

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

Training Module 2.1

Le cycle de vie des Données et Métadonnées Publiques Liées Ouvertes

Les objectifs de cette formation

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

- Ce qu'est un cycle de vie des pour des données publiques liées ouvertes gouvernemental (LOGD).
- La différence entre l'offre et la demande de données.
- Les différentes étapes du cycle de vie des LOGD.
- Les outils et les meilleures pratiques associées à chaque étape du cycle de vie.

Contenu

Ce module contient ...

- Un aperçu du cycle de vie des données existantes pour les données liées ouvertes gouvernemental (LOGD).
- Un cycle de vie hybride pour les LOGD et les métadonnées, couvrant à la fois l'offre et la demande.
- Une vue d'ensemble des technologies existantes pour les LOGD et métadonnées - y compris la « Open Data Interoperability Platform » (plateforme d'interopérabilité Open Data) (ODIP).

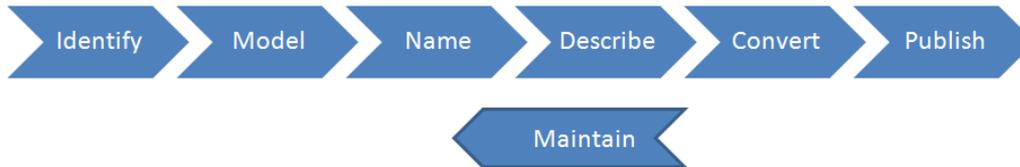
Différents cycles de vie LOGD

L'état des lieux

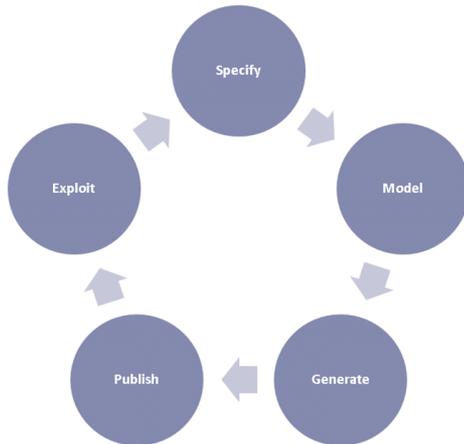
Voir aussi:

http://www.w3.org/2011/gld/wiki/GLD_Life_cycle

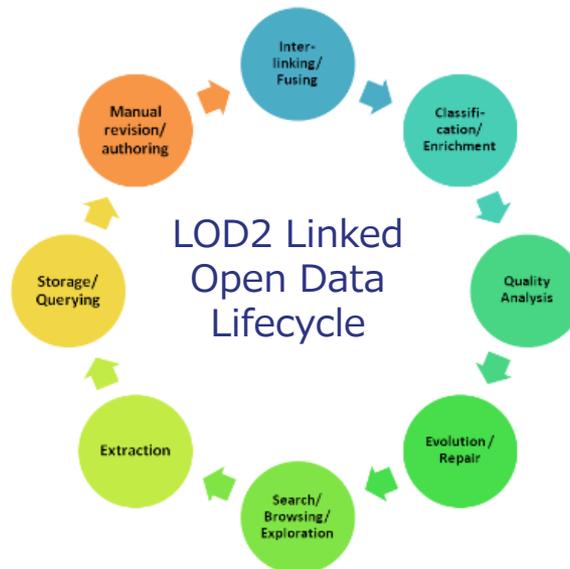
Hyland et al.



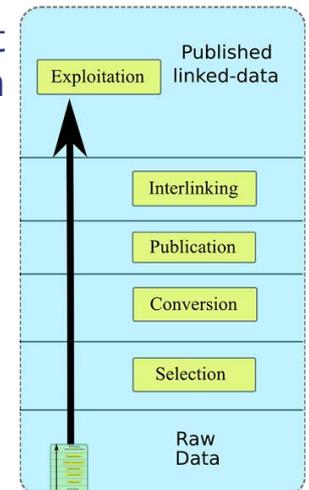
Hausenblas et al.



Villazon-Terrazas et al.



Datalift Vision



Différents cycles de vie LOGD

Observations

- Pas de cycle de vie LOGD standardisé.
- La plupart des approches s'accordent sur un ensemble de phases de base, comme par ex: identifier le modèle, et publier.
- Les cycles de vie actuels se concentrent de fournir « open data »:
 - Identification et sélection des LOGD.
 - Modélisation et nettoyage des LOGD.
 - Publication et liaison des données.
- Mais que dire du côté de la demande?
 - Trouver et extraire des LOGD.
 - Intégrer et réutiliser des données ouvertes.
 - Fournir un retour d'information sur les LOGD.

Qu'est ce que les métadonnées?

“Les métadonnées sont de l'information structurée qui décrit, explique, localise ou facilite autrement l'obtention, l'utilisation ou la gestion d'une ressource d'information. Les métadonnées sont souvent appelées données des données ou information sur l'information.”

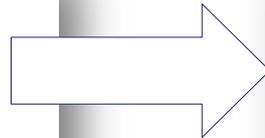
-- National Information Standards Organization

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Description de l'ensemble des données (DCAT)

```
:weather1-7 a dcat:Dataset ;
dct:title "Measurements from weather stations 1-7" ;
dct:description "Data from seven weather stations
                showing temparture, humidity,
                wind direction and wind speed" ;
dct:modified "2013-07-01" ;
dct:publisher <http://myweather.com/id/myweather> ;
dcat:keyword "weather" ;
dcat:landingpage <http://myweather.com/stations1-7.html> ;
dcat:distribution :weatherdata-xlsx
.

:weatherdata1-7-xlsx a dcat:Distribution ;
dct:format <http://publications.europa.eu/resource/authority/file-type/XLSX> ;
dct:licence <http://creativecommons.org/licenses/cc0> ;
dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```



Ensemble des données

	Temp. °C	Humidity %	Wind direction	Wind speed km/h
Station 1	18.1	60	WSW	18
Station 2	17.5	59	WSW	20
Station 3	18.2	55	SW	22
Station 4	19.0	62	SW	18
Station 5	18.0	65	WSW	19
Station 6	18.2	63	SSW	21
Station 7	17.9	61	SW	22

Les meilleures pratiques pour la publication de vos données et métadonnées

Livre de cuisine du W3C pour les données liées

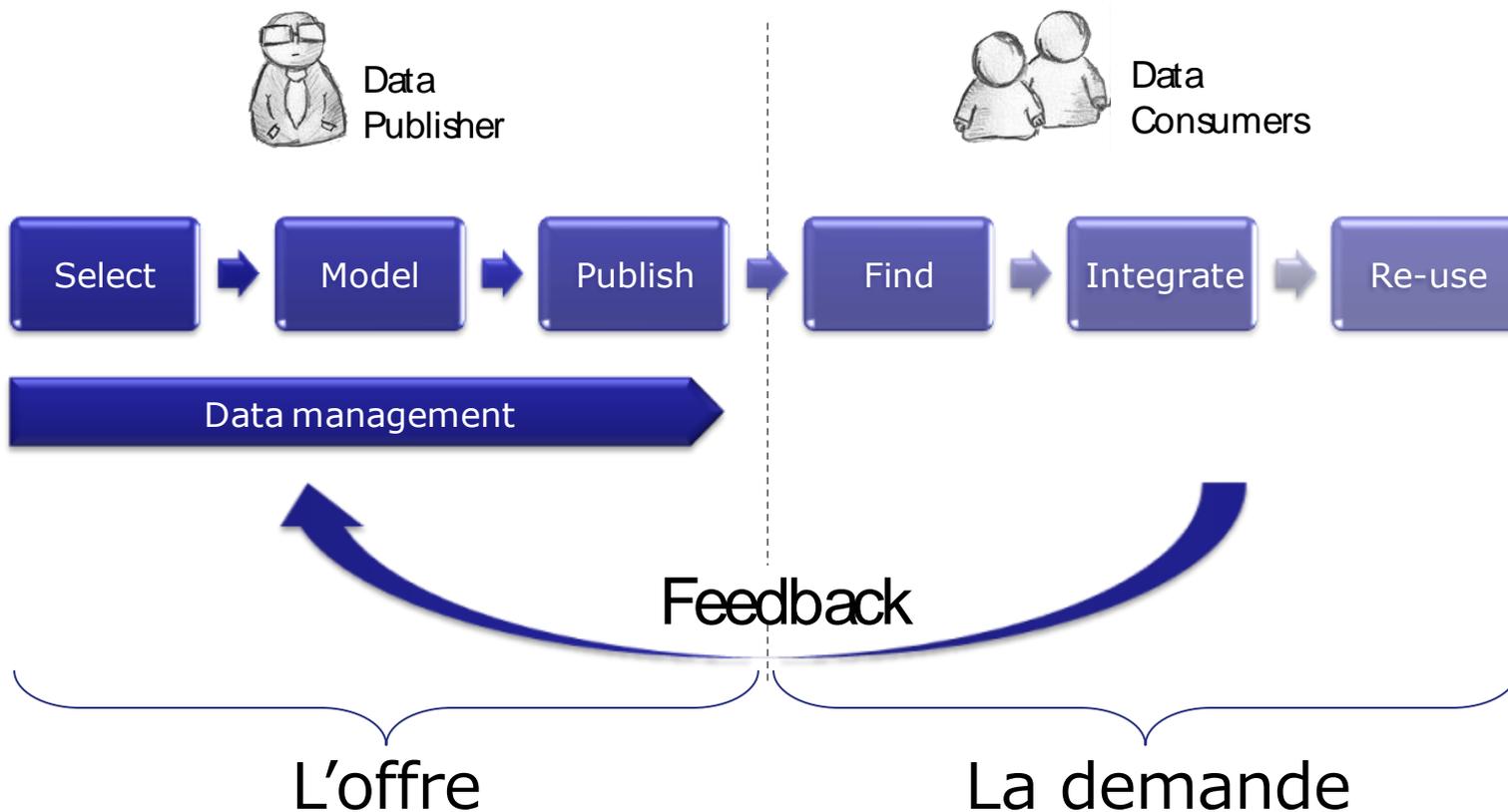
1. Modèle des données;
2. Nomer les choses avec des URIs;
3. Réutiliser les vocabulaires existants dans la mesure du possible;
4. Publier des descriptions lisibles par l'homme et la machine - **métadonnées**;
5. Convertir les données en RDF;
6. Spécifier une licence appropriée;
7. Héberger l'ensemble des données liées et leurs métadonnées publiquement et l'annoncer!

Voir aussi:

<http://www.w3.org/TR/gov-data/>

http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

LOGD et le cycle de vie des métadonnées avec l'accent sur l'offre et la demande



LOGD & l'offre de metadonnées

Les gouvernements rendent publiques leurs données et les publient sous forme de données liées accompagnées de descriptions métadonnées appropriées.

Sélection des données à haute valeur ajoutée

Plusieurs dimensions peuvent être prises en compte dans le processus de sélection des données publiques liées ouvertes, à la fois du point de vue du publicateur et du ré-utilisateur:

- **Transparence:** La publication de l'ensemble de données augmente-t-il la transparence et l'ouverture du gouvernement envers ces citoyens?
- **Les exigences légales:** Y a-t-il une loi qui rend obligatoire la publication ouverte ou n'y a-t-il pas d'obligation spécifique?
- **Relation avec mission de service public:** Est-ce que les données sont le résultat direct issu de la mission publique primaire du gouvernement ou sont-elles le produit d'une activité non essentielle?
- **Situation actuelle de la publication ouverte:** Est-ce que les données sont déjà disponibles ouvertement ou doit-il encore être ouvert?
- **Type de valeur:** Les données sont-elles utiles pour l'engagement social ou ont-t-elles une valeur commerciale?
- **Audience:** Est-ce que les données sont principalement destinées au public ou sont-elles destinées principalement à l'intégration « back-office »?

Sélection basée sur la transparence

Dans certains cas, la publication de l'ensemble de données augmente la transparence et l'ouverture du gouvernement envers ces citoyens, par ex.:

- Données parlementaires, telles que les résultats des votes.
- La façon dont est dépensé le budget gouvernemental.
- Coûts de personnel des administrations publiques

Tout les exemples sus-mentionnés contribuent à la transparence du fonctionnement des administrations publiques.

Sélection basée sur les exigences juridiques

Certaines données peuvent être couvertes par une loi ou un règlement qui impose leur publication de manière ouverte, par ex:

- Texte de lois, directives, règlements, etc
- Les propositions et les délibérations du Parlement et de comités.
- Les résultats des élections.
- Les budgets et dépenses publique.
- Les appels d'offres et l'attribution des contrats.

D'autres données peuvent être le sous-produit de l'activité gouvernementale et il serait utile aux citoyens et aux entreprises d'y avoir accès, par ex:

- État des infrastructures et espaces publics (routes, arbres).
- Horaires des transports en commun et horaires de collecte des ordures.

Sélection basée sur la relation avec les missions de service public

Certaines données peuvent être le résultat direct d'une fonction primaire du gouvernement, par exemple les fonctions énumérées dans la COFOG:

- Les organes législatifs et exécutifs, les affaires financières ou fiscales etc.
- L'ordre public et la sécurité.
- La protection de l'environnement.
- La santé.
- La culture.
- L'éducation.

D'autres données produites par le gouvernement qui sont non essentielles (elles peuvent être - et sont parfois - fournies par le secteur privé), par ex:

- La cartographie pour la navigation (cf. Google Street View)
- Les prévisions météorologiques (cf. Weather Channel)

Sélection basée sur le statut de l'ouverture

Certaines données sont déjà publiées ouvertement et par voie électronique, par ex: (dans certains pays):

- Les informations cadastrales.
- Les cartes topographiques.
- Les informations de trafic.
- Les prévisions météorologiques.

D'autres données peuvent encore être cachées au public (peut-être parce qu'elles sont difficiles à publier ou comportent des données personnelles, des données sensibles ou en partie soumises à des licences de tiers).

Sélection basée sur le type de valeur

Certaines données peuvent avoir une valeur essentiellement sociale, par ex:

- Les lois et les données parlementaires (par ex: les votes exprimés par les représentants)
- L'information préélectorale (par ex: les programmes des partis politiques)
- e-démocratie et e-participation (par ex: les consultations publiques)

*D'autres données peuvent avoir plus **de valeur commerciale** (modèle d'affaires), par ex:*

- Les cartes routières et l'information trafic en temps réel
- L'information météo en temps réel
- L'information économique

Sélection basée sur le type de public

*Certaines données sont destinée à la **société** (citoyens et entreprises), par ex:*

- Informations légales.
- e-démocratie, e-participation et consultation publique.
- Marchés publics.

*D'autres données sont axées sur l'utilisation interne ou pour l'intégration dans le « **back-office** », par ex:*

- Diverses sources qui sont utilisées pour l'application de la loi.
- Indicateurs de performance de service.
- Les descriptions de fonctions publiques

Sélection basée sur la taille de l'audience

*Certaines données sont destinées à un **large public et les marchés de masse**, par ex:*

- Les informations sur le trafic.
- Transports en commun.
- Données électorales.

*D'autres données sont essentielles pour de **petits groupes de personnes et des marchés de niche**, par exemple:*

- Informations sur les installations et le soutien financier pour les personnes ayant des besoins spéciaux.
- Statistiques économiques.
- Décisions de justice.

Forte value du point de vue des ré-utilisateurs

*Du point de vue d'un ré-utilisateur, la valeur d'un ensemble de données dépend principalement de son **potentiel d'utilisation et de ré-utilisation**, ce qui peut amener à créer de (nouveaux) **modèles d'affaires**.*

Le potentiel d'utilisation et de ré-utilisation d'un ensemble de données est défini par:

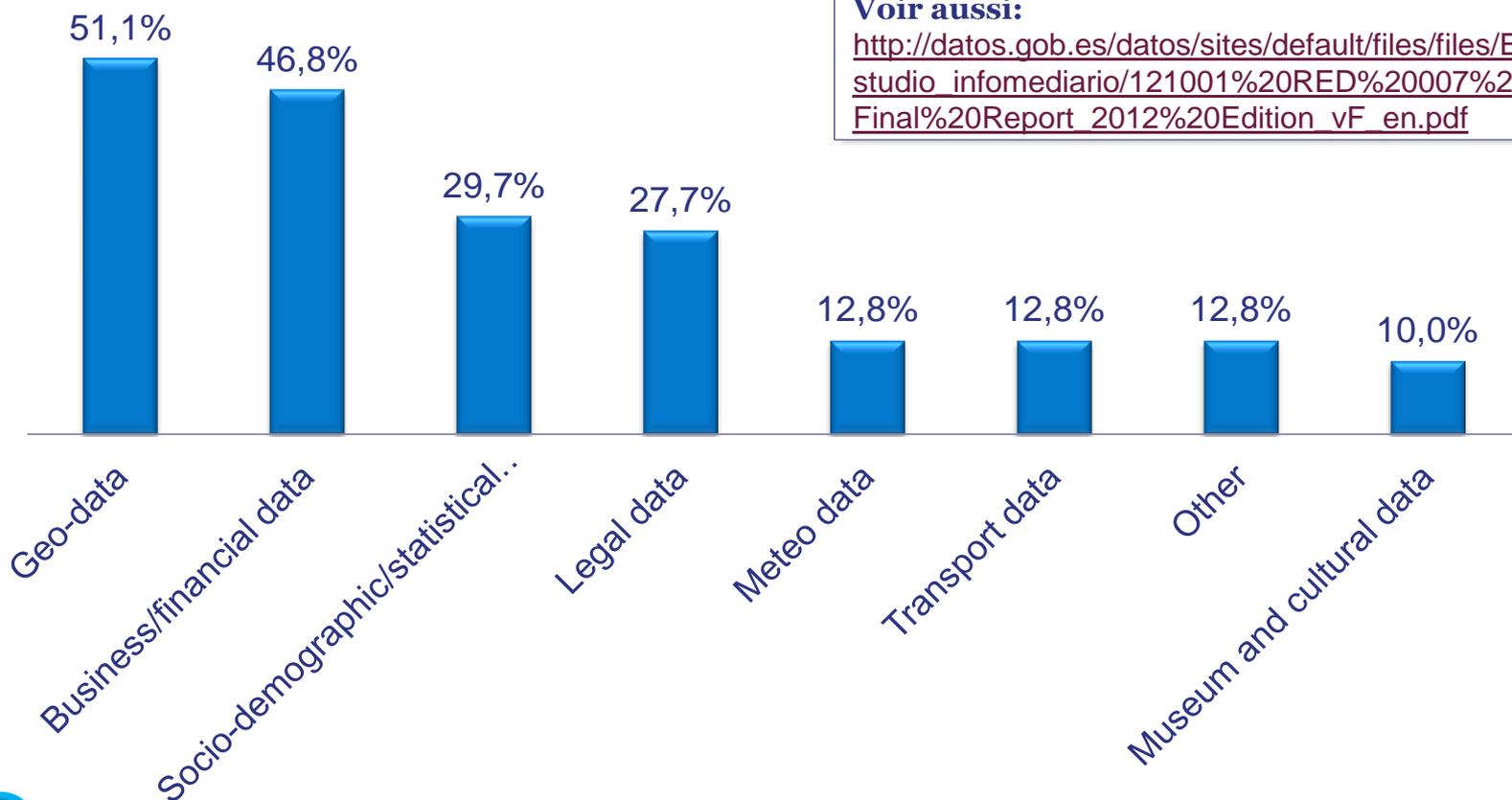
- La taille et le dynamisme de la population cible de l'ensemble de données; et
- Le nombre de systèmes et services, nouveaux et existants, qui utilisent l'ensemble de données.

*L'ouverture d'ensembles de données à fort potentiel d'utilisation et de ré-utilisation amène à la **création de nouveaux produits et/ou services** qui ont un **impact économique ou social** direct ou indirect, et/ou des **externalités économiques positives**.*

Sélection basée sur les besoins de l'audience

Quelles données ont les réutilisateurs besoin / envie?

Selon une étude espagnole, les types d'informations suivants sont les plus réutilisés par les entreprises:



Voir aussi:

http://datos.gob.es/datos/sites/default/files/files/Estudio_infomediario/121001%20RED%20007%20Final%20Report_2012%20Edition_vF_en.pdf

Types d'ensembles de données sur les portails de données européens

Tags

[View all tags](#)

population local-government care disclosure s
health children health-and-social
transparency communities Spend
health-well-being-and-care **child finance** educa
NERC_DDC nhs spend-transaction

Source:[<http://data.gov.uk/data>]

Search Europe's Public Data

Find datasets

- Finance and Budgeting (436)
- Social Questions (226) Environment (215)
- Transportation (196)
- Education and Communication (188)
- Agriculture, Fisheries, Forestry (176)
- Population (145)
- Economy and Industry (114) Health (79)

Source:[<http://publicdata.eu/>]

Keywords

- geospatial data (82)
- international trade (81)
- agriculture (77)
- environmental statistics (51)
- water (44)
- vector data (41)
- agricultural structure (39)
- balance of payments (38)
- EEA owned data sets (38)
- raster data (34)

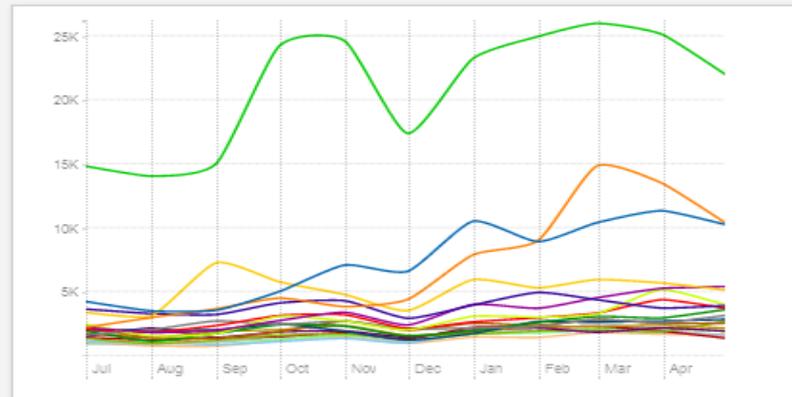
Source:[<http://open-data.europa.eu/en/data/dataset>]

Metadonnées



Quels ensembles de données sont principalement consultés sur Data.gov.uk

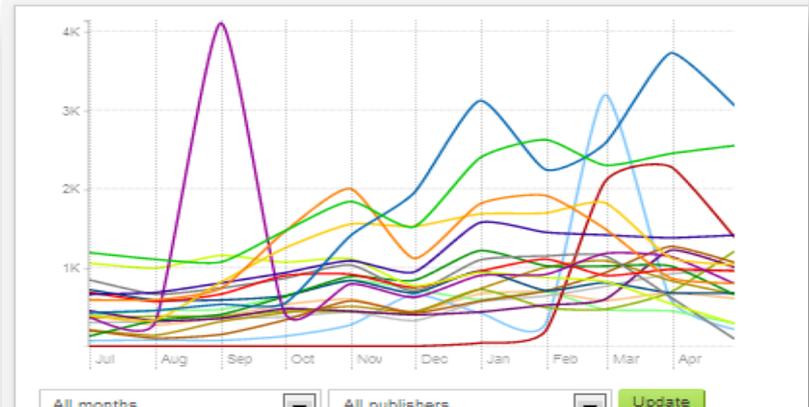
Par éditeur



Statistics for

Publisher	Dataset Views
Office for National Statistics	230861
Department for Communities and Local Government	82783
NHS Information Centre for Health and Social Care	79447
Department for Transport	56362
Department for Environment, Food and Rural Affairs	42654
Department for Business, Innovation and Skills	41579
Home Office	33115
Department of Health	31702
British Geological Survey	28812
Department for Children, Schools and Families	25635
Ordnance Survey	24427
Environment Agency	24039
Cabinet Office	21773
Transport Direct	20471
Department for Education	18744
Vehicle and Operator Services Agency	18569
Department of Energy and Climate Change	17973
Department for Work and Pensions	16719
Ministry of Defence	15520
Ministry of Justice	14203

Par ensemble de données



Statistics for all months:

Dataset	Views	Downloads
Lower Layer Super Output Area (LSOA) boundaries	20787	4203
English Indices of Deprivation 2010	20447	5752
UK Tariff Codes	13474	2907
Social Trends	13340	3020
Building Price and Cost Indices	12498	3259
Road Safety Data	11682	6355
Organogram and staff pay data for DEFRA	9622	2202
National Public Transport Access Nodes (NaPTAN)	9488	1758
Open Access Non-VOSA Sites	8937	1236
Average Earnings Index	8197	1355
Organogram and staff pay data for Environment Agency	8003	1913
Regional Household Income	6445	2392
Energy Consumption in the UK	6437	1717
NHS Payment by Results 2010-11 National Tariff Information	6407	1625
VAT registered businesses	6179	363
Bona Vacantia Unclaimed Estates and Adverts	6160	11930
Active Vehicle Testing Stations in Great Britain	6054	1627
Population estimates by broad age band	6050	882
GP practice prescribing data - Chemical level	5990	1292
National Public Transport Data Repository (NPTDR)	5901	1452

La modélisation de vos données et métadonnées a pour but de...

- Rendre vos données disponibles d'une manière **structurée, compréhensible et lisible par machine**.
- **Réutiliser** ce qui existe déjà en termes de vocabulaires et des données de référence.
- Atteindre un bon niveau de qualité en **nettoyant** vos données.
- Fournir des **informations sur les licences** afin que les consommateurs de données connaissent les conditions de réutilisation.
- Fournir une description riche (**métadonnées**).
- Utiliser les technologies sémantiques (RDF, HTTP URIs ...) pour décrire vos données.

Modéliser vos données - réutiliser si possible, créer si nécessaire

- **Réutiliser** les vocabulaires existants **autant que possible**.
 - Si vous constatez qu'il n'existe aucune source officielle de référence réutilisable pour le domaine spécifique, **créez la vôtre en utilisant**:
 - Schéma RDF (RDFS): vocabulaire RDF de base pour décrire les classes et les propriétés des classes.
 - Langage d'ontologie Web (OWL): langage de représentation du savoir pour décrire les ontologies.

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/model-your-data-metadata>
<http://www.w3.org/TR/owl-features/>
<http://www.w3.org/TR/rdf-schema/>

Réutiliser les vocabulaires communs pour modéliser et décrire vos données ... (en RDF)



Dublin Core
Metadata Initiative®

Vocabulaires d'usage général: DCMI, RDFS



Pour nommer des ressources: rdfs:label, foaf:name, skos:prefLabel



Pour décrire des personnes: FOAF, vCard, Core Person Vocabulary



Pour décrire des organisations enregistrés: Registered Organisation Vocabulary



Pour décrire des adresses: vCard, Core Location Vocabulary



Pour décrire des services publics: Core Public Service Vocabulary



... et les métadonnées ...



Pour décrire des ensembles de données (métadonnées) : DCAT, DCAT Application Profile, VoID



Pour décrire des projets: DOAP, ADMS.SW

Pour décrire des ressources d'interopérabilité: ADMS

Trouver des vocabulaires réutilisables

Joinup

- Plateforme en ligne pour la recherche et le partage de ressources d'interopérabilité décrites avec ADMS.

- Développée par le Programme ISA de la Commission Européenn

More targeted.

Refine the search results via the faceted search filters.

1 Enter a search keyword to find interoperability assets available on different websites.

2

3

More focused.

More targeted.

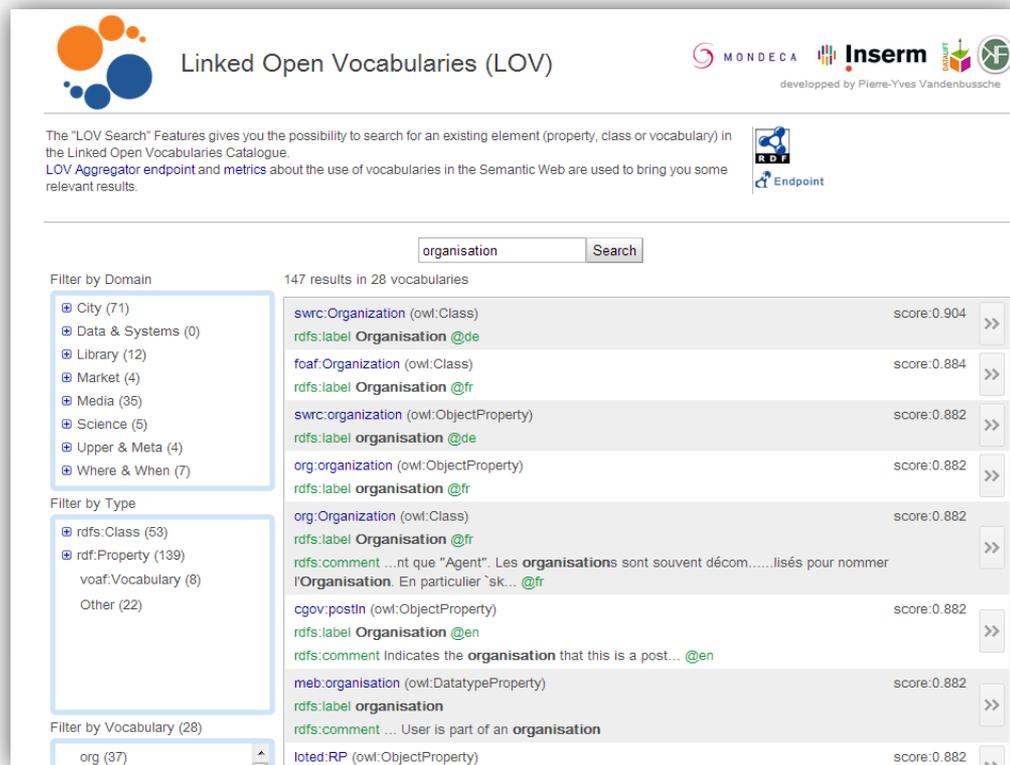
More relevant.

<http://joinup.ec.europa.eu/>

Trouver des vocabulaires réutilisables

“Linked Open Vocabularies”

- Fournit des méthodes d'accès facile à l'écosystème de vocabulaires.
- Permet de relier les uns aux autres de manière explicite.
- Fournit des mesures sur la façon dont ils sont utilisés dans le nuage LOGD.
- Développé par “Open Knowledge Foundation.”



The "LOV Search" Features gives you the possibility to search for an existing element (property, class or vocabulary) in the Linked Open Vocabularies Catalogue. LOV Aggregator endpoint and metrics about the use of vocabularies in the Semantic Web are used to bring you some relevant results.

organisation Search

Filter by Domain

- City (71)
- Data & Systems (0)
- Library (12)
- Market (4)
- Media (35)
- Science (5)
- Upper & Meta (4)
- Where & When (7)

Filter by Type

- rdfs:Class (53)
- rdf:Property (139)
- voaf:Vocabulary (8)
- Other (22)

Filter by Vocabulary (28)

- org (37)

147 results in 28 vocabularies

swrc:Organization (owl:Class)	score:0.904	>>
rdfs:label Organisation @de		
foaf:Organization (owl:Class)	score:0.884	>>
rdfs:label Organisation @fr		
swrc:organization (owl:ObjectProperty)	score:0.882	>>
rdfs:label organisation @de		
org:organization (owl:ObjectProperty)	score:0.882	>>
rdfs:label organisation @fr		
org:Organization (owl:Class)	score:0.882	
rdfs:label Organisation @fr		
rdfs:comment ...nt que "Agent". Les organisations sont souvent décom...lisés pour nommer l'Organisation. En particulier "sk... @fr		>>
cgov:postIn (owl:ObjectProperty)	score:0.882	
rdfs:label Organisation @en		
rdfs:comment Indicates the organisation that this is a post... @en		
meb:organisation (owl:DatatypeProperty)	score:0.882	>>
rdfs:label organisation		
rdfs:comment ... User is part of an organisation		
loted:RP (owl:ObjectProperty)	score:0.882	>>

<http://lov.okfn.org/>

Le nettoyage de vos données et métadonnées

Pour s'assurer que les données et les métadonnées peuvent être publiées avec un niveau de qualité approprié et un minimum d'erreurs.

Cela signifie:

- Corriger les erreurs.
- Transformer / homogénéiser les formats.
- Aligner les incohérences dans les données et métadonnées.
- Supprimer les informations dupliquées / redondantes.
- Ajouter les informations manquantes.
- Vérifier que les informations soient à jour.

Voir aussi:

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

Cleanse your data with Open Refine (Google Refine) -

<https://code.google.com/p/google-refine/>

Le nettoyage des données - exemple

Duplicata

Problème de formatage

erreur

Company name	Registration date	Country	E-mail	# Establishments
Nikè	1991-04-28	Belgium	niké	7
BARCO	15 September 1986	BE	Barco@email.be	2
Nikè		België		
Coca-Cola		United States	coca@cola.com	3

Informations manquantes

Incohérences

Information redondantes

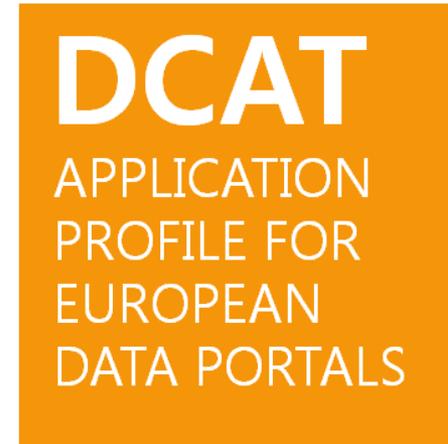
Opérations de nettoyage

Company name	Registration date	Country	E-mail
Nikè	1991-04-28	BE	niké@sport.org
BARCO	1986-09-05	BE	Barco@email.be
Coca-Cola	1964-03-26	US	coca@cola.com

Modéliser vos métadonnées

Le profil d'application DCAT pour les portails de données en Europe (DCAT-AP) est une spécification basée sur le « Data Catalogue vocabulary » (DCAT) pour décrire des ensembles de données du secteur public en Europe.

DCAT-AP améliore la découverte des ensembles de données du secteur public au delà des frontières et à travers les secteurs.



Voir aussi:

https://joinup.ec.europa.eu/asset/dcat_application_profile/description

Utiliser les identifiants persistants uniformes de ressources (URI) pour nommer des ressources

Des URI persistants définissent les bases des données liées.



Follow the pattern

e.g. `http://{domain}/{type}/{concept}/{reference}`

Re-use existing identifiers

e.g. `http://education.data.gov.uk/id/school/123456`

Link multiple representations

e.g. `http://data.example.org/doc/foo/bar.html`

e.g. `http://data.example.org/doc/foo/bar.rdf`

Implement 303 redirects for real-world objects

e.g. `http://www.example.com/id/alice_brown`

Use a dedicated service

i.e. independent of the data originator

10 rules for persistent URIs



Avoid stating ownership

e.g. `http://education.data.gov.uk/ministryofeducation/id/school/123456`

Avoid version numbers

e.g. `http://education.data.gov.uk/doc/school/v1/123456`

Avoid using auto-increment

e.g. `http://education.data.gov.uk/id/school1/123456`

e.g. `http://education.data.gov.uk/id/school1/123457`

Avoid query strings

e.g. `http://education.data.gov.uk/doc/school?id=123456`

Avoid file extensions

`http://education.data.gov.uk/doc/schools/123456.cs`

Voir aussi:

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

<https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

Attribuer une licence à vos données et métadonnées a pour but de...

- Informer des utilisateurs potentiels sur la façon dont les **données et les métadonnées** peuvent être (ré) utilisées et / ou adaptées.
- Ne pas lier vos **données et métadonnées** avec des informations de licence est un obstacle important pour la réutilisation et réduit donc la valeur que peut créer l'ouverture de vos données.
- Les données ouvertes doivent être, par définition, publiées sous une **licence ouverte**.
- Les métadonnées devrait être publiée sous une licence indiquant qu'elles font partie du domaine public pour renforcer la réutilisation et faciliter la recherche de vos données.

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/licence-your-data-metadata>

Les licences ouvertes

Voir aussi:

http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf

- Creative Commons (CC) (<http://creativecommons.org/licenses/>)
 - Paternité (BY): Le créateur de l'œuvre doit être mentionné.
 - Non Commercial (NC): Le travail ne peut pas être utilisé à des fins commerciales.
 - Pas de produits dérivés (ND): Les travaux ne peuvent pas être adaptés ou fusionnés avec d'autres travaux.
 - Share Alike (SA): Le travail peut être adapté, mais doit être attribué la même licence si vous le mettez à disposition.
 - CC Zero (CC0): Le travail est du domaine public.
- Open Data Commons (<http://opendatacommons.org/licenses/>)
 - Open Data Commons Attribution Licence (ODC-By): compatible avec CC BY
 - Open Data Commons Open Database Licence (ODC-ODbL): compatible avec CC BY SA)
 - Public Domain Dedication Licence (PDDL): compatible avec CC Zero
- The Open Government Licence (<http://www.nationalarchives.gov.uk/doc/open-government-licence/>)

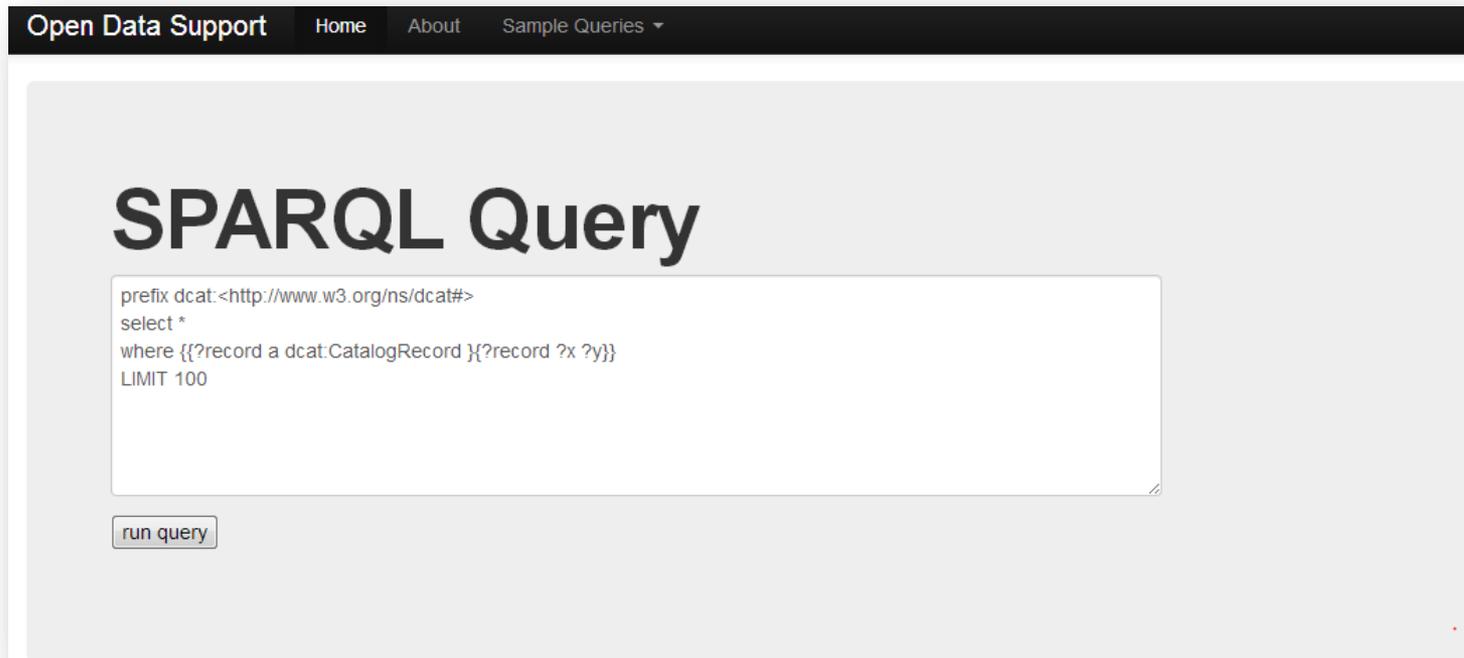
La pose d'une licence sur vos données et métadonnées a pour but de...

Briser les murs des silos afin de créer plus de valeur.

- Rendre vos données et métadonnées publiquement et facilement accessibles sur le Web.
- Lier vos données et métadonnées à d'autres données (ou métadonnées) dans le but de:
 - Donner du sens et du contenu à celles-ci.
 - Donner un contexte.
 - Les enrichir.
 - Permettre aux gens de découvrir d'autres choses.

Fournir un “SPARQL Endpoint”

Le “SPARQL Endpoint” est un service qui permet à d'autres d'interroger vos données liées (et / ou métadonnées).



The screenshot shows the Open Data Support website's SPARQL Query interface. The navigation bar at the top includes 'Open Data Support', 'Home', 'About', and 'Sample Queries'. The main heading is 'SPARQL Query'. Below it is a text input field containing a SPARQL query:

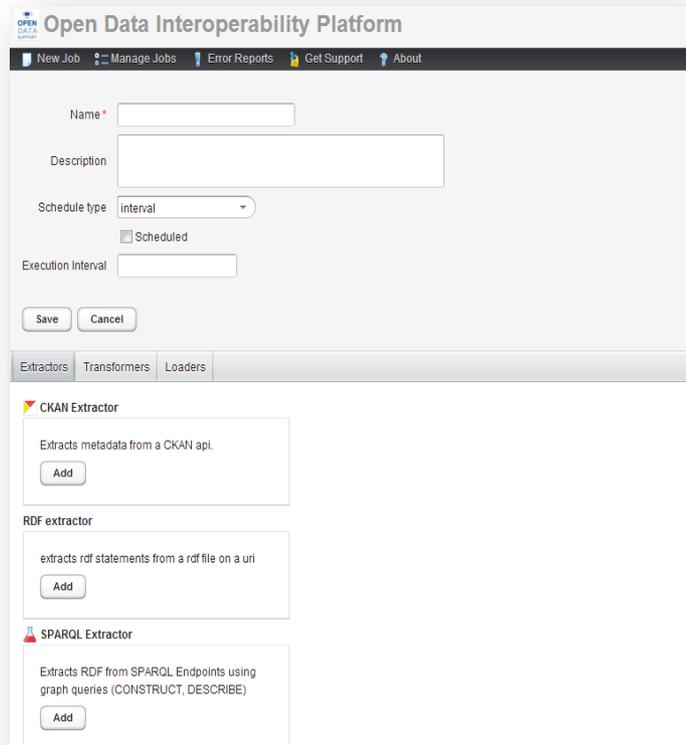
```
prefix dcat:<http://www.w3.org/ns/dcat#>
select *
where {{?record a dcat:CatalogRecord }}{{?record ?x ?y}}
LIMIT 100
```

 A 'run query' button is located below the input field.

<http://data.opendatasupport.eu>

Publier vos métadonnées

Publiez vos métadonnées sur une plateforme centrale de partage des données pour lui donner plus de visibilité et augmenter la réutilisation de vos bases de données.



The screenshot shows the Open Data Interoperability Platform (ODIP) interface. At the top, there is a navigation bar with links for 'New Job', 'Manage Jobs', 'Error Reports', 'Get Support', and 'About'. Below this is a form for creating a new job. The form includes a 'Name' field, a 'Description' field, a 'Schedule type' dropdown menu (set to 'interval'), a 'Scheduled' checkbox, and an 'Execution Interval' field. There are 'Save' and 'Cancel' buttons. Below the form, there are tabs for 'Extractors', 'Transformers', and 'Loaders'. Under the 'Extractors' tab, there are three sections: 'CKAN Extractor' (Extracts metadata from a CKAN api), 'RDF extractor' (extracts rdf statements from a rdf file on a uri), and 'SPARQL Extractor' (Extracts RDF from SPARQL Endpoints using graph queries (CONSTRUCT, DESCRIBE)). Each section has an 'Add' button.

The Open Data Interoperability Platform (ODIP):

- ODIP est une plateforme centrale de partage des données développée par la Commission européenne pour permettre la recherche européenne d'ensembles de données de manière transfrontalière.
- Des éditeurs de données ODIP et des portails de données pour publier centralement les métadonnées descriptives pour les ensembles de données.

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/promoting-the-re-use-of-open-data-through-odip>

La gestion des données et des métadonnées a pour but de...

- Gérer un cycle de vie de données – créer, mettre à jour, de garder a jour et mise au rebut d'ensemble de données.
- Gestion du cycle de vie des métadonnées.
- Mettre en place des processus pour s'assurer que vos données et métadonnées a un niveau de qualité approprié.
- Indiquant la propriété des données (ensembles) et des métadonnées.

Voir aussi:

<http://www.slideshare.net/OpenDataSupport/introduction-to-metadata-management>

Recueillir les commentaires des réutilisateurs de vos données

Demandez leur avis aux utilisateurs (potentiels) de données:

- Quels sont les données dont ils ont besoin?
- Comment ont-ils utilisé les données?
- Que pensent-ils de la qualité?
- Assurez-vous que les demandes et les corrections arrivent chez vous – communautarisez le contrôle de la qualité des données!

Behoeftte aan meer data?

Via onderstaand formulier kunt u een verzoek om overheidsdata insturen. We zullen dan bemiddelen tussen u en de overheidsinstantie van wie u data zou willen hebben. We kunnen geen garantie geven dat een aanvraag ook leidt tot het beschikbaar komen van data.

Voor vragen kunt u contact opnemen met paul.suijkerbuijk@minbzk.nl.

We houden ons aan de [e-mailgedragslijn voor overheden](#) van burger@overheid.nl. Lees hier hoe Overheid.nl met uw [persoonsgegevens](#) omgaat. Velden aangegeven met een * zijn verplicht.

Uw naam *

Uw organisatie *

data.overheid.nl



data.gov.uk

Demande de LOGD

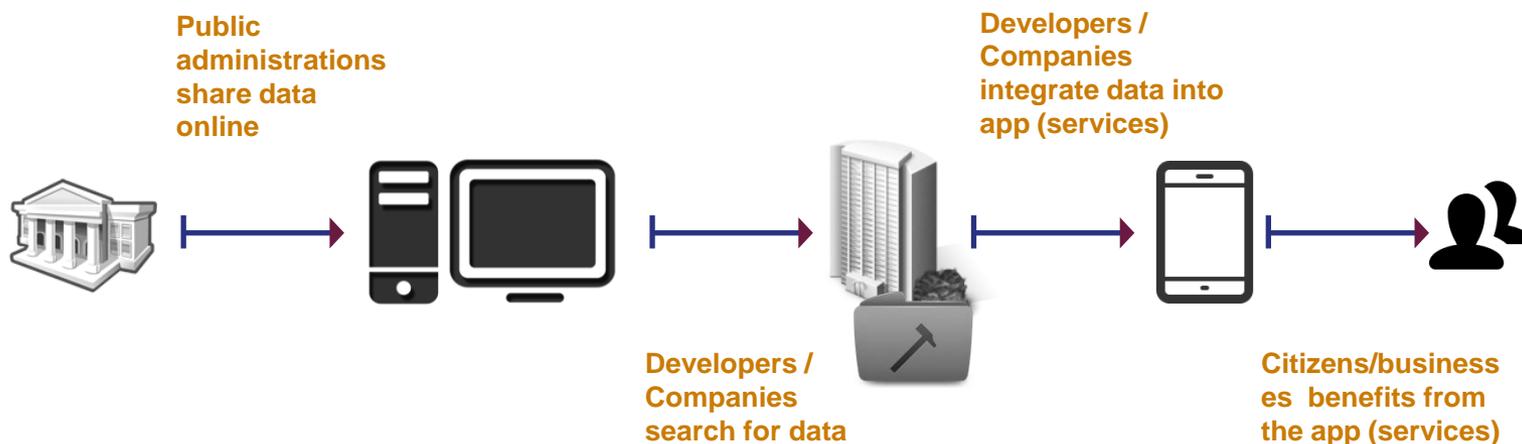
Les besoins des entreprises, des entrepreneurs, des chercheurs et des gouvernements pour les données liées.

Le côté de la demande du cycle de vie LOGD porte sur ...

Réutilisateurs de données peuvent:

- trouver des ensembles de données appropriés;
- utiliser les ensembles de données pour des analyses, la création d'applications et de services.
- Savoir ce que leur gouvernement fait (transparence)
- Réaliser des économies

Publishing data



Reusing data

Où trouver des ensembles de données?

Les ensembles de données sont disponibles sur différentes plateformes réparties à travers l'Europe.

*« Une plateforme centrale de partage des données collecte les **métadonnées** de différents portails de données ouvertes et les publie à l'aide d'un **modèle commun de métadonnées**. De cette façon, les ensembles de données sont **disponibles de façon uniformes** à partir d'un **point d'accès unique**. »*

- Portails locaux de données ouvertes:
 - opendatamanchester.org.uk
 - Data.gent.be
- Portails régionaux de données ouvertes:
 - opendata.regionpaca.fr
 - Open public data of the government of Catalonia
- Portails nationaux de données ouvertes:
 - Opendata.at
 - opendata.lu
- Portails européens de données ouvertes:
 - Open-data.europa.eu
- Plateformes de partage des données:
 - Publicdata.eu
 - ODIP

Utilisez un “SPARQL endpoint” ou un “faceted browser” pour trouver des ensembles de données

Un utilisateur peut exécuter une requête SPARQL à l'aide d'un « **SPARQL endpoint** » pour trouver des ensembles de données, ou «filtrer son chemin» à travers la collection d'ensembles de données à l'aide d'un « **faceted browser** »

The screenshot shows the 'SPARQL endpoint' page on data.gov.uk. It features a search bar for entering a SPARQL query, a dropdown for 'Data sources' (currently set to 'Reference'), and a dropdown for 'Data format' (currently set to 'XML'). Below these are fields for 'Enter your SPARQL query:' with a list of prefixes: PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>, PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>, PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>, PREFIX owl: <http://www.w3.org/2002/07/owl#>, and PREFIX skos: <http://www.w3.org/2004/02/skos/core#>. A 'submit' button is at the bottom.

<http://data.gov.uk/sparql>

The screenshot shows the 'Faceted browser' page on data.gov.uk. It displays a search bar with '110 Results' and a 'Search' button. Below the search bar, there are filters for 'Format' (set to 'RDF') and 'Tag' (set to 'accountability'). A 'Sort by' dropdown is set to 'Popularity'. The main content area shows a list of results, including 'Licence' (Open Government Licence), 'Tags' (accountability, transparency, senior-civil-service, civil-service, british-civil-service, organograms-and-salaries), and 'Resource Format' (RDF, CSV, PDF, PPT, HTML, XLS). The results are categorized by publisher: Department for Education, Department for Work and Pensions, and Foreign and Commonwealth Office.

<http://data.gov.uk/data/search>

Intégration des ensemble de données et la création d'applications et de services

Quelques outils pour l'intégration d'ensemble de données:

- Karma (<http://www.isi.edu/integration/karma/>)
- Talend (<http://www.talend.com/products/data-integration>)

Voir aussi:
<http://www.slideshare.net/OpenDataSupport/introduction-to-linked-data-23402165>

Publier les LOGD en utilisant “Open Refine”

Utiliser “Open refine” pour modéliser publier des données ouvertes

Pour débiter

1. Installez Open Refine à partir de: <https://github.com/OpenRefine>
2. Installez l'extension RDF : <http://refine.deri.ie/>

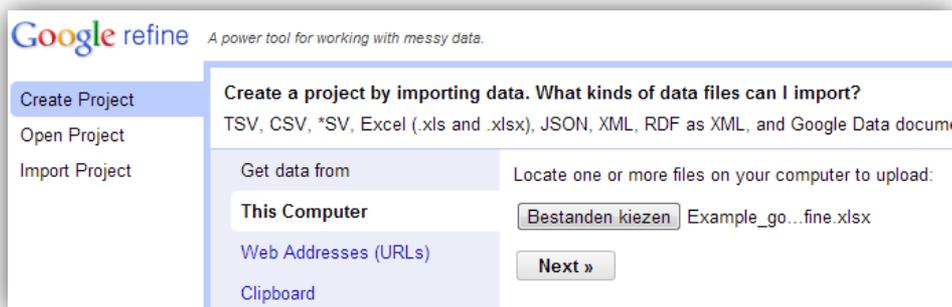
Et puis ...

- 1** Décrivez vos données dans un tableur.
- 2** Créez un projet et transférez le dans « Open Refine ».
- 3** Faites correspondre vos données avec les classes et propriétés RDF appropriées .
- 4** Exportez les données en tant que fichier RDF.

1 *Décrivez vos données dans un tableur*

Company_name	Registration date	Country	E-mail
Nikè	1991-04-28	BE	niké@sport.org
BARCO	1986-09-05	BE	Barco@email.be
Coca-Cola	1964-03-26	US	coca@cola.com

2 Créez un projet et transférez le dans « Open Refine »



Google refine A power tool for working with messy data.

Create Project
Open Project
Import Project

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents

Get data from: This Computer
Locate one or more files on your computer to upload:
Bestanden kiezen Example_go...fine.xlsx

Web Addresses (URLs)
Clipboard

Next »

Chargez le
tableur

Sélectionnez
l'onglet
approprié



Google refine A power tool for working with messy data.

Create Project « Start Over Configure Parsing Options
Open Project
Import Project

	URI	Company_name	Registration_date	Country	E-mail
1.	http://www.example.org/org/1234	Nikè	1991-04-28	BE	niké@sport.org
2.	http://www.example.org/org/2345	BARCO	1986-09-05	BE	Barco@email.be
3.	http://www.example.org/org/3456	Coca-Cola	1964-03-26	US	coca@cola.com

Créez le
projet



3 rows

Show as: rows records Show: 5 10 25 50 rows

	All	URI	Company_name	Registration_da	Country	E-mail
☆ ↻	1.	http://www.example.org/org/1234	Nikè	1991-04-28	BE	niké@sport.org
☆ ↻	2.	http://www.example.org/org/2345	BARCO	1986-09-05	BE	Barco@email.be
☆ ↻	3.	http://www.example.org/org/3456	Coca-Cola	1964-03-26	US	coca@cola.com

3

Faites correspondre vos données avec les classes et propriétés RDF appropriées (modélisez vos données)

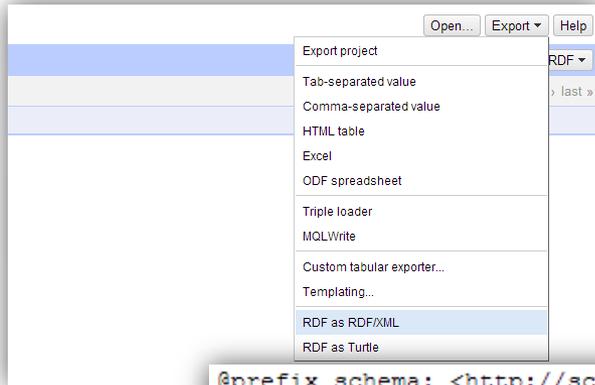
Définir un squelette pour transformer vos données de tableur à RDF

The screenshot shows the 'RDF Schema Alignment' dialog box in the Freebase editor. The dialog is titled 'RDF Schema Alignment' and contains the following elements:

- Base URI:** <http://www.w3.org/TR/vocab-regorg#> [edit](#)
- Available Prefixes:** schema rov vcard dcterms + add prefix manage prefixes
- Left Column (URI URI):**
 - [xrov:RegisteredOrganisation](#) [add rdf.type](#)
- Right Column (RDF Class/Property):**
 - [Company_name URI](#) [add rdf.type](#)
 - [Registration_date URI](#) [add rdf.type](#)
 - [Country URI](#) [add rdf.type](#)
 - [E-mail URI](#) [add rdf.type](#)
- Mapping Arrows:**
 - [xrov:legalName](#) → [Company_name URI](#)
 - [dcterms:created](#) → [Registration_date URI](#)
 - [vcard:country](#) → [Country URI](#)
 - [vcard:email](#) → [E-mail URI](#)
- Buttons:** OK, Cancel, Save

4

Exportez les données en tant que fichier RDF/XML ou Turtle



```
@prefix schema: <http://schema.org/> .
@prefix rov: <http://www.w3.org/TR/vocab-regorg/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix dcterms: <http://purl.org/dc/terms/> .

<http://www.example.org/org/1234> a <http://www.w3.org/TR/vocab-regorg#RegisteredOrganisation> ;
  <http://www.w3.org/TR/vocab-regorg#legalName> <http://www.w3.org/TR/vocab-regorg#Nikè> ;
  dcterms:created <http://www.w3.org/TR/vocab-regorg#1991-04-28> ;
  vcard:country <http://www.w3.org/TR/vocab-regorg#BE> ;
  vcard:email <http://www.w3.org/TR/vocab-regorg#niké@sport.org> .

<http://www.example.org/org/2345> a <http://www.w3.org/TR/vocab-regorg#RegisteredOrganisation> ;
  <http://www.w3.org/TR/vocab-regorg#legalName> <http://www.w3.org/TR/vocab-regorg#BARCO> ;
  dcterms:created <http://www.w3.org/TR/vocab-regorg#1986-09-05> ;
  vcard:country <http://www.w3.org/TR/vocab-regorg#BE> ;
  vcard:email <http://www.w3.org/TR/vocab-regorg#Barco@email.be> .

<http://www.example.org/org/3456> a <http://www.w3.org/TR/vocab-regorg#RegisteredOrganisation> ;
  <http://www.w3.org/TR/vocab-regorg#legalName> <http://www.w3.org/TR/vocab-regorg#Coca-Cola> ;
  dcterms:created <http://www.w3.org/TR/vocab-regorg#1964-03-26> ;
  vcard:country <http://www.w3.org/TR/vocab-regorg#US> ;
  vcard:email <http://www.w3.org/TR/vocab-regorg#coca@cola.com> .
```

La pile LOD2

Outils pour la publication et l'interrogation des LOGD via des requêtes

Publication de vos données avec la pile LOD2

« La pile LOD2 est une distribution intégrée d'outils alignés qui soutiennent le cycle de vie des données liées (ouvertes), de l'extraction, de la création, l'enrichissement, l'interconnexion, la fusion à la visualisation et à l'entretien. La pile comprend des outils des partenaires LOD2 et des tiers. »



Source: [<http://stack.lod2.eu/>]

Silk – Un outil pour lier vos données

“Le “Silk framework” est un outil pour découvrir les relations entre les éléments de données au sein de différentes sources de données liées. Les éditeurs de données peuvent utiliser le “Silk framework” pour créer des liens RDF entre leurs données et d'autres données de sources différentes sur le Web .”

Pour télécharger ou pour de plus amples informations:

<http://wifo5-03.informatik.uni-mannheim.de/bizer/silk>

Conclusions

- Le cycle de vie des LOGD et des métadonnées doit traiter à la fois l'offre et la demande.
- Choisir les données et métadonnées à ouvrir signifie qu'il faut tenir compte de différentes dimensions.
- Modéliser a pour but d'obtenir des données et des métadonnées structurées et d'atteindre un niveau de qualité approprié.
- La publication a pour but de rendre les données et les métadonnées publiques, facilement accessibles et interrogeables.
- La gestion des données et des métadonnées doit assurer que les processus et les politiques sont en place pour régir le cycle de vie des données et des métadonnées.
- L'éditeur des données doit permettre à au réutilisateur des données d'envoyer son feedback, de sorte à pouvoir mesurer la demande et la qualité du « crowdsourcing ».
- Plusieurs outils sont disponibles pour la modélisation et la publication LOGD- mais peu ont la qualité nécessaire pour être utilisé en production.

Questions de groupe



<http://www.visualpharm.com>

Avez-vous une méthodologie de gestion des données et/ou des métadonnées au niveau de votre organisation?



<http://www.visualpharm.com>

Y a-t-il une offre et une demande pour les données publiques ouvertes (liées) dans votre pays? Si oui, qui fournit quoi à qui?



<http://www.visualpharm.com>

D'après vous, quels sont les obstacles principaux pour la publication de données publiques ouvertes (liées) dans votre pays?

Faîtes aussi 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 and Stijn Goedertier

Références

Slide 5:

- GLD Life cycle. W3C. http://www.w3.org/2011/gld/wiki/GLD_Life_cycle

Slide 8:

- Linked Data Cookbook. W3C. http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

Slide 14:

- United Nations Statistics Division. COFOG (Classification of the Functions of Government). <http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=4>

Slide 21:

- Characterization Study of the Infomediary Sector - 2012 Edition. Datos.gob.es. http://datos.gob.es/datos/sites/default/files/files/Estudio_infomediario/121001%20RED%20007%20Final%20Report_2012%20Edition_vF_en.pdf

Slide 21:

- <http://data.gov.uk/data>
- <http://publicdata.eu/>
- <http://open-data.europa.eu/en/data/dataset>

Slide 21:

- <http://data.gov.uk/data/site-usage/publisher?month=>
- <http://data.gov.uk/data/site-usage/dataset>

Slide 24-25:

- Cookbook for translating Data Models to RDF Schemas. IAS Programme. <https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas>

Slide 26:

- ADMS Brochure. ISA Programme. <https://joinup.ec.europa.eu/elibrary/document/adms-brochure>

Slide 27:

- <http://lov.okfn.org/>

Slide 29:

- DCAT application profile for data portals in Europe. ISA Programme. https://joinup.ec.europa.eu/asset/dcat_application_profile/description

Slide 31:

- 10 Rules for Persistent URIs. ISA Programme. <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

Slide 32-33:

- Licensing Open Data: A Practical Guide. Naomi Korn and Professor Charles Oppenheim. http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf

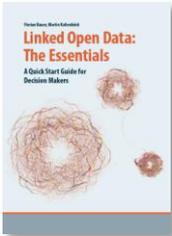
Slide 51:

- Announcement of intermediate LOD2 Stack release, March 2012. Martin Kaltenboeck. <http://lod2.eu/BlogPost/1034-announcement-of-intermediate-lod2-stack-release-march-2012.html>

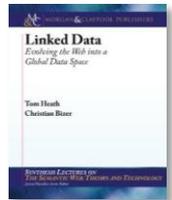
Slide 52:

- Silk - A Link Discovery Framework for the Web of Data. University of Mannheim. <http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/>

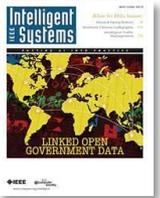
Lecture supplémentaire (2/2)



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 (1)



LOD2 Technology Stack, <http://stack.lod2.eu/>



Open Data Publishing Pipeline DERI, <http://sw.deri.ie/content/odpp>



W3C Linked Data Cookbook,
[http://www.w3.org/2011/gld/wiki/Linked Data Cookbook](http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook)



Cookbook for translating Data Models to RDF Schemas,
<https://joinup.ec.europa.eu/community/semic/document/cookbook-translating-data-models-rdf-schemas>

Projets et initiatives apparentés (2)



EUCLID FP7 Project, <http://projecteuclid.org/>



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



Generic Statistical Business Process Model,
<http://www1.unece.org/stat/platform/display/GSBPM/Generic+Statistical+Business+Process+Model+Paper>

Be part of our team...

Find us on



Open Data Support

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



Open Data Support

<http://goo.gl/y9ZZI>

Follow us



@OpenDataSupport

Join us on



joinup

<http://www.opendatasupport.eu>

Contact us

contact@opendatasupport.eu