

# *Präsentation Metadaten*

Open Data Support wird von der Europäischen Kommission finanziert, gemäß 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' (Vertrag No. 30-CE-0530965/00-17).

© 2013 European Commission



# OPEN DATA SUPPORT

## Trainingsmodul 2.3

# Langlebige URIs entwerfen und verwalten

# Lernziele

Am Ende dieses Trainingsmoduls sollten Sie verstehen:

- Was ein Uniform Resource Identifier (URI) ist .
- Warum die URI Langlebigkeit wichtig ist.
- Wie Sie langlebige URIs für Daten-Ressourcen entwerfen und verwalten können.

# Inhalt

Dies Modul enthält ...

- Eine Einführung in die Uniform Resource Identifier (URI).
- Eine Reihe von Design-Prinzipen für den Aufbau von langlebigen URIs.
- Service-Anforderungen für langlebige URIs.

# Uniform Resource Identifiers (URIs)

*Als gemeinsame Bezeichnung für Dinge, z.B. Personen, Gebäude, Standorte, Informationsquellen...*

## Was ist eine URI?

Eine URI ist

**“eine kompakte Abfolge von Zeichen, die eine abstrakte oder physische Ressource bezeichnet”**

*[TBL et al, 2005].*

- **“Kompakt”** bedeutet, dass die Zeichenfolge keine Leerstellen enthalten darf;
- **“Abstrakt oder physisch”** bedeutet, dass die URI zu einem realen Objekt (oder Ding) verweisen kann, z.B. zu einer Person, einem Gebäude oder sogar zu abstrakten Ideen, wie beispielsweise einer Dienstleistung oder einem Web-Dokument.

## Zum Beispiel...



Ein Land, z.B. Belgien

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



Eine Organisation, z.B. Das Amt für Veröffentlichungen

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



Ein Datensatz, z.B. Länder Benannt Behörde Liste

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

# Grundprinzipien

- **Langlebig**, d.h. eine URI ordnet permanent auf eine bestimmte Ressource zu. Sie ist stabil und ändert sich nicht und verschwindet nicht über die Zeit.
- **Dereferenzierbar**, d.h., ein User-Agent kann eine Anfrage über das Internet an diese URI machen und eine aussagekräftige Antwort zurückempfangen.
  - Wenn der User-Agent ein *Webbrowser* ist, dann sollte das, was zurückkommt, ein von Menschen lesbares HTML-Dokument sein.
  - Wenn der User-Agent ein RDF-Auftraggeber ist, dann sollte RDF von *derselben URI* zurückgegeben werden.
- **Eindeutig**, d.h. Hier sollte es keine Verwechslung zwischen Identifier für Web-Dokumenten und Identifier für andere Ressourcen geben.
  - Es müsste eine unterschiedliche URI für die Referenzierung des Autors einer Webseite und der Webseite selbst geben.

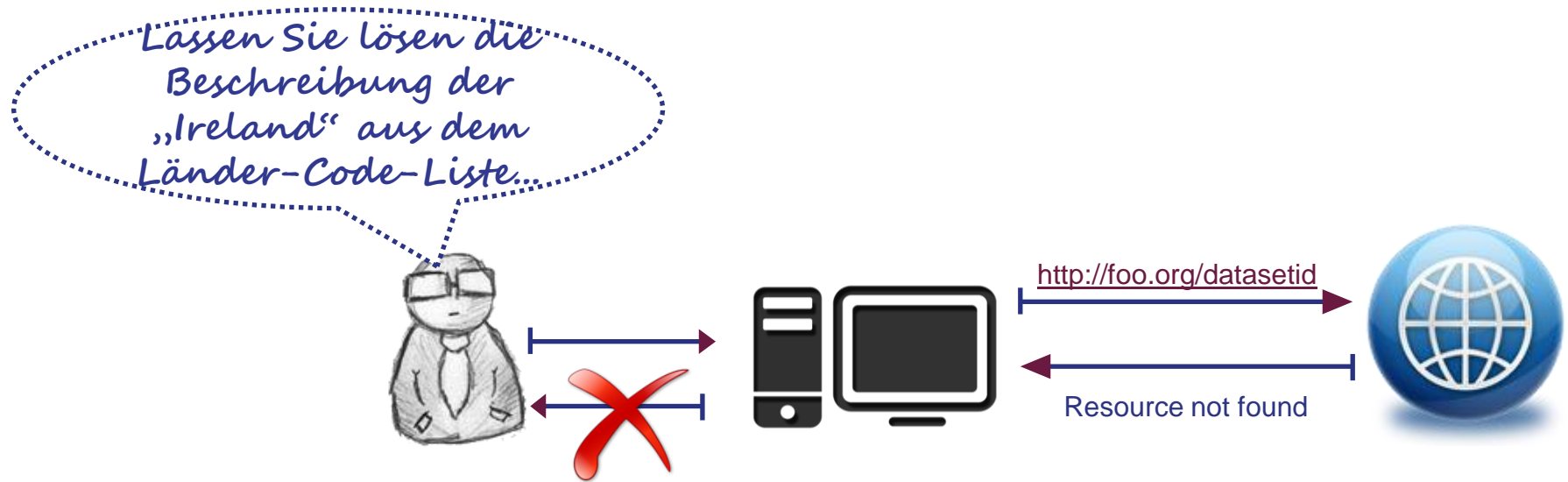
## Wesentliche Voraussetzungen

- Um URIs zu erstellen und zu verwalten, **solte** man **der Betreiber** der jeweiligen Internet-Domain **sein**, und Administratorenrechte haben.
- Für Regierungsdomains ist es sehr wahrscheinlich, dass diese **von einer zentralen Behörde verwaltet werden**. Bevor Sie starten, sollten Sie dies mit Ihren Kollegen überprüfen.
- Langlebige und dereferenzierbare URIs müssen **von einer vertrauenswürdigen zugrundeliegenden Web-Infrastruktur** unterstützt werden. Eine solche Infrastruktur kann innerhalb Ihrer Organisation verfügbar sein oder von einer anderen Organisation gestellt werden: z.B. als gemeinsame Ressource. Bevor Sie starten, sollten Sie dies mit Ihren IT-Kollegen überprüfen.



# Was passiert, wenn eine URI nicht dereferenzierbar und/oder langlebig ist?

Stellen Sie sich die folgende Situation vor...



# Das Entwerfen von langlebigen URIs für Datensätze

*10 Dos und Dont's*

## Folgen Sie einem generischen URI-Format

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

- `{domain}` ist eine Kombination aus dem Host und dem relevanten Sektor.
- `{type}` sollte einer von einer kleinen Anzahl von möglichen Werten sein, die die Art von Ressourcen erklären, die genannt werden. Typische Beispiele sind:
  - `,id`` oder `,item`` für realen Objekte;
  - `,doc`` für Dokumente, die diese Objekte beschreiben;
  - `,def`` für Konzepte;
  - `,set`` für Datensätze;
  - eine Reihe spezifischer Merkmale im Kontext.
- `{concept}` könnte eine Sammlung, die Art des genannten realen Objekts oder der Name des Konzept-Schemas sein;
- `{reference}` ist ein bestimmtes Element, Begriff oder Konzept.

## *Prägen Sie URIs, indem Sie bestehende Bezeichnungen wiederverwenden*

- Bestehende Bezeichnungen von Ressourcen, z.B. Datenbankschlüsseln, sollten in die URI eingebaut werden.
  - Verwenden Sie diejenigen Bezeichnungen wieder, die wahrscheinlich selbst langlebig sind.
  - Verwenden Sie Standard- Bezeichnungen anstatt interne System-spezifische Codes.
- Zum Beispiel, wenn die Bezeichnung eines Unternehmens in einem nationalen Geschäftsregister ein String wie AB123456 ist, dann könnte die URI für dieses Unternehmen sein:

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

## Implementieren Sie 303 URIs für reale Ressourcen

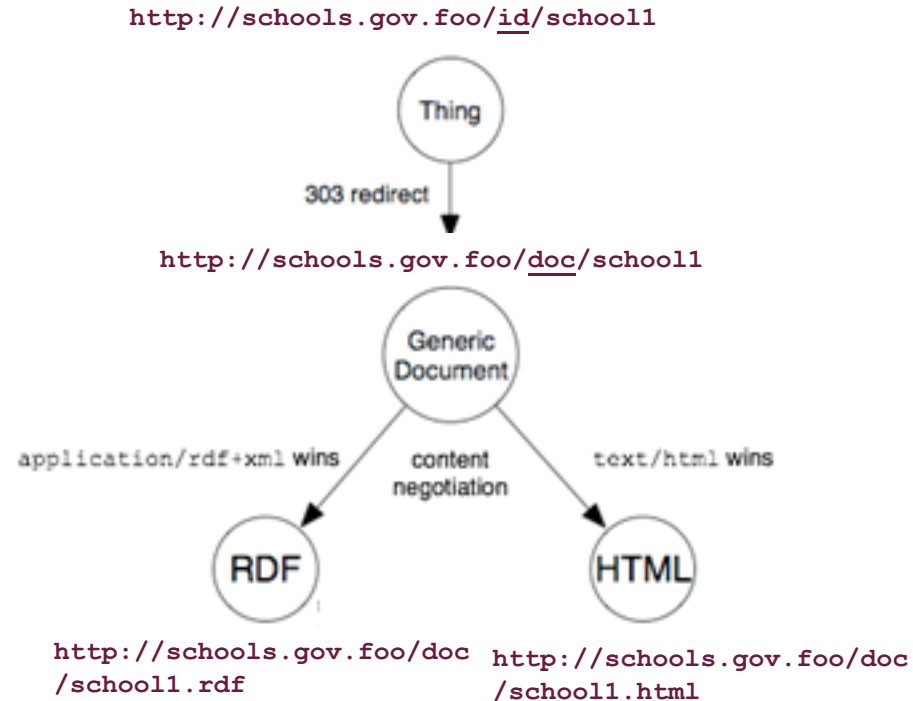
- Wenn **keine geeignete Darstellung für eine reale Ressource existiert** (d.h. eine Ressource , die nicht auf einem Dokument basiert, wie eine Person, ein Unternehmen, ein Standort...), ist es nützlich, zu einem Web-Dokument, das Information über diese Ressource hält, geleitet zu werden.
- Es **vermeidet Mehrdeutigkeit** zwischen der realen Ressource und dem Dokument, das sie darstellt.
- Zum Beispiel, wenn eine Regierung beschließt, 303 URIs zu erstellen, um Grundschulen zu vertreten, kann das Ergebnis sein:
  - `http://schools.gov.foo/id/school1`
  - `http://schools.gov.foo/id/school2`

### Siehe auch:

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

# Die Dereferenzierung von 303 URIs und Content-Übertragung

- Bei der Dereferenzierung, sollten die URIs dieser Ressourcen mit **HTTP 303** auf ein Dokument antworten, das das Objekt beschreibt.
- Der Web-Server muss konfiguriert werden, um von `http://schools.gov.foo/id/school1` nach `http://schools.gov.foo/doc/school1` zu führen.
  - Eine **URI-Regel** zum Wiederschreiben ersetzt typischerweise die URI {type} von 'id' mit 'doc'.
- **Verschiedene Darstellungen** sind möglich , z.B. RDF, XML, HTML...



## Siehe auch:

Cool URIs for the Semantic Web.

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

## Vermeiden Sie es, Versionsnummern in URIs einzubeziehen.

- Datensätze, Konzeptschemata, Ontologien, Taxonomien und Vokabulare werden in aufeinanderfolgenden Versionen nach sich wiederholenden Änderungs-/Update-Zyklen veröffentlicht.
- Die URIs sollte **zwischen Versionen beständig bleiben**.
  - Versionsnummern und Status-Informationen **sollte nicht** in die URI **aufgenommen werden**.
- Zum Beispiel beim Aufzeigen von zwei aufeinanderfolgenden Versionen v0.01 und v0.02 des Schul-Datensatzes: Wenn hier Versionsinformationen in der URI enthalten waren, dann muss die URI des Datensatzes jedes Mal, wenn eine neue Version herauskommt, geändert werden.
  - `http://schools.gov.foo/set/0.01/schools`
  - `http://schools.gov.foo/set/0.02/schools`

## *Vermeiden Sie es „auto-increment“ zu benutzen, wenn sie neue URIs prägen.*

- Bei der Erstellung von URIs für einen großen Datensatz mag es einfach sein, eine Gegenmaßnahme zu inkrementieren, doch dies kann zu ernststen Problemen führen.
  - Was passiert, wenn der Datensatz aktualisiert wird und URIs wieder zugeordnet werden müssen. Wie können wir sicherstellen, dass die Sequenz die gleiche sein wird?

*Bedeutet dies, dass ich es nie tun sollte?*

Die Verwendung von auto-increment in URIs kann in Betracht kommen, wenn:

- der Prozess niemals wiederholt wird;
- der Prozess wiederholt werden kann, um genau dieselben URIs für dieselben Eingabedaten mit neuen URIs zu entwickeln, die nur für neue Elemente geprägt werden.



## Vermeiden Sie die Nutzung von Query-Strings

- Ein Query-String (z.B. '?param=Wert') ist ein Text, der an das Ende eines URL angehängt wird und der Daten enthält, die an Web-Anwendungen weitergegeben werden, z.B. Suchparameter, um Begriffe in einer Datenbank nachzuschlagen.
  - Query-Strings sind nicht langlebig, da sie **auf bestimmten Implementierungen beruhen**. Daher sollten sie von URIs vermieden werden.
- Stellen Sie sich zum Beispiel vor, die URI eines von einem nationalen Unternehmensregister (NBR) veröffentlichten Unternehmens war

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

statt

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



## *Vermeiden Sie es, Informationen über die Eigentümerschaft einzubeziehen*

- Eine langlebige URI-Vorlage sollte **nicht den Namen der Organisation oder des Projekts**, das die URI geprägt hat, einbeziehen.
- Stellen Sie sich zum Beispiel vor, die URI eines von einem nationalen Unternehmensregister (NUR) veröffentlichten Unternehmens war.

<http://businessdata.gov/NBR/id/company/AB123456>



- Nach ein paar Jahren wird NBR umbenannt in Nationales Gesellschaftsregister (NGR). Infolgedessen müssen alle URIs aktualisiert werden.
- In diesem Fall wäre eine auf Langlebigkeit entworfene URI:

<http://businessdata.gov/id/company/AB123456>

## *Vermeiden Sie es, Dateierweiterungen zu verwenden*

- Dateierweiterungen zeigen den Dateityp eines bestimmten Dokuments.
- Die Verwendung von Dateierweiterungen sollten in langlebigen URIs vermieden werden.
- Zum Beispiel sollte die URI von einem Datensatz, der die Liste von Schulen in einem Mitgliedsstaat enthält, eher
  - `http://data.gov.foo/set/schools`sein, als
  - `http://data.gov.foo/set/schools.csv`
- Die Dateierweiterung kann Teil der Metadaten des Dokuments sein.
  - Z.B. `dcat:mediaType` im Datenkatalog Vokabular von W3C für die Beschreibung der Datensätzen.

# Langlebige URIs für Datensätze anbieten

## *Verwenden Sie einen speziellen Dienst*

- Es muss ein spezieller zuverlässiger Dienst eingerichtet werden, der **unabhängig vom Daten-Urheber** ist,.
- Dieser Dienst sollte, wenn notwendig, **leicht** auf jemand anderes **übertragen** und von ihm ausgeführt werden können.
  - Dublin Core verwendet purl.org
  - data.gov.uk und publications.europa.eu sind auch unabhängig von einer bestimmten Behörde
- Es ist nicht notwendig, einen einzigen Dienst für mehrere Datenversorger einzusetzen.
  - Ein höheres Risiko, weil es eine einzelne Bruchstelle wäre, aber:
  - Es ist leichter zu verwalten und kosteneffizienter.

# Schlussfolgerungen

Eine URI ist “eine kompakte Zeichenfolge, die eine abstrakte oder physische Ressource benennt”.



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`

**Siehe auch:**

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

## Gruppe Fragen



<http://www.visualpharm.com>

Hat Ihr Land eine nationale URI-Politik? Wenn ja, was sind die Grundprinzipien?



<http://www.visualpharm.com>

Verfügt Ihr Land über einen speziellen Dienst für URI-Langlebigkeit? Wenn ja, welche Organisation verwaltet diesen Dienst? Wenn nein, warum nicht?

**Nehmen Sie auch die Online-Test!**

# Vielen Dank!

## ...und jetzt IHRE Fragen?



# *Diese Präsentation wurde von Open Data Support erstellt*

## **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.

Autoren:

Michiel De Keyzer, Nikolaos Loutas and Stijn Goedertier

# Referenzen

Folie 6:

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

Folien 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>

Folien 14-15:

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

# Weiter lesen (1/2)

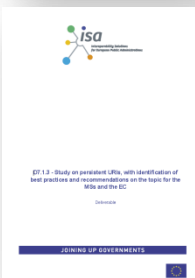


T. Berners-Lee, R. Fielding and L. Masinter (2005) "Uniform Resource Identifier (URI): Generic Syntax". <http://tools.ietf.org/html/rfc3986>



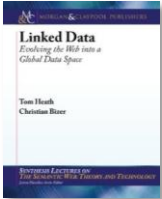
UK Government, CTO Council, Designing URI sets of the UK Public Sector.

[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>

## Weiter lesen (2/2)



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

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

## Verwandte Projekte und Initiativen



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>



# Werden Sie Teil unseres Teams...

*Finden Sie  
uns auf*



Open Data Support

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



Open Data Support

<http://goo.gl/y9ZZI>

*Folgen Sie  
uns auf*



@OpenDataSupport

*Begleiten Sie  
uns auf*



joinup

<http://www.opendatasupport.eu>

*Kontaktieren  
Sie uns unter*

[contact@opendatasupport.eu](mailto:contact@opendatasupport.eu)