XKOS
Extending SKOS for Describing Statistical Classifications
Dan Gillman, Franck Cotton, Yves Jaques
Overview

• What are statistical classifications and what are they used for?
• Why publish them as linked data and why SKOS is not enough?
• XKOS and what was added
• XKOS in action
What are statistical classifications?

- Hierarchical structures of concepts
  - One or several levels of details
- Covering a specific field
  - eg. economic activity, occupation...
  - Usually exhaustively and mutually exclusively
- Living objects
  - Published in major versions (classification schemes)
  - Minor modifications on a regular basis (notes)
- Linked by correspondence tables
Standard data model for classifications

Neuchâtel model, v2.1

[Diagram of data model]
Statistical classification examples

- Economic activities: ISIC, NAICS, NACE, ANZSIC, etc.
- Products
  - Industrial production or services, exchange of goods...
  - HS, CPC/CPA, PRODCOM, etc
- Occupations: ISCO, SOC, NOC...
- Health, Education, etc.
Examples of versions and correspondences

- **Versions**

- **Correspondences**
  - ISIC Rev.4 and ISIC Rev.3.1
  - ISIC Rev.4 and CPC Ver.2
  - ISIC Rev.4 and NACE Rev.2
Uses of classifications

- Data collection
  - Stratification of the population
  - Response values
    - Directly
    - Through automated coding

- Data dissemination
  - Coded dimensions in series, tables or cubes
  - Coded values for measures
Why publish classifications as linked data?

- They structure a great part of the statistical data
  - Dimensions or observations values
- They often form a system which is articulated at the international level
  - See an example on the next slide
- So they allow linking of data between sources and countries
Main classifications are internationally organized

<table>
<thead>
<tr>
<th>World level</th>
<th>EU level</th>
<th>National level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISIC</td>
<td>NACE</td>
<td>National versions of NACE</td>
</tr>
<tr>
<td>CPC</td>
<td>CPA</td>
<td>National versions of CPA</td>
</tr>
<tr>
<td>Goods</td>
<td>PRODCOM</td>
<td>National versions of PRODCOM</td>
</tr>
</tbody>
</table>

- ISIC: Internationally standardized industrial classification
- NACE: European classification for economic activities
- CPC: Classification of the productive community
- PRODCOM: Common nomenclature for industrial classification
- HS: Harmonized System
- SITC: Standard International Trade Classification
- CN: Common Nomenclature

Is the reference classification. Classifications are linked by the structure.

Is the reference classification. Classifications are linked by conversion table.

Classifications are linked by conversion tables.
Why is SKOS not enough?

- The classification levels
  - Not your ordinary skos:Collection
- The correspondences
  - You want to group them, document them
  - You sometimes need n-to-m relations
- The explanatory notes
  - Not your ordinary skos:note or skos:scopeNote
- The semantic relations
  - You need more precise semantics
XKOS

• Work mostly done during two workshops in Dagstuhl (September 2011 and October 2012)
  – Organized by the DDI Alliance
  – Statisticians, data archivists, semantic web specialists
  – The other line of work lead to Disco

• Drawing from
  – SKOS, Neuchâtel Model
  – ISO standards on terminology, such as ISO 704 and ISO 1087-1
XKOS: what is added?

• Classification levels
  – Subclass of skos:Collection
  – Bear a xkos:depth property (and others)
  – Organized as a RDF List attached to the classification scheme (skos:ConceptScheme)
XKOS: what is added?

• Correspondences
XKOS: what is added?

- Textual properties
  - Fixed-length labels
  - Explanatory notes
- Refined types
- Links
XKOS: semantic properties

- Hierarchical
  - generalizes/specializes
  - hasPart/isPartOf

- Associative
  - causal
    - Causes/causedBy
  - sequential
    - precedes/follows (previous/next)
    - temporal (before/after)
  - disjoint
XKOS in action: querying the ISIC

- The ISIC is published as HTML pages on the UN Statistics Division web site
- Application created to do extract and download the data as XKOS in a RDF triple store
XKOS in action: querying the ISIC

• It is then possible to execute SPARQL queries that would not be possible with only XKOS
  - Query on inclusion notes and not exclusion notes
    • Example: « wholesale of office furniture » (in 4659 and not in 4669)
  - Query on notes about correspondances
    • Example: « repair of weapons » between Rev3.1 and Rev.4 (moved from 2927 to 3311).
XKOS in action: browsing the NAF

- The NAF is the French national version of the NACE
- It is published in XKOS format on http://rdf.insee.fr
- A simple browser application was created to illustrate the possibilities of XKOS
- See demo