Desired portal functionality for effective resource discovery on Official Statistics (OS) data across Europe

Marion Wittenberg
(Data Archiving and Networked Services - DANS)
Contributors

• Jesús Bouso Freijo (CIS)
• Luis Antonio Castrillo Aguilera (CIS)
• Camelia Chirniciuc (CNPS-INS)
• Maria Radulescu (CNPS-INS)
• Mike Priddy (DANS)
• Marion Wittenberg (DANS)
• Christopher Guerke (Destatis)
• Maurice Brandt (Destatis)
• Arofan Gregory (MT)
• Atle Alvheim (NSD)
• Ørnulf Risnes (NSD)
• John Shepherdson (UKDA)
The European landscape
The European landscape

- Many NSIs, & Archives (sources)
- CESSDA portal
- 23 official languages
- Eurostat portal with aggregate data
- Different levels of documentation
- Metadata on the same datasets in various locations/instances
Objectives of DwB
for the resource discovery portal

• To investigate the possibilities and problems associated with harvesting NSI metadata on OS data and making them available through an enhanced CESSDA portal.

• To create a metadata model incorporating SDMX and DDI as well as any system-specific enrichment required to deliver extended portal functionality.

• To develop functional requirements for effective resource discovery on data harvested from the NSIs.
Resource discovery - the difficulties

- Resource discovery can often be serendipitous
- Silos of data & metadata
- Multiple sources of information can relate to the same dataset
- Diversity of Metadata (describing OS data)
- Technical diversity of sources
Diversity of Metadata (describing OS data)

• NSIs
• Data archives
• DwB
• Other sources
Diversity in Metadata: NSI’s

- SDMX (full or large extent compliant)
- SDMX (partially or small extent compliant)
- DQAF/SDDS
- Metadata registry standard
- Data Document Initiative DDI
- Dublin Core DC
- Neuchatel classification model
- Neuchatel variable model
- Other

Bar chart showing the diversity in metadata standards among NSI's.
NSI: Export/interchange & Dissemination Formats

- CSV
- Excel
- HTML
- Word
- XML
- PDF
- txt

- Export/Interchange
- Dissemination

DwB
### Diversity of Metadata: Archives

**Labour Force Survey**

<table>
<thead>
<tr>
<th>Study</th>
<th>Section</th>
<th>Variable</th>
<th>Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ9801:</td>
<td></td>
<td></td>
<td>WISDOM</td>
</tr>
<tr>
<td>MZ7402:</td>
<td></td>
<td></td>
<td>WISDOM</td>
</tr>
<tr>
<td>MZ9601:</td>
<td></td>
<td></td>
<td>WISDOM</td>
</tr>
<tr>
<td>MZ7901:</td>
<td></td>
<td></td>
<td>WISDOM</td>
</tr>
<tr>
<td>MZ9901:</td>
<td></td>
<td></td>
<td>WISDOM</td>
</tr>
<tr>
<td>Quarterly Labour Force Survey, June - August, 2008: Ethnicity Teaching Dataset</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Statens Arkiver**

Dansk Data Arkiv

**Søgningen efter studier med arbejd fandt 1248 datamateriale(r).**

Se her for information om søgeresultatet

Du ser post 1 til 12.

**Bestil:**

**Titel:**
Multiple metadata instances
For 3 years
For multiple years, we don’t want this situation
Clear provenance
What does the researcher expect of a portal

- To search on variables
- Information on breaks/gaps in longitudinal data
- Information on comparability between countries
- Overview of the differences between the data files
  ➢ Wikis - user generated content
Portal user requirements

- Interviews with researchers
- Information from NDS’s & archives
- User stories
- Testing user stories with stakeholders
Official Statistics Object Model
The metadata model

- SDMX & DDI (codebook and lifecycle)
- Aggregate data & micro data
- thesauri & language support
Outcomes

- One (of many) portals
  - Many specialized resource discovery portals?

- No data - only metadata
We don’t want to end up with this situation:

[Image of a comic strip showing the proliferation of standards, with the text:

SITUATION: THERE ARE 14 COMPETING STANDARDS.

14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES. YEAH!

SOON:

SITUATION: THERE ARE 15 COMPETING STANDARDS.]

http://xkcd.com/927/
WORK IN PROGRESS
Thanks for Listening

Contact: mikke.priddy@dans.knaw.nl & marion.wittenberg@dans.knaw.nl

Website: http://www.dwbproject.org/