The Modernisation of the Central Metainformation System in Accordance with the LOD and FAIR Principles

aniko.mohay@ksh.hu
Anikó Mohay, Zsuzsana Mahr Oberné

12 April 2024, Paris
Outline

• The basic of the Central Metainformation System
  • Types of metadata
  • Users of metadata
  • Technical background
  • Relationships
  • Metadata-driven

• Linked Open Data – FAIR – Semantic Technology

https://creativecommons.org/licenses/by/4.0/
Outline

• The present situation in Central Metainformation System

• LOD architecture design
  • Interpretation of data
  • Linking of data
  • Accessibility and discovery of metadata

• Development - Conclusion
Types of Metadata

Grouping metadata according to their purpose and subject

• Structural metadata: describe individual pieces of information and their relationships.
• Descriptive metadata
  • Content identifier
  • Content description (definitions of concepts and nomenclature elements, which can be textual or algorithmic)
  • Metadata describing the data production process and applied methods
  • Metadata describing quality
Types of Metadata

Grouping metadata according to their purpose and subject

• Control metadata
  - Control of data collection
  - Control of data preparation
  - Control of processing
  - Control of data warehousing/querying
Users of Metadata

Who are the users?

• Non-expert users
  - Clear description

• Expert users
  - Detailed description (methodology, quality)
Users of Metadata

Who are the users?

• Statisticians, experts
  - Detailed description, connections

• IT experts responsible for data processing
  ✓ Description of production database
  ✓ Description of data warehouse

• Softwares, programmes
  - Information for operation
Types of Metadata

**DESCRIPTIVE METADATA**
Metadata describing process and quality concepts etc.

**Users**
- Non-expert users
- Expert users

**STRUCTURAL METADATA**
- Code List
- Variable

**Users**
- Data producers / statisticians
- Internal users

**META DATA USED BY SYSTEMS**
- Code List

**Users**
- Experts who have access
- IT systems
Basic of Central Metainformation System

Aims:

• Give information to users on content, quality and process method of data
• Ensure documentation for statisticians and experts
• Ensure operation information for softwares, programmes
• Meet international needs
• Metadata-driven:

Based on the descriptions above, it is evident that the metadata recorded in the Central Metainformation System currently control and support various tools for statistical business process. In the operation and development of the Central Metainformation System, we continue to strive for achieving metadata-driven processes, which means the determination of how software and programs operate through control metadata.
Technical Background

Base on:

- ORACLE SQL Developer

Three-tier environment:

- Developer
- Testing
- Production

User interface:

- SQL Forms

Development environment:

- Windows
LOD – FAIR DATA – Semantic Technology

Quality

Ontology Vocabulary

URI/IRI

SPARQL

RDF

Semantic Technology

Findable
Accessible
Interoperable
Reusable

Research DATA
Open DATA
(META) DATA
FAIR DATA

Source: 5-star Open Data (5stardata.info)
The Present Situation in Central Metainformation System

What is missing?

• Institutional Background

• Standards

• Metadata harvesting

• Metainformation on microdata level visible and accessible on the website

• Permalink (URI/IRI, Cool URI)

• Reference

• Advanced search possibility
Present Situation in Central Metainformation System

What is missing?

• Standard machine-readable format

• Data visualization

• Contextualisation (with concepts, other data or site, etc.)

• Structure reflection (hierarchy - classifications)

• Quality of metainformation

• Search engine optimisation

• Language equivalence
## LOD Architecture Design

### Interpretation of Metadata

Legend: ● = fulfill criterion, ○ = partially fulfill criterion, – = not fulfill

<table>
<thead>
<tr>
<th>Interpretation of data</th>
<th>Central Metainformation System</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fields describing and specifying metadata</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(Legal context, creation, responsible, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistent metadata</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Neutral metadata</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Metadata publication</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

### Tasks

- Defining mandatory fields
- Database cleansing
- Creation of quality indicators
- Handling homonyms and synonyms
- Utilization of multilingual thesaurys
- Usage terms

https://creativecommons.org/licenses/by/4.0/
## LOD Architecture Design

### Linking of Metadata

<table>
<thead>
<tr>
<th>Linking of data</th>
<th>Central Metainformation System</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversion of metadata</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Standardized metadata</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Structured metadata</td>
<td>⚫</td>
<td>⚫</td>
</tr>
<tr>
<td>Machine readable metadata</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The possibility of linking</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Legend: ⚫ = fulfill criterion, ○ = partially fulfill criterion, – not fulfill

### Tasks

- Conversion
- Machine readable metadata
- Linking metadata to other metadata and data (data enrichment)

[License](https://creativecommons.org/licenses/by/4.0/)
**LOD Architecture Design**

**Accessibility and discovery of metadata**

Tasks

- Development of Central Metainformation System
- Search Engine Optimization
- Defining entity types – URI/IRI
- Data enrichment
- Utilization of dictionaries and ontologies

<table>
<thead>
<tr>
<th>Accessibility and discovery of data</th>
<th>Central Metainformation System</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating metadata is easy</td>
<td>☐</td>
<td>–</td>
</tr>
<tr>
<td>Metadata is easily accessible,</td>
<td>☐</td>
<td>–</td>
</tr>
<tr>
<td>discoverable, and prominently</td>
<td></td>
<td></td>
</tr>
<tr>
<td>displayed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOD representation (search,</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>browse, query)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Downloading metadata</td>
<td>☐</td>
<td>-</td>
</tr>
<tr>
<td>Matching metadata in multiple</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>languages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storing metadata, version</td>
<td>☐</td>
<td>-</td>
</tr>
<tr>
<td>management</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# LOD Architecture Design

<table>
<thead>
<tr>
<th>ID denomination</th>
<th>ID</th>
<th>Denomination of concept</th>
<th>Short denomination of concept</th>
<th>Denomination of concept in English</th>
<th>Start of validity</th>
<th>End of validity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example</strong></td>
<td>5837</td>
<td>Bevásárlóközpontok</td>
<td>Üzletközpontok</td>
<td>Shopping centers</td>
<td>Shopping centers</td>
<td>04-JAN.-01</td>
</tr>
</tbody>
</table>

![Diagram](https://creativecommons.org/licenses/by/4.0/)

- **Bevásárlóközpont**
- **Üzletközpontok**
- **Shopping centers**

**shortName**

**englishName**
Questions of the operation/planning

What metadata have to be stored?
Who establishes metadata?
Who describes the metadata?
Who accepts metadata?

Viewpoint
Frequent use
Functions of the business processes (applications) have to be built on metadata
Development

Guarantee for the Quality

• Developing metainformation system in project team (meta expert, IT expert, statisticians, website: external users)
• Coordination of operation by a central unit (Methodology department or IT department)
• Training and consultancy for stakeholders (statisticians, IT and external users)
• Regular use of metadata in daily statistical works

https://creativecommons.org/licenses/by/4.0/
Conclusion

• Linking entities
• Interpretation of chatbot RDFs
• Assignment of terms of use
• Linking to other datasets
Future
Thank You
For Your Attention!

aniko.mohay@ksh.hu