HUNGARIAN CENTRAL STATISTICAL OFFICE

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

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

- Control metadata

  ❑Control of data collection

  ❑Control of data preparation

  ❑Control of processing

  ❑Control of data warehousing/querying
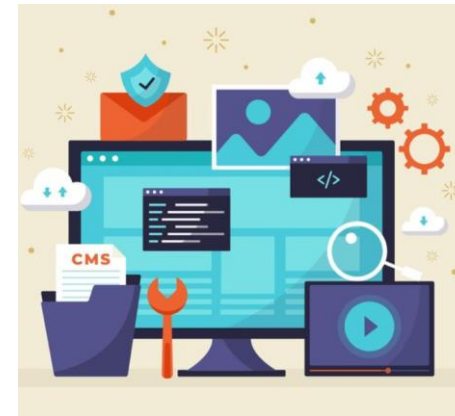
Users of Metadata

Who are the users?

External users

- Non-expert users

   - Clear description

- Expert users

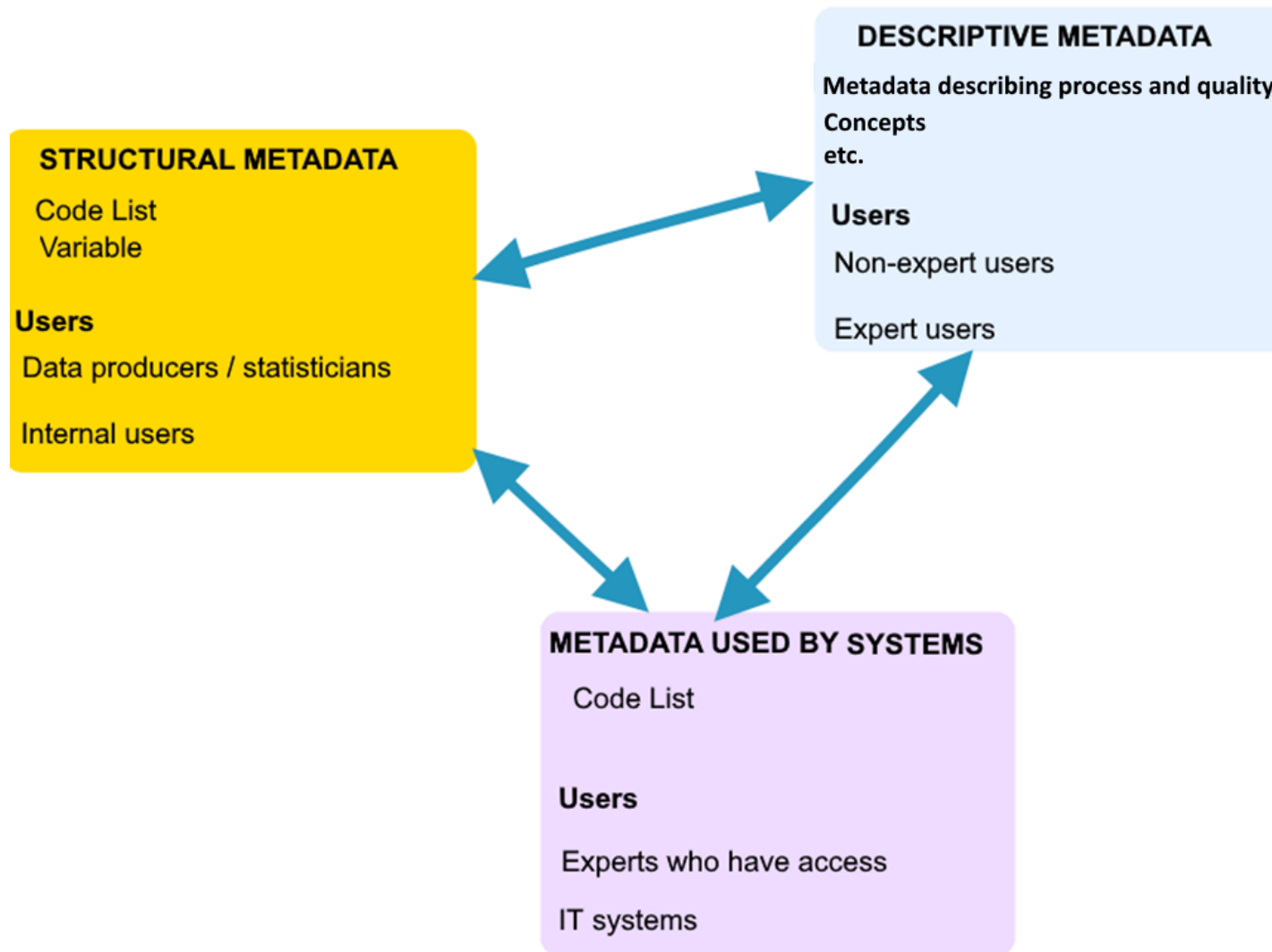   - Detailed description (methodology, quality)

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

7

Types of Metadata

STRUCTURAL METADATA

Code List
Variable

Users

Data producers / statisticians

Internal users

DESCRIPTIVE METADATA

Metadata describing process and quality
Concepts
etc.

Users

Non-expert users

Expert users

METADATA 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

- **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.

10

Technical Background

Three-tier environment:
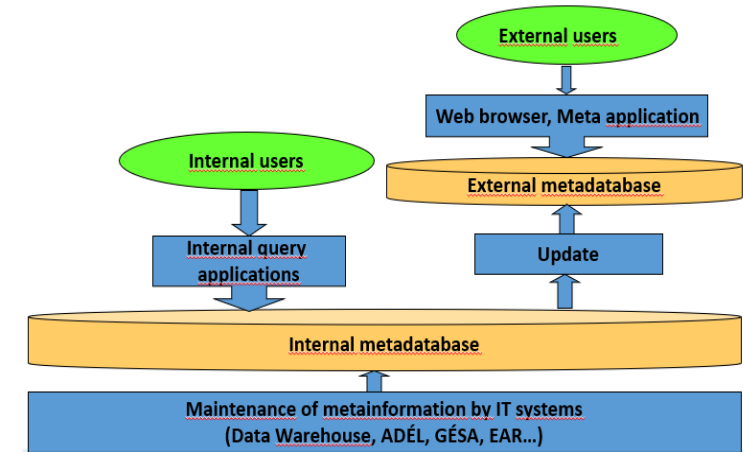
- Developer
- Testing
- Production

Base on:

- ORACLE SQL Developer

User interface:

- SQL Forms

Development envinronment:

- Windows

11

LOD – FAIR DATA – Semantic Technology
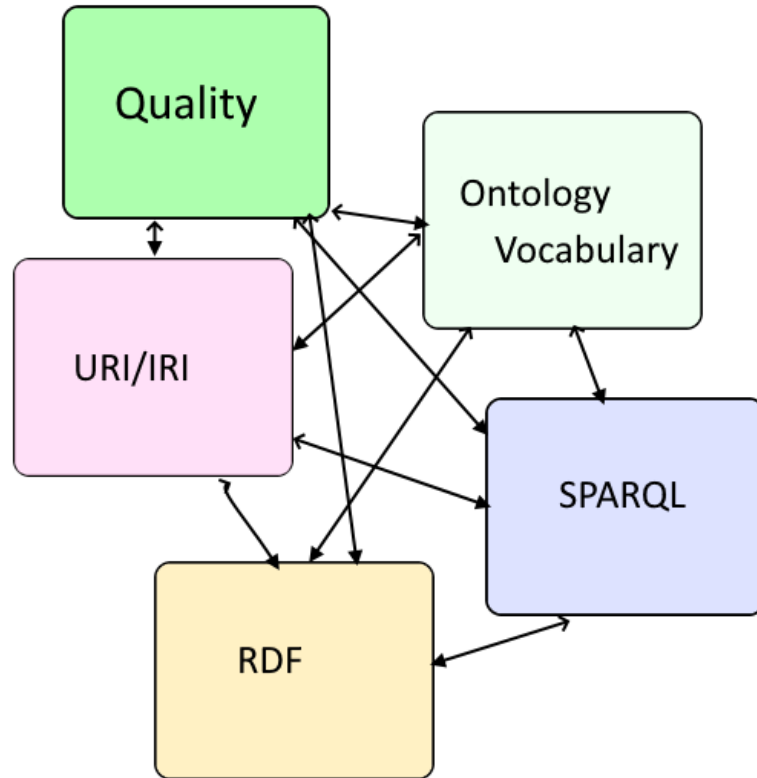
Semantic Technology

Findable

Accessible

Interoperable

Reusable

Research DATA

Open DATA

(META) DATA

FAIR DATA

Source: 5-star Open Data (5stardata.info)

KSH

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

Interpretation of data	Central Metainformation System	Website
Fields describing and specifying metadata (Legal context, creation, responsible, etc.)	○	○
Consistent metadata	–	–
Neutral metadata	●	●
Metadata publication	○	○

Tasks

- Defining mandatory fields

- Database cleansing

- Creation of quality indicators

- Handling homonyms and synonyms

- Utilization of multilingual thesaury

- Usage terms

KSH

LOD Architecture Design

Linking of Metadata

Linking of data
Legend: ●=fulfill criterion, ○ = partially fulfill criterion, – not fulfill

Linking of data	Central Metainformation System	Website
Conversion of metadata	–	–
Standardized metadata	–	–
Structured metadata	●	●
Machine readable metadata	○	○
The possibility of linking	○	○

Tasks

- Conversion

- Machine readable metadata

- Linking metadata to other metadata and data (data enrichment)

LOD Architecture Design

Accessibility and discovery of metadata

Accessibility and discovery of metadata

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

Accessibility and discovery of data	Central Metainformation System	Website
Creating metadata is easy	○	–
Metadata is easily accessible, discoverable, and prominently displayed	○	–
LOD representation (search, browse, query)	-	-
Downloading metadata	○	-
Matching metadata in multiple languages	-	-
Storing metadata, version management	○	-

Tasks

- Development of Central Metainformation System

- Search Engine Optimization

- Defininig entity types – URI/IRI

- Data enrichment

- Utilization of dictionaries and ontologies

18

LOD Architecture Design

	FAZON	MEGNEV	ROVMEGNEV	A_R_MEGNEV	ANGMEGNEV	TOL	IG
ID denomination	ID	Denomination of concept	Short denomination of concept	Short denomination of concept in English	Denomination of concept in English	Start of validity	End of validity
Example	5837	Bevásárló-központok	Üzletköz-pontok	Shopping centers	Shopping centers	04-JAN.-01	10-JÚL. -01
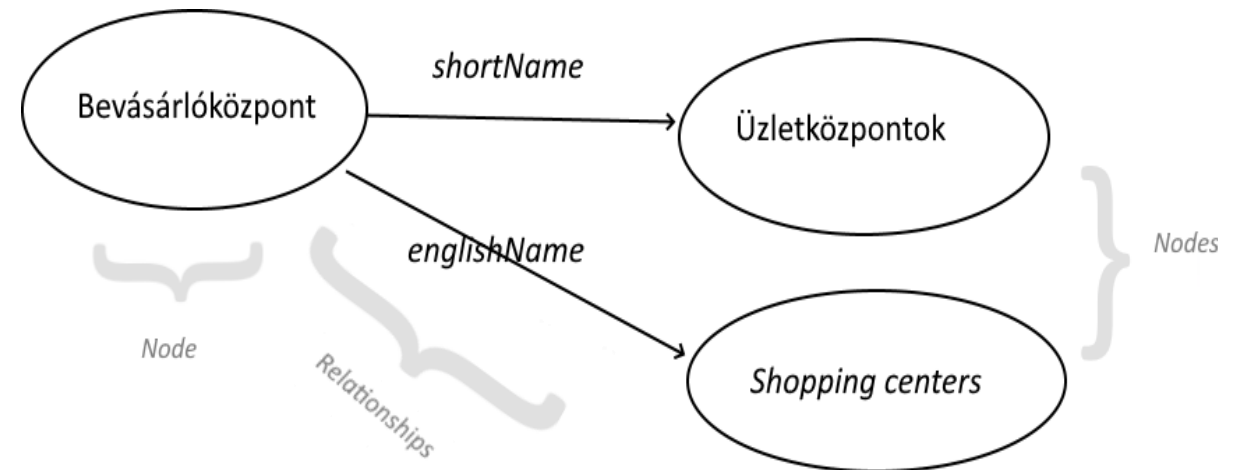
LOD Architecture Design

Cost of data

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

Cost of data	Central Metainformation System	Website
The cost of metadata production is low	○	–
The time required for metadata production is low	○	–

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

Conclusion

- Linking entities

- Interpretation of chatbot RDFs

- Assignment of terms of use

- Linking to other datasets

Future

23
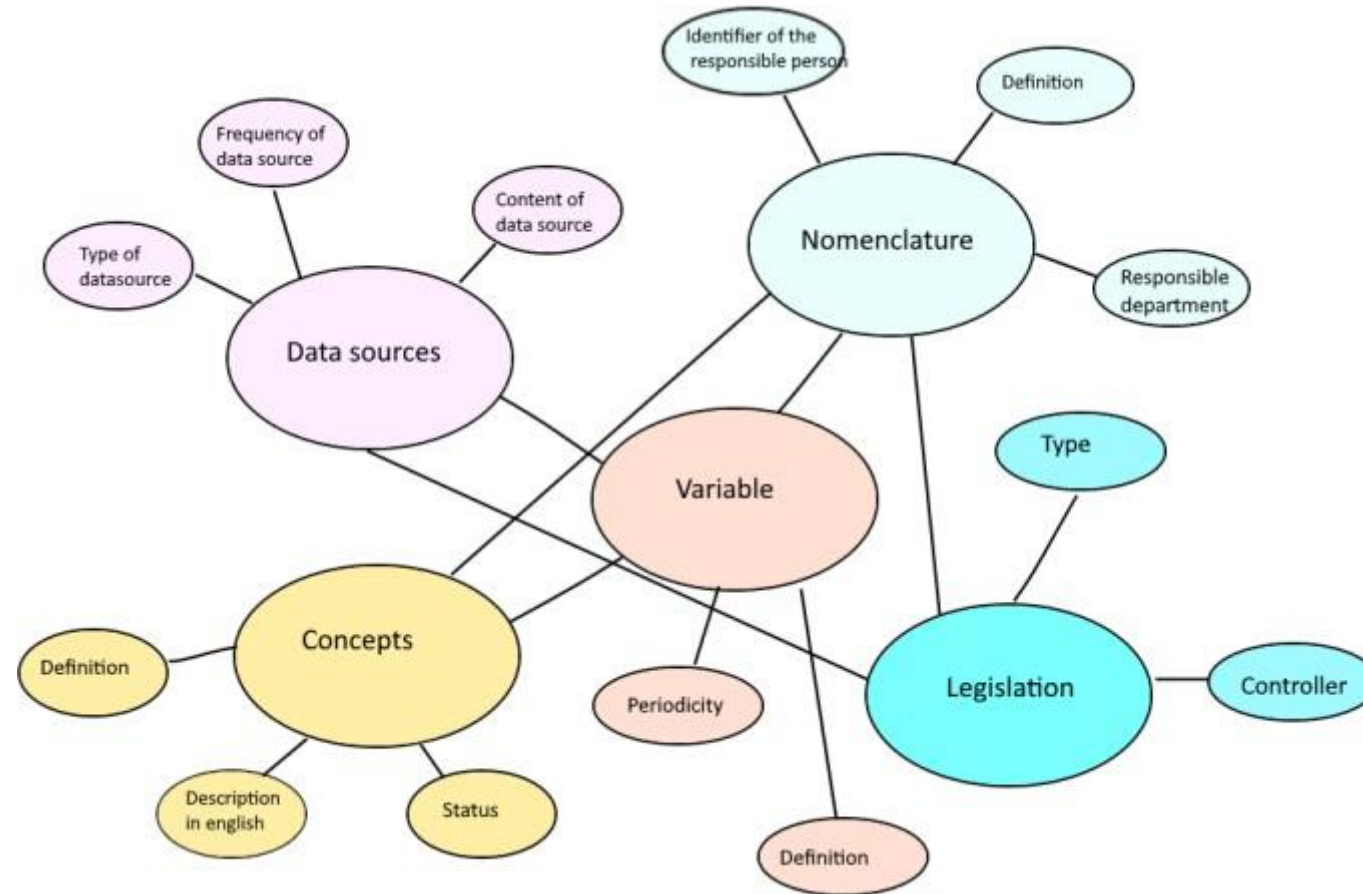
Thank You
For Your Attention!

aniko.mohay@ksh.hu