

# *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 2.3

## Concevoir et gérer des URIs Persistants

## *Objectifs de cette formation*

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

- Qu'est-ce qu'un identificateur de ressources uniformes (URI)?
- Pourquoi la persistance d'un URI est importante?
- Comment concevoir et gérer des URIs persistants pour les ressources de données.

# Contenu

Ce module contient ...

- Une introduction aux identificateurs de ressources uniformes (URI).
- Un ensemble de principes de conception pour construire un URI persistant.
- Le service minimum pour avoir des URI persistants.

# Identificateurs de ressources uniformes (URIs)

*Comme identifiants communs pour les choses, par ex. les personnes, les bâtiments, les lieux, les ressources d'information*

...

## Qu'est-ce qu'un URI?

Un URI est

« **une séquence restreinte de caractères qui identifie une ressource abstraite ou physique** »

*[TBL et al, 2005].*

- «**restreinte**» signifie que la chaîne de caractères ne doit contenir aucun remplissage ou padding d'espace blanc;
- «**abstraite ou physique**» signifie que l'URI peut se référer à un objet (ou une chose) du monde réel, par exemple une personne, un bâtiment ou même des idées abstraites comme un service, ou un document Web.

## Par exemple ...



### Un pays, par ex. la Belgique

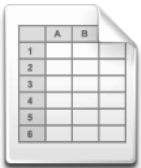
- <http://publications.europa.eu/resource/authority/country/BEL>



Publications Office

### Une organisation, par ex. le Bureau des Publications de l'UE

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



### Un ensemble de données, par ex. la liste des codes-pays

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

## Les principes clés

- **Persistant**, soit un URI affecté en permanence à une ressource particulière. Il est stable, ne change et ne disparaît pas au fil du temps.
- **Déréférençable**, c-à-d un agent utilisateur peut faire une demande à cet URI sur Internet et recevoir une réponse significative en retour.
  - Si l'agent utilisateur est un *navigateur Web*, alors ce qui est reçu doit être un document HTML lisible par l'homme.
  - Si l'agent utilisateur est un client RDF, alors un RDF doit être *renvoyé à partir de la même URI*.
- **Univoque**, c-à-d qu'il devrait n'y avoir aucune confusion entre les identifiants pour les documents Web et les identificateurs d'autres ressources.
  - Il devrait y avoir un URI différent pour le référencement de l'auteur d'une page Web et la page Web elle-même.

## Les hypothèses clés

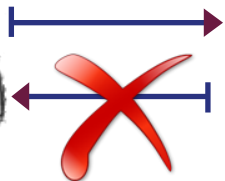
- Afin de créer et gérer des URIs, il **faudrait être le propriétaire** du domaine Internet concerné et en avoir les droits d'administrateur.
- Pour les domaines gouvernementaux, il est très probable qu'ils soient **gérés par une agence centrale**. Veillez à consulter vos collègues avant de commencer.
- Les URIs persistants et déréférençables doivent être supportés par **une infrastructure Web sous-jacente fiable**. Une telle infrastructure peut-être disponible au sein de votre organisation ou peut être fournie par une différente organisation – par ex. en tant que ressource partagée. Veillez à consulter vos collègues d'IT avant de commencer.



# Que faire si un URI n'est pas déréréférençable et / ou persistant?

Imaginez la situation suivante ...

Réolvons la description de « Irlande » à partir de la liste des codes-pays



<http://foo.org/datasetid>

Resource not found



# Concevoir des URIs persistants pour les ensembles de données

*10 choses à faire et ne pas faire*

## *Suivre un format générique d'URI*

`http://{domain}/{type}/{concept}/{reference}`

`{domain}` est une combinaison du nom de l'hôte et du secteur concerné.

- `{type}` devrait être l'un d'une petite série de valeurs possibles qui déclarent le type de ressource qui est identifiée. Les exemples typiques incluent:
  - `'id'` ou `'item'` pour les objets du monde réel;
  - `'doc'` pour les documents qui décrivent ces objets;
  - `'def'` pour des concepts;
  - `'set'` pour les ensembles de données;
  - une chaîne de caractères spécifique au contexte.
- `{concept}` pourrait être une collection, le type d'objet du monde réel ou le nom du schéma conceptuel;
- `{reference}` est un élément, terme ou concept spécifique.

## Créer un URI en réutilisant des identifiants existants

- Les identificateurs existants de ressources, par exemple des clés primaires utilisées dans une bases de données, devraient être incorporés dans l'URI.
  - Réutilisez des identificateurs qui sont eux-mêmes susceptibles d'être persistants.
  - Réutilisez des identificateurs normalisés plutôt que des codes spécifiques au système interne.
- Par exemple, si l'identificateur d'une entreprise dans un registre national des affaires est une suite comme AB123456, alors l'URI pour cette société pourrait être:

**`http://businessdata.gov/id/company/AB123456`**

# *Mettre en œuvre des URI 303 pour les ressources du monde réel*

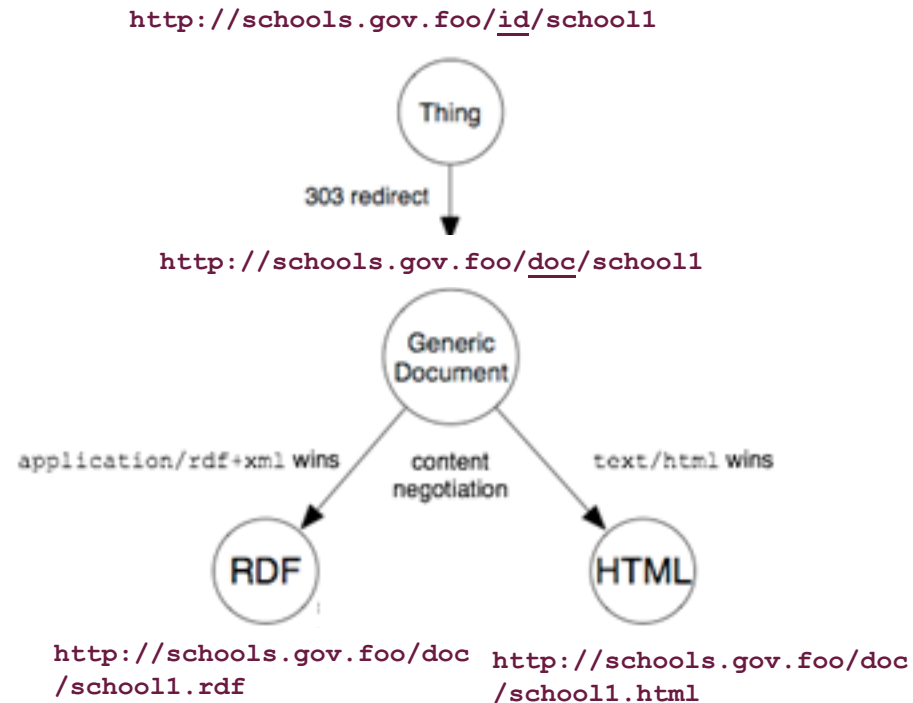
- Comme il n'existe **pas de représentation appropriée pour des ressources du monde réel** (tel qu'une ressource non-document comme un personne, une entreprise, un lieu ...), il est utile d'être dirigé vers un document Web qui contient des informations sur cette ressource.
- **Eviter l'ambiguïté** entre la ressource dans le monde réel et le document qu'il la représente.
- Par exemple, si un gouvernement décide de créer des URI 303 pour représenter les écoles primaires, le résultat pourrait être:
  - `http://schools.gov.foo/id/school1`
  - `http://schools.gov.foo/id/school2`

**Voir aussi:**

Cool URIs for the Semantic Web. <http://www.w3.org/TR/cooluris>

# Déréférencement URI 303 et la négociation de contenu

- Si déréférencés, les URIs de ces ressources devraient répondre avec un code **HTTP 303** à un document qui décrit l'objet.
- Cela signifie que le serveur Web doit être configuré pour **rediriger** de `http://schools.gov.foo/id/school1` à `http://schools.gov.foo/doc/school1`
  - Une **règle de réécriture** URI, remplaçant généralement l'URI {type} 'id' par 'doc'
- **Différentes représentations** peuvent être possibles, par ex. RDF, XML, HTML ...



## Voir aussi:

Cool URIs for the Semantic Web.

<http://www.w3.org/TR/cooluris>

## Évitez d'intégrer les numéros de version dans les URIs

- Ensembles de données, schémas conceptuels, ontologies, taxonomies et vocabulaires sont libérés dans des versions successives, suivant des cycles itératifs de changement/mise à jour
- Les URIs devraient rester **stable entre les versions**.
  - Les numéros de version et les informations d'état **ne devraient pas être inclus** dans l'URI.
- Par exemple, en imaginant deux versions consécutives, v0.01 et v0.02 d'un ensemble de données d'une école. Si les informations de version avaient été incluses dans l'URI, l'URI de l'ensemble de données doit changer à chaque fois qu'une nouvelle version sort.
  - `http://schools.gov.foo/set/0.01/schools`
  - `http://schools.gov.foo/set/0.02/schools`

## *Évitez d'utiliser l'auto-incrémentation lors de la création de nouveaux URIs*

- L'incrémentation d'un compteur lors de la création d'URIs pour de grands ensembles de données peut être simple, mais elle peut entraîner de graves problèmes.
  - Qu'advient-il si l'ensemble de données est mis à jour et les URIs doivent être attribués à nouveau? Comment pouvons-nous nous assurer que la séquence sera la même?

Est-ce ça veut dire que je n'aurais jamais à le faire?

L'utilisation de l'auto-incrémentation dans les URIs peut être envisagée, lorsque:

- le processus ne sera jamais répété;
- le processus peut être répété pour créer exactement le même URI pour les mêmes données d'entrée avec de nouveaux URIs créés seulement pour de nouvelles ressources.



## Évitez d'utiliser des requêtes à base de chaînes de caractères

- Une requête à base de chaînes de caractères (par exemple « ?param=value' ) est un texte ajouté à la fin d'une URL qui contient les données à transmettre aux applications Web, par exemple les paramètres de recherche pour trouver des termes dans une base de données.
  - Ces requêtes ne sont pas persistantes car **elles reposent sur des implémentations particulières**. Par conséquent, elles devraient être évitées dans la création d'URIs.
- Par exemple, imaginez que l'URI d'une entreprise publié dans un registre national des entreprises (NBR) ait été

`http://businessdata.gov/NBR/id/company?id="AB123456"`

au lieu de

`http://businessdata.gov/NBR/id/company/AB123456`



## Évitez d'inclure des informations sur la « propriété »

- Un modèle d'URIs persistants **ne doit pas inclure le nom de l'organisation ou du projet** qui a produit l'URI.
- Par exemple, imaginez que l'URI d'une entreprise publiée par un registre national des entreprises (NBR) ait été  
`http://businessdata.gov/NBR/id/company/AB123456` ❌
- Après plusieurs années NBR est renommé registre des entreprises nationales (RCN). Dans ce cas tous les URI doivent être mis à jour.
- Dans ce cas, un URI conçu pour être persistant serait  
`http://businessdata.gov/id/company/AB123456`

## Évitez d'utiliser des extensions de fichiers

- Les extensions de fichier dévoilent le type de fichier de documents spécifiques.
- L'utilisation d'extensions de fichier doit être évitée pour des URIs persistants
- Par exemple, l'URI d'un ensemble de données contenant la liste des écoles dans un État membre serait plutôt
  - `http://data.gov.foo/set/schools`que
  - `http://data.gov.foo/set/schools.csv`
- L'extension de fichier peut faire partie des métadonnées du document.
  - par ex., `dc:mediaType` dans le Vocabulaire pour Catalogues de Données du W3C pour décrire des ensembles de données.

# Fournir des URIs persistants pour les ensembles de données

## *Utilisez un service dédié*

- Un service dédié, de confiance, qui est **indépendant de l'émetteur des données** doit être mis en place.
- **Facile à être transféré** et géré par quelqu'un d'autre si nécessaire.
  - Dublin Core utilise purl.org
  - data.gov.uk et publications.europa.eu sont eux aussi indépendants d'un ministère spécifique.
- Il n'est pas nécessaire d'adopter un seul service pour plusieurs fournisseurs de données.
  - Des risques plus élevés car il n'y aurait qu'un point de défaillance unique, mais
  - Plus facile à gérer et plus rentable.

# Conclusions

Un URI est "une séquence restreinte de caractères qui identifie une ressource physique ou abstraite".



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/school11/123457`

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:**

10 Rules for Persistent URIs. <https://joinup.ec.europa.eu/node/53858>

## Questions de groupe



Votre pays a-t-il une politique URI nationale? Si oui, quels en sont les principes clés?



Votre pays a-t-il un service dédié pour la persistance des URIs? Si oui, quel organisation gère ce service? Si non, pourquoi?

**Faites 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 6:

- T. Berners-Lee, R. Fielding and L. Masinter (2005) "Uniform Resource Identifier (URI): Generic Syntax". <http://tools.ietf.org/html/rfc3986>

Slides 11-22:

- UK Government, CTO Council, Designing URI sets of the UK Public Sector. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/60975/designing-uri-sets-uk-public-sector.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-uri-sets-uk-public-sector.pdf)
- EC ISA Programme, Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

Slides 14-15:

- Cool URIs for the Semantic Web, <http://www.w3.org/TR/cooluris>

# Lecture supplémentaire (1/2)

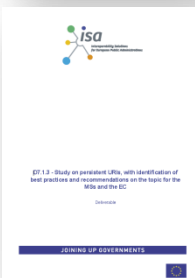


T. Berners-Lee, R. Fielding and L. Masinter (2005) "Uniform Resource Identifier (URI): Generic Syntax". <http://tools.ietf.org/html/rfc3986>



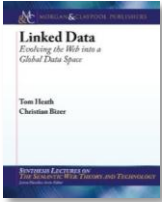
UK Government, CTO Council, Designing URI sets of the UK Public Sector.

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/60975/designing-uri-sets-uk-public-sector.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-uri-sets-uk-public-sector.pdf)



EC ISA Programme, Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. <https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris>

## Lecture supplémentaire (2/2)



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

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

# Projets et initiatives apparentés



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



W3C Cool URIs for the Semantic Web

- <http://www.w3.org/TR/cooluris>
- <http://www.w3.org/wiki/GoodURIs>



URI Design Principles: Creating Unique URIs for Government Linked Data, <http://logd.tw.rpi.edu/instance-hub-uri-design>



Publications Office of the European Commission,  
<http://publications.europa.eu>



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



*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](mailto:contact@opendatasupport.eu)