

Missing Pieces in the Global Metadata Landscape

Stuart Weibel
Tokyo
December 2010

Fifteen years later...

Dublin Core metadata emerged 15 years ago, and one of the major hopes of that early effort has gone unrealized:

A world-wide, Web-wide metadata infrastructure effective across domains, cultures, languages and political boundaries

We have succeeded to a degree, but...

What is still missing?

Three Questions

1. What is the most important missing infrastructural link in establishing globally interoperable metadata systems?
2. Are there conceptual or theoretical impediments that remain to be solved?
3. What are the organizational or institutional impediments that must be overcome to improve prospects?

The Respondents:

Some practitioners, some theoreticians,
mostly who were at DC 2010

Emmanuelle Bermes, BNF

Alex Haffner, DDB

Karen Coyle

Diane Hillmann & Jon Phipps

Gordon Dunshire

Andy Powell, Eduserv

Kai Eckert, Mannheim
University

Jodi Schneider, DERI

Muriel Foulonneau

Paul Walk, UKOLN

Jeff Young, OCLC

What is the most important missing infrastructural link for globally interoperable metadata systems?

- Tools to support and encourage the reuse of terms, concepts, schemas, ontologies (eg., metadata registries, and more)
- Widespread, cross disciplinary adoption of a common metadata approach (Linked Data)
- Query languages for the open web (SPARQL) are not fully mature
- Trust and provenance Infrastructure
- Nothing missing... just use RDF, Linked Data, and the open web. The key is broad adoption, and that requires better tools and applications. Its a ***social*** problem, not a technical problem

Missing Infrastructure (continued)

- The ability to identify 'universals' across languages, disciplines, and cultures - revive Ranganathan's facets?
- Terminology services have long been proposed as important services, but they are expensive to create, curate, and manage, and the economic models are weak
- Stuff that does not work is often obvious. We need usage data to see what **does** work, and amplify it

The soundbites:

- The Web has what we need in infrastructure (mostly)
- The key is broad adoption of standards, conventions, and common practice

Are there conceptual or theoretical impediments that remain to be solved?

- Better research and metrics about what works, what is used
- Research on the management and use of decentralized data
- Choosing vocabularies, or aligning related vocabularies is not well understood
- Working across high level conceptual models (such as RDF, CIDOC, CRM) is difficult: Models are important, but they also impose create silos that are difficult to connect. There is no universal data model

Theoretical impediments (continued)

- Standardized ways to design, describe, and declare application profiles (or ontologies, or domain models) are necessary if reuse is to happen
- The declaration and management of provenance information. Mashups are fine, but we need to be able to unmash, as well, with transparent accountability about the provenance of data

What are the organizational or institutional impediments that must be overcome to improve prospects?

Economic models can cripple broad use. Metadata is costly to create, and those who create or manage it want to be paid for its use (which, in turn inhibits use, and stifles innovation)

Centralized curation is important for quality and persistence, but is costly, and leads to protectionism

Funding of ongoing services (metadata maintenance, terminology services, authority services) is difficult to secure, and even harder to keep going

Institutional Impediments (continued)

Getting organizations to move in similar directions (adopting a common model for metadata distribution or disclosure) is very difficult: **Institutional inertial** is always a problem, especially when the paths are less than clear.

Standards solutions need to be open, and created and maintained by a broad cross section of practitioners and theoreticians. Leave domain silos and join the web!

My favorite version of the institutional impediments question

Fear:

- of loss of control
- of commercial exploitation of *commons* data
- Is the semantic web just another *technology du jour* ?

Ignorance:

- of the power of current semantic enabling technologies
- of terminology issues - common terminologies for modeling, for example

Bemusement:

- about modeling
- lack of agreement on recommended practices

The short version of this talk

- Better tools and applications
- Alignment of economic models and funding with the new technological environment
- Identify research and applications that will promote better sharing capabilities
- Understand and address provenance issues
- Improve understanding and acceptance of the benefits (and weaknesses) of data modeling
- Let no proposed standard escape the critical eye of working code that solves problems

どうもありがとうございました！

Thank you for your attention

And thanks to all my respondents,
who kindly took the time to reply to my questions.

I hope I did not distort the clarity of their thinking