Vocabulary Evolution and Management for the Digital Cultural Heritage Sector

Emma Tonkin
Interoperability Focus

UKOLN is supported by:

www.ukoln.ac.uk

A centre of expertise in digital information management
About UKOLN

- Digital information management -
- Provide advice and services to communities inc.:
  - Library
  - Information
  - Education
  - Cultural Heritage
- Jointly funded by the MLA and the JISC
Starting with the obvious...

• What is metadata?
  – “It's been in the Guardian, so everybody knows what metadata is”

• Document surrogates
  – Why?
  – What?
Metadata: data about data

• What information do we need to know – or share – about this object?
• Lots of details are irrelevant
• Lots are potentially relevant
• Some are observations, some are interpretations or reactions
• What details do we choose?
Elements, Schemas and APs

- Each detail is an element
- Each collection of facts is a schema
- Each description characterising a way of using a collection is an application profile (DC)
Types of metadata

- DC shares a lot with other types of metadata, on this level
- Shared aims:
  - Promotion of reuse;
  - Encouraging consistent use within a context;
  - High quality metadata
How do we ensure good-quality metadata?

• Well-trained staff for indexing work
• Lots of readable documentation:
  – The metadata
  – Its use (including examples)
  – The underlying conceptual model-how it all fits together
A familiar model?

• Jakob Nielsen: “Speak the users' language”
• But isn't this an argument against reusability?
  – In digital cultural heritage, there are many specialised groups
  – Each has a specialised use of vocabulary
Effects of specialisation

• Inconsistency:
  – Use of language can vary between and within groups or contexts (inter-indexer)
  – Furthermore, an individual asked to contribute metadata for the same object on two occasions may not produce the same metadata twice! (intra-indexer)
How do we mitigate all this?

• One familiar structure is the metadata schema registry, to support the processes of:
  – Developing APs/schemas/elements
  – Publishing what has been completed
  – Supporting people in its use
  – Sharing experiences and notes
  – Encouraging reuse or partial reuse
How is an MSR technically structured?

• Could be an Excel sheet
• Could be an XML document
• Could be an RDF-backed aggregation of a back-end service and a collection of front-end interfaces, plus links to user-targeted services.

See ISO 11179
The process of creating a metadata AP

- Collaborative: many interested parties.
- A compromise between relevant groups:
  - End-users
  - Developers (technical feasibility)
  - Domain experts
  - etc
Negotiating a compromise

Developers: practical interests, computability, etc

Domain experts: accurate representation using domain-specific terminology

End-users: adequate representation - using appropriate terminology and describing the right details

www.ukoln.ac.uk
A centre of expertise in digital information management
Maintenance

• Things change.
  – “Change is a constant“ - OCLC, about the Dewey Decimal Classification
• Models reflect our preconceptions; attitudes, theories, understanding
• Change is often prescribed, or can be observed in action
Comparison to an informal approach: social tagging

• Currently very popular
  – Cheerful anarchy!
• Annotations containing whatever comes into the user's mind, given the interface and a resource
• Easy for the user to write, but computationally difficult
A spectrum of annotations

Explicit, formal, fully defined, computationally tractable
RDF/OWL
DC metadata
Microformats
Social Tagging

Implicit, informal, contextual, more difficult to analyse
What's the Right Answer?

• No one right answer
• Different contexts – diverse successful approaches.
• Situated in a cultural/organisational context
• It's all about compromise and understanding your requirements, so often the right answer will involve several technologies or approaches
Discussion

- Advantages and disadvantages of each type of metadata
- Interface considerations and opportunities
- Making the most of the metadata -
  - and the communities!
Example: Steve.museum

• Jennifer Trant's work on linking of social tagging with terms provided by trained cataloguers
• Shows the benefits and disadvantages of each, particularly accessibility benefits through “speaking the users' language”
Conclusion

• There exist useful tools and best practices for managing vocabularies: the metadata registry concept, for example
• It's important to use the right approach for each context, which means understanding the context and stakeholders
• There are excellent case studies available that give ideas, recommendations and advice

www.ukoln.ac.uk