Elena Shulman, PhD, MLIS

D.E.Solution sprl

BUILDING A COMMON LANGUAGE FOR SEMANTIC INTEROPERABILITY

Enabling Semantic Interoperability

Example: Digital transformation of European Commission (EC) public facing information system

- A large variety of websites with their own vocabularies or no vocabularies
- More than 50 different sites not speaking the same machine understandable language

Semantic Interoperability Best Practices

- Holistic approach looking at the task from four vantage points:
  - context
  - content
  - end-user searching for content
  - tagger
- Identify systems to manage controlled vocabularies
  - Establish a centralized system if one is not available
  - Information models, documentation, training, and tools

Core questions to consider

Context: Is interoperability important in this context? What kind of services will a common language need to support?

Content: What kind of content needs to be described and what needs to be said about that content?

End-user: Who is the target audience? What is known about user behavior and user satisfaction?

Taggers: Who is going to be tagging the content and what information and tools do they need? Are they professional librarians? Are they authors who know their subject matter? Are they webmasters?

Context: EC Digital Strategy

Make content ‘open’, ‘borderless’ and ‘interoperable’, use **centralized** services, **re-use**…

But what does it mean to make content ‘open’?

- Just publishing it on a public website?

Interoperability: Enabled when content providers use a common language to describe content

Vocabularies are building blocks of a **common language**

A ‘common’ set of controlled vocabularies

The Publications Office of the EU already managed a number of relevant vocabularies and provided a platform to manage customized vocabularies

Having a central managing authority is essential

Specific requirements of public websites (and who would be doing the tagging) pointed to a need for a customised locally owned but interoperable vocabulary.

Having a common understanding/training for using tagging -

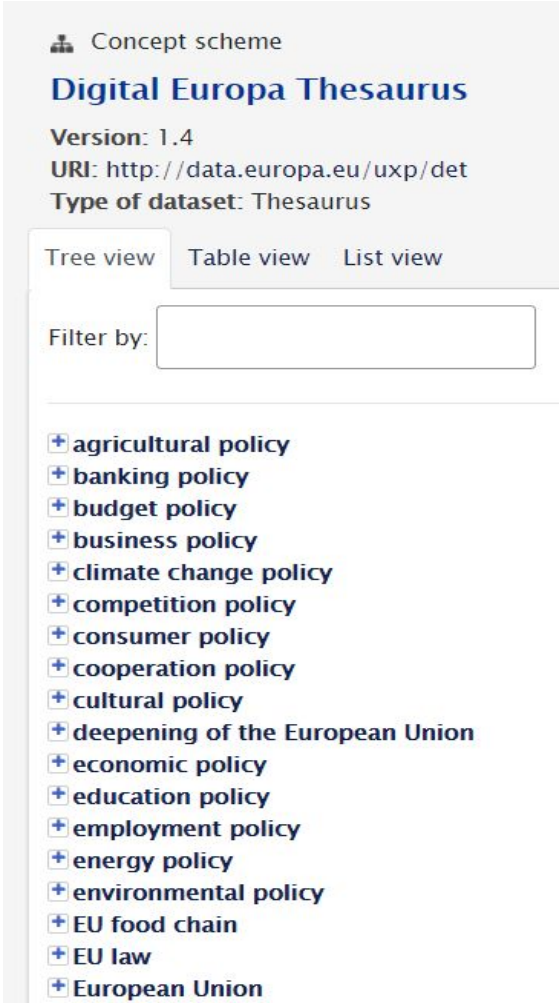
- The objective of tagging is to retrieve similar content together

New subject vocabulary: Digital Europa Thesaurus

A controlled vocabulary with broader/narrower (parent/child) term relationships

Concepts are available in 24 languages

May or may not have “synonyms” to point to the correct, preferred terms (also in multiple languages).



Basic question

Do you need a new taxonomy, or should you reuse an already existing taxonomy that provides a common language with other content?

Broader context: It is important to think of the content being discoverable in systems outside your own.

Is this vocabulary understandable by machines and humans outside of the context in which the vocabulary was initially used?

If you choose very narrow or ambiguous descriptors, it will be difficult for other systems to re-use the terminology to deliver this content in other platforms.

If the terms are very generic, your content will be 'lost' in an ocean of similarly tagged content when merged with other databases.

What about Artificial Intelligence?

Metadata (controlled vocabularies) is the foundation upon which AI and machine learning produce better user experience.

Artificial Intelligence cannot initially accomplish what a human can by creating categorizations that are context specific.

The return on investment (ROI) for building and managing a vocabulary is the scenario where it is used by an author/tagger to describe the main subject matter of content with terminology that may not be present (or sparingly present) in the content being described so that similar content can be retrieved together.

This adds an extra layer of meaning to the metadata that is not possible to generate by mining the text for common terminology.

What ensures consistent tagging?

Concepts must be self-explanatory (given the larger context and content) and unambiguous (not necessarily narrow) and must stand on their own.

User-centric tagging tools that make it easy for taggers to see hierarchies and relationship between concepts (and any available definitions) to understand their context and rules for using them

Training - shared understanding of purpose

Tagging strategy

Taxonomy governance, tools and management

1. Taxonomy documentation provides:

An explanation of the taxonomy's context, purpose, ownership, etc.

Taxonomy maintenance and editorial style policies/guidelines for its evolution

Tagging strategy (as opposed to vocabulary building strategy) with clear guidelines using the taxonomies to tag content

2. Taxonomy building & maintenance tools

Most desirable but complex - Taxonomy/thesaurus management platforms:

eg: VocBench - A central authoritative database for hosting and sharing taxonomies

Performance in the long-term

1. Tools and workflows must be in place for controlled evolution of vocabularies
2. Tools and training must be provided for taggers as an ongoing process
3. The semantic interoperability ecosystem must be continuously managed

ENDORSE.

THE EUROPEAN DATA CONFERENCE ON REFERENCE DATA AND SEMANTICS

Publications Office
of the European Union

ISA²