



# Public Linked Open Data - the Publications Office Contribution to the Semantic Web

Audience: W3C Workshop: Making the Multilingual Web Work

Delivered by: Peter Schmitz - Publications Office of the  
European Union (Publications Office)

Date: 13/03/2013



## Publisher of the EU Institutions

mandatory publications: legislation

non-mandatory publications: books, eBooks, posters, leaflets...

### *..but also:*

allocation of international standard numbers, cataloguing, indexation, distribution, promotion, sale, storage and archiving

### *..and finally:*

in the middle of a major change initiative: from paper to electronic, from a publisher to an information provider and in this context also towards a provider of multilingual open data





Dealing with multilingualism is core business!

Daily simultaneous and partially synoptic publication  
in up to 23 languages:

Bulgarian Czech Danish Dutch English Estonian Finnish  
French German Greek Hungarian Irish Italian Latvian  
Lithuanian Maltese Polish Portuguese Romanian Slovak  
Slovene Spanish Swedish





## Multilingual public websites



Access to European Union law



European public procurement



CORDIS – Community Research and Development Information Service



All EU publications



CELLAR

Common repository for content and metadata

EU Data Portal

Open Data Hub of the European Union

EuroVoc

Multilingual thesaurus

MDR

Metadata Registry



The metadata and content from daily production + backlog of the Publications Office represents over 43 million files, over 13 TB and growing...

# Cellar

**CCR** Common CONTENT repository

Fedora Commons 3.5

EMC2

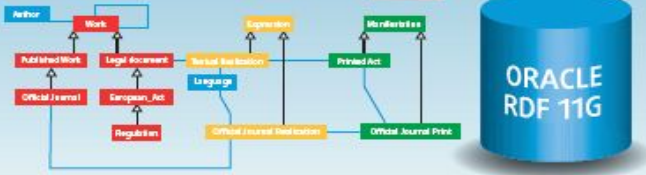


jBoss Spring Hibernate Pellet Jena joseki SunOS 5.10

**CMR** Common METADATA repository

FRBR

RDF



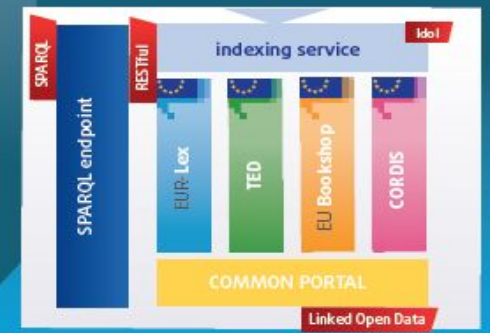
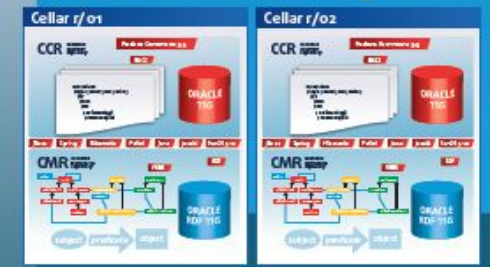
**ZIP** **METS**

```

<met>
<metsHdr/>
<dmdSec/>
<amdSec/>
<fileSec/>
<structMap/>
<structLink/>
<behaviorSec/>
</mets>
  
```

Long-term Archiving

ORACLE DataGuard replication



**OWL**

Common Data Model (ontology)

**OWL SKOS-core**

Named Authority Lists

**OWL SKOS-XL SKOS-core**

EuroVoc (thesaurus)

internet

# CELLAR – Figures

Start of initial load: May 2012

Status at the 7th of March 2013:

Count (\*) of triples: **719.916.730**

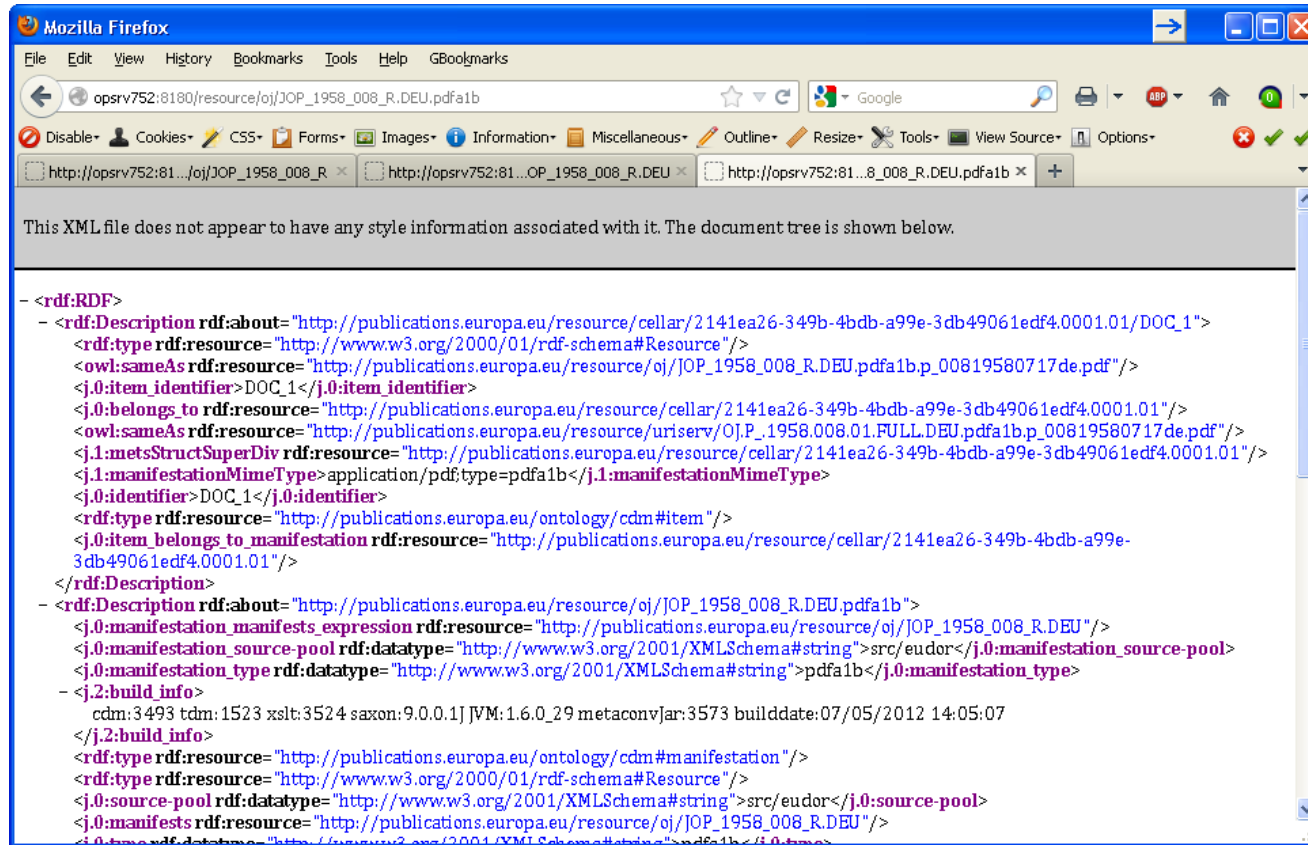
Count of manifestations per format:

14.007.390	pdf
8.952.408	xml
8.369.374	html
4.429.379	pdfa1b
3.275.685	xhtml
2.391.283	fmx4
1.646.626	jpg
971.465	pdfa1a
782.065	tif
28.056	pdfx
13.644	doc



# CELLAR – Open Data

...through RESTful API



The screenshot shows a Mozilla Firefox browser window with the address bar displaying `oprsrv752:8180/resource/oj/JOP_1958_008_R.DEU.pdf#1b`. The browser interface includes a menu bar (File, Edit, View, History, Bookmarks, Tools, Help, GBookmarks), a search bar with the Google logo, and a toolbar with various icons. The main content area displays the XML document tree for the loaded PDF file. The XML structure is as follows:

```
- <rdf:RDF>
- <rdf:Description rdf:about="http://publications.europa.eu/resource/cellar/2141ea26-349b-4bdb-a99e-3db49061edf4.0001.01/DOC_1">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
  <owl:sameAs rdf:resource="http://publications.europa.eu/resource/oj/JOP_1958_008_R.DEU.pdf#1b_p_00819580717de.pdf"/>
  <j.0:item_identifier>DOC_1</j.0:item_identifier>
  <j.0:belongs_to rdf:resource="http://publications.europa.eu/resource/cellar/2141ea26-349b-4bdb-a99e-3db49061edf4.0001.01"/>
  <owl:sameAs rdf:resource="http://publications.europa.eu/resource/uriserv/OJ.P_1958.008.01.FULL.DEU.pdf#1b_p_00819580717de.pdf"/>
  <j.1:metsStructSuperDiv rdf:resource="http://publications.europa.eu/resource/cellar/2141ea26-349b-4bdb-a99e-3db49061edf4.0001.01"/>
  <j.1.manifestationMimeType>application/pdf;type=pdf#1b</j.1.manifestationMimeType>
  <j.0:identifier>DOC_1</j.0:identifier>
  <rdf:type rdf:resource="http://publications.europa.eu/ontology/cdm#item"/>
  <j.0:item_belongs_to_manifestation rdf:resource="http://publications.europa.eu/resource/cellar/2141ea26-349b-4bdb-a99e-3db49061edf4.0001.01"/>
</rdf:Description>
- <rdf:Description rdf:about="http://publications.europa.eu/resource/oj/JOP_1958_008_R.DEU.pdf#1b">
  <j.0:manifestation_manifests_expression rdf:resource="http://publications.europa.eu/resource/oj/JOP_1958_008_R.DEU"/>
  <j.0:manifestation_source_pool rdf:datatype="http://www.w3.org/2001/XMLSchema#string">src/eudor</j.0:manifestation_source_pool>
  <j.0:manifestation_type rdf:datatype="http://www.w3.org/2001/XMLSchema#string">pdf#1b</j.0:manifestation_type>
- <j.2:build_info>
  cdm:3493 tdm:1523 xslt:3524 saxon:9.0.0.1J JVM:1.6.0_29 metaconvjar:3573 builddate:07/05/2012 14:05:07
</j.2:build_info>
<rdf:type rdf:resource="http://publications.europa.eu/ontology/cdm#manifestation"/>
<rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
<j.0:source_pool rdf:datatype="http://www.w3.org/2001/XMLSchema#string">src/eudor</j.0:source_pool>
<j.0:manifests rdf:resource="http://publications.europa.eu/resource/oj/JOP_1958_008_R.DEU"/>
<j.0:type rdf:datatype="http://www.w3.org/2001/XMLSchema#string">pdf#1b</j.0:type>
```





# CELLAR – Open Data

...through Sparql endpoint

The screenshot shows the 'Cellar Admin Interface' in Mozilla Firefox. The browser address bar shows 'opsrv880:8180/admin/cmr/sparql/endpoint'. The page title is 'Cellar Admin Interface' and the URL is 'opsrv880:8180/admin/cmr/sparql/endpoint'. The interface includes a navigation menu with buttons for Dashboard, Audit/Log, Sparql, Indexing, Log, Configuration, Tracker, Archive, Search, Content Visibility, Work Embargo, NALs, Ontologies, Languages, and Validator. The 'Sparql' button is highlighted. The main content area is titled 'Sparql Endpoint' and contains a sidebar with 'Sparql Endpoint' and 'Configuration' options. The main area displays 'Used prefixes:' with a list of URIs for 'oracle', 'rdfs', 'rdf', 'xsd', 'module', 'joseki', 'ja', 'cdm', and 'owl'. Below this is an 'Example query:' dropdown menu set to 'None'. A large text area contains the query 'SELECT ?x ?y ?z WHERE ( ?x ?y ?z )'. To the right of the text area is a 'Submit Query' button. Further right is a box for 'JSON output:' and 'XSLT style sheet:' with a dropdown menu. The footer of the page reads 'Cellar Admin Interface - build: 0.1'.



## EU Data Portal

<http://open-data.europa.eu/open-data/>

### Commission reuse policy:

Framework:

#### Decision of the 11th of December 2011

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:330:0039:0042:EN:PDF>

Tool:

#### EU Data Portal

Article 5 of the decision:

The Commission shall set up a data portal as a single point of access to its structured data so as to facilitate linking and reuse for commercial and non-commercial purposes.



- RDF based interface for upload of metadata

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:ecodp="http://ec.europa.eu/open-data/ontologie/ec-odp#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dct="http://purl.org/dc/terms/" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:dcat="http://
  www.w3.org/ns/dcat#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:skos-xl="http://www.w3.org/2008/05/skos-xl#"
  xmlns:dcam="http://purl.org/dc/dcam/" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns="http://ec.europa.eu/open-data/ontologies/ec-odp-example#"> <!--
  Dataset type and URI, don't fill in the CKAN catalog identifier here -->
<dcat:Dataset rdf:about="http://publications.europa.eu/resource/authority/language/"> <!-- Title, 1 ... 1 can be filled in in multiple translations through -->
<dct:title xml:lang="en">Languages Name Authority List</dct:title> <!-- Alternative Title(s), 0..* -->
<dct:alternative xml:lang="en">Languages NAL</dct:alternative> <!-- Description, 1..* -->
<dct:description xml:lang="en">The Languages name authority list or authority table is a controlled vocabulary listing languages with their authority code and their label in
  the 23 official languages of the EU when available. The table is based on the ISO 639 standards and provides mappings between them. This table is maintained by the
  Publications Office of the EU in its Metadata Registry (MDR)</dct:description> <!-- Keyword, 0..* -->
<ecodp:keyword xmlns:ecodp="http://ec.europa.eu/open-data/ontologies/ec-odp#" xml:lang="en">authority data</ecodp:keyword>
<ecodp:keyword xmlns:ecodp="http://ec.europa.eu/open-data/ontologies/ec-odp#" xml:lang="en">controlled vocabulary</ecodp:keyword> <!-- Identifier, 0..* -->
<dct:identifier xml:lang="en">language</dct:identifier> <!-- Interoperability level, 0..1 -->
<ecodp:interoperabilityLevel xmlns:ecodp="http://ec.europa.eu/open-data/ontologies/ec-odp#"
  <skos:Concept rdf:about="http://ec.europa.eu/open-data/kos/interoperability-level/Semantic" />
</ecodp:interoperabilityLevel>
```

- Metadata stored in triple store and accessible through Sparql endpoint
- Already several data providers provide RDF





Eurovoc available in SKOS/RDF or XML format

Integrated in the semantic model of the CELLAR

Outlook:

EuroVoc will be stored and accessed  
through the CELLAR



MDR – multilingual controlled vocabularies  
<http://publications.europa.eu/mdr/index.html>

Interinstitutional Metadata Registry: controlled vocabulaires, exchange protocol

Approved data is public available

Outlook:

Publication and access through CELLAR, individual URIs for controlled vocabularies will become resolvable, CELLAR ontology will be integrated.



# MDR – multilingual controlled vocabularies

<http://publications.europa.eu/mdr/index.html>

## Extract of MDR skos:

```

dc:identifier>ENG</dc:identifier>
<at:start.use>1950-05-09</at:start.use>
<skos:prefLabel xml:lang="es">inglés</skos:prefLabel>
<skos:prefLabel xml:lang="da">engelsk</skos:prefLabel>
<skos:prefLabel xml:lang="de">Englisch</skos:prefLabel>
<skos:prefLabel xml:lang="el">αγγλική</skos:prefLabel>
<skos:prefLabel xml:lang="en">English</skos:prefLabel>
<skos:prefLabel xml:lang="fr">anglais</skos:prefLabel>
<skos:prefLabel xml:lang="it">inglese</skos:prefLabel>
<skos:prefLabel xml:lang="nl">Engels</skos:prefLabel>
<skos:prefLabel xml:lang="pt">inglês</skos:prefLabel>
<skos:prefLabel xml:lang="fi">englanti</skos:prefLabel>
<skos:prefLabel xml:lang="sv">engelska</skos:prefLabel>
<skos:prefLabel xml:lang="ga">Béarla</skos:prefLabel>
<skos:prefLabel xml:lang="cs">angličtina</skos:prefLabel>
<skos:prefLabel xml:lang="et">inglise keel</skos:prefLabel>
<skos:prefLabel xml:lang="lv">angļu valoda</skos:prefLabel>
<skos:prefLabel xml:lang="lt">anglų kalba</skos:prefLabel>
<skos:prefLabel xml:lang="hu">angol</skos:prefLabel>
<skos:prefLabel xml:lang="mt">Ingliz</skos:prefLabel>
<skos:prefLabel xml:lang="pl">angielski</skos:prefLabel>
<skos:prefLabel xml:lang="sk">angličtina</skos:prefLabel>
<skos:prefLabel xml:lang="sl">angleščina</skos:prefLabel>
<skos:prefLabel xml:lang="bg">английски</skos:prefLabel>
<skos:prefLabel xml:lang="ro">engleză</skos:prefLabel>
<skos:prefLabel xml:lang="tr">İngilizce</skos:prefLabel>
<skos:prefLabel xml:lang="is">enska</skos:prefLabel>
<skos:prefLabel xml:lang="no">engelsk</skos:prefLabel>
<skos:prefLabel xml:lang="ru">Английский язык</skos:prefLabel>
<skos:prefLabel xml:lang="hr">engleski</skos:prefLabel>
<skos:prefLabel xml:lang="mk">англиски</skos:prefLabel>
<skos:prefLabel xml:lang="sh">engleski jezik</skos:prefLabel>
<skos:prefLabel xml:lang="sr">Енглески језик</skos:prefLabel>
<skos:prefLabel xml:lang="hsb">Jendźelšćina</skos:prefLabel>
<skos:prefLabel xml:lang="uk">Англійська мова</skos:prefLabel>
<at:op-mapped-code>

```



# European Legislation Identifier (ELI)

EU Council has published the definition European Legislation Identifier (ELI) and invites for introduction

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:325:0003:0011:EN:PDF>

For the data exchange to become more efficient, ELI metadata elements may be serialised in compliance with the W3C Recommendation "RDFa in XHTML: Syntax and Processing".

ELI uses "HTTP URIs" to specifically identify all online legal information officially published across Europe. These URIs are formally described by machine-readable URI templates (IETF RFC 6570), using components that carry semantics both from a legal and an end-user point of view. Each Member State will build its own, self-describing URIs using the described components as well as taking into account their specific language requirements.

All the components are optional and can be selected based on national requirements and do not have a pre-defined order. To enable the exchange of information the chosen URI template must be documented using the URI template mechanism, see example below:

```
/eli/{jurisdiction}/{agent}/{sub-agent}/{year}/{month}/{day}/{type}/{natural identifier}/{level1...}/{point in time}/{version}/{language}
```



Thank you for your attention!

