



***Multilinguism and Linked Open  
Data in the EU Open Data Portal  
and other projects of the  
Publications Office***

**Audience:** W3C Workshop: The Multilingual Web – Linked Open Data  
and MultilingualWeb-LT Requirements

**Presented by:** Peter Schmitz

**Date:** 11/06/2012



- Commission's decision on the reuse of Commission documents (2011/833/EU, OJ L 330, page 39):
  - Increase efficiency of Commission's reuse regime with the objective to achieve broader reuse of such documents for commercial and non-commercial purposes
  - Extension of reuse to data and commitment for the set-up of a data portal
  - Standard conditions for reuse:
    - Acknowledgement of the source
    - No distortion of the original meaning or message
    - Non-liability of the Commission for any consequence stemming from the reuse
  - Free of charge



# Positioning of the Publications Office

- Publications Office (PO) = publisher of the EU Institutions
- Daily publication of the Official Journal of the EU in 23 languages

## Main public online services



# Positioning of the Publications Office



- Commission reuse policy includes documents published by the Publications Office (PO) on behalf of the Commission
- PO is in charge of technical implementation of the Commission Open Data Portal (ODP) = single point of access to the structured data available for reuse
- PO is also setting up a “common portal” as a single point of access to all content and metadata managed by PO, i.e. related to official publications of the EU
- PO involved in metadata standardisation on the EU Institution level, including member states (government services)



# Availability and re-use of language resources in the NLP domain

## Possible contributions of PO:

- Multilingual thesaurus (EuroVoc), including a Thesaurus Alignment Environment  
Objective: extension of general domains/concepts through alignment with specialized thesauri
- Multilingual controlled vocabularies and taxonomies (“common authority tables”)
- Linked multilingual XHTML content (Official Journal, Case law...)



# Current and future content delivery infrastructure for Linked-Multilingual Web Data

## PO contributions (implementation ongoing):

- Storage and dissemination of metadata (for content and datasets) as well as of controlled vocabularies in RDF
- Set-up of a public Sparql endpoint
- Provision of persistent URIs  
(URI prefix: <http://publications.europa.eu/>)
- Support and encouragement of data providers to represent data in RDF
- Provision of visualisation tools based on RDF (open data portal)



# Organisational and technical issues relating to collaboration and access control



## For the Open Data Portal:

- On the portal level there will be the possibility to contribute: « ideas », « forum »
- Data providers will have access to the catalogue data related to their own data sets

## For the general dissemination architecture of PO:

- Scalability and access control through replication, in particular of the triple store (read-only nodes)



# Crowd-sourced content annotation and adaptation of LOD language resources

Annotation of official content is problematic!

Are there use cases, requirements?

Are there ideas/strategies to organise participative collaboration between online communities and public service organisations?

Support of crowd-sourced activities by public services?

How to define and to support quality notions?





## Provenance tracking, history and storage of LOD resources

Linked open data and authenticity:

- Is there something like a responsible organisation for particular linked open data?
- How to implement this concept?
- How to approve it?

This also includes preservation of history and storage of these data.

Examples in PO context:

- Release management for LOD resources (EuroVoc, controlled vocabularies)
- Preservation of historical data and mapping to previous codes (for controlled vocabularies)



## Stable and authoritative source for URIs of quality data « Guaranteed » and authorised relationships

- Official translations or even legal equivalence of resources in different languages (for example: EU legislation)
- Federation of open data based on linked open data (catalogue data) (pan-European data portal)
- Standardisation and linking of relevant public sector information.

Example: **European Legislation Identifier (ELI)**, which is supposed to identify in a unique way, legislation adopted at regional, national and European level

### Directive 2008/98/EC

ELI URI: <http://eurlex.europa.eu/eli/dir/2008/98>

**dir**: directive **2008**: year **98**: number





# Linked Open Data and it's role in MultilingualWeb-LT metadata

Integration of linked open data through MultilingualWeb-LT metadata - standardized "hooks".

Reference from Web content items

- to entries in multilingual thesauri or authority tables
- to information about the trustworthiness and authenticity
- to legal identifiers, provenance information

Enrichment of Web content with high quality information resources as basis for improving machine translation and other multilingual technologies as well as localization workflows.

(Contribution of F. Sasaki)



Thank you for your attention!

