/document/linking-data-about-applications-and-decisions-authorisation-ppp

Slide 40:

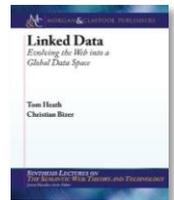
- Bibliotheksverbund Bayern, <http://lod.b3kat.de/doc>
- Agenzia per l'Italia Digitale, <http://spcdata.digitpa.gov.it/data.html>
- NL – Building and address register, <http://lod.Geodan.nl>
- UK Ordnance Survey, <http://data.ordnancesurvey.co.uk/>
- UK Companies House, <http://companieshouse.gov.uk/>

Lecture supplémentaire



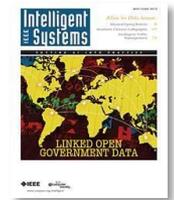
Linked Open Data: The Essentials. Florian Bauer, Martin Kaltenböck.

<http://www.semantic-web.at/LOD-TheEssentials.pdf>



Linked Data: Evolving the Web into a Global Data Space. Tom Heath and Christian Bizer.

<http://linkeddatabook.com/editions/1.0/>



Linked Open Government Data. Li Ding Qualcomm, Vassilios Peristeras and Michael Hausenblas.

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6237454>



EUCLID - Course 1: Introduction and Application Scenarios

<http://www.euclid-project.eu/modules/course1>

Projets et initiatives apparentés



LOD2 FP7 project, <http://lod2.eu/>



The Open Knowledge Foundation, <http://okfn.org/>



W3C Semantic Web, <http://www.w3.org/standards/semanticweb/>



EUCLID, <http://projecteuclid.org/>



ISA Programme, <http://ec.europa.eu/isa/>



W3C LOGD WG, http://www.w3.org/2011/gld/wiki/Main_Page



LOD Around The Clock FP7 project, <http://latc-project.eu/>



Data.gov.uk, <http://data.gov.uk/linked-data>



Faites partie de notre équipe ...

Retrouvez-nous



Open Data Support

<http://www.slideshare.net/OpenDataSupport>



Open Data Support

<http://goo.gl/y9ZZI>

Suivez-nous



@OpenDataSupport

Rejoignez-nous



joinup

<http://www.opendatasupport.eu>

Contactez-nous

contact@opendatasupport.eu