Dublin Core at 21

Reminiscence and Reflections

Stuart Weibel
OCLC Research (retired)
Dublin Core at 21

I  How did we get started, and why?

II  The conditions for success?

III  The major impediments

IV  Personal reflections
When I started at OCLC in 1985:

- I was 4 years away from my first email address
- 1 GB of storage was about $50,000 and was the size of a small refrigerator
- Cameras still used film
- Cell phones were the size of briefcases
- MARC cataloging stood alone as the discovery tool for intellectual assets of libraries
- No end-user access to the global library catalogs, and Google hadn’t been even dreamed of
In 1993…

- The Web exploded into popular awareness in January, 1993

- Terry Noreault, then Director of Research at OCLC, asked if I might be interested in serving as liaison between the OCLC and the Web community?

- I was, and the two of us visited NCSA a few weeks later

- The first World Wide Web conference took place at CERN, near Geneva in May

- The second WWW conference took place in Chicago in October
WWW-2: Chicago

- Congress Hotel, booked on short notice

- run down, shabby downtown hotel infamous for having been the site of the 1968 Democratic National Convention (ouch)

- 500,000 web documents in the world at that time, without much search capability
Casual Coffee-break discussion between sessions

- Joseph Hardin,
  Yuri Rubinski,
  Terry Noreault,
  Eric Miller, and myself

- “There must be something we can do to make things easier to find…”

- Yuri turned to me and suggested a workshop? I looked at Terry, he smiled and nodded, and I said “Eric and I will do that”
Fools Rush In

- WWW-II: mid October, WWW-III was planned for April in Darmstadt
- We wanted to present results in Darmstadt: 6 months to organize and distill the results
DC-1: An Invitational Workshop
In Dublin, Ohio

- 52 geeks, freaks, and people with sensible shoes

- An amalgam of internet types, content and markup specialists, and librarians

- Goal: promote a consensus on a **core set of metadata elements** to describe networked resources.

- A set of 13 elements emerged from the workshop, later expanded to 15
Founding Principles

- Consensus Driven
- Cross Disciplinary but Web-centric
- **International** and Multi-lingual

- The first word in metadata, but NOT the last: an **extensible** element set
We reported the workshop results at WWW-3 in Darmstadt.

Uhmm what’s next? (no clue)

Lorcan Dempsey, then of UKOLN at the University of Bath called… might we be interested in a follow-up workshop in the UK?
One thing led to another

- The workshop led to a workshop series, which led to the Dublin Core Metadata Initiative, and the workshops became conferences.
- Held in North America, Europe, Asia, and South America to reflect the global nature of the Web and its stakeholders
- DCMI workshops and conferences became the premier events for both metadata practitioners and theoreticians
Metadata is about *social* as well as *technical* engineering

- The products:
  - **Semantics** - What do you want to say about a digital resource?
  - **Syntax** - What are the grammatical and technical guidelines for expressing semantics?
  - **Consensus** - How can many stakeholders and communities of practice be brought together under a single tent, so as to increase the *network effects* of the standards?
  - **Community** - Provide a gathering place where researchers and practitioners may explore ideas, share successes, and nurture one another
Part II: The necessary conditions for success

- An institutional Home
- Need
- Juice
- Standardization
- People of good will
- A global perspective
OCLC: An Institutional Home

- OCLC was DCMI’s home for 12 years
- Probably no other possible home for its beginning
  - International stature
  - Trusted neutral party (especially from the POV of the web community)
  - Deep expertise in resource description
  - Willingness to invest
The Stages of Enthusiasm

I. Benign neglect
   - Dublin what? Is that ours?

II. Oh! They are talking about us!
   - DCMI acquired a measure of Web-luster, and lots of good publicity
   - DCMI was a major bridge between the library community and the Web community for many years

III. How much longer (and how much) do we have to pay for this?
Terry Noreault, Director of Research for DCMI’s early years was a great supporter of the effort, and nurtured with a light touch.

Jay Jordan, OCLC’s CEO from 1998 to 2012, was a strong supporter of DCMI.

Many others, especially in Research: Eric Miller, Jean Godby, Eric Childress, Todd Matola, and more.
Is the need real?

- We had the first workshop... ok... now what?
- Lorcan’s midsummer call from UKOLN
- Roland Schwänzl and Judith Plümer - an independent deployment of DC metadata for the description of mathematics archives (1996) - They saw it and recognized it as a useful solution and implemented it
- Researchers in the UK, Germany, the Nordic countries, and Australia were eager to participate and contribute
Stakeholders

- A global community of practice coalesced around DCMI, starting with projects, but growing into communities with shared understanding of resource description needs, and a recognition of the benefits of sharing in the creation of solutions

- museums, professional societies, government agencies, educational technologists, and more (including libraries)
Juice

- DCMI was spawned by the Web, the need for shared description standards for electronic resources.

- To prosper it needed practitioners — people to populate metadata stores and to support appropriate infrastructure.

- But it also needed a community of researchers and developers to experiment, to create the infrastructure, to explore the limits of the technology within the rapidly changing Web.
Standardization

- Standardization of the Dublin Core was time consuming, often contentious, but necessary to promote adoption by others:
  - IETF RFC (Internet Engineering Task Force)
  - NISO (North American)
  - ISO (International)
People of good will and common purpose

- DC brought together librarians, web technologists, and content/domain specialists from many communities, cultures, and languages.
- Reaching consensus was often fraught with conflict and division.
- A great deal of work was done in workshop sessions, over lap-tops in bars at all hours, and by email and teleconference.
Work-at-a-distance requires trust
Establishing trust requires presence

- The disagreements were often profound, and complicated by different languages and different disciplines

- The workshops were intense, often messy affairs: we always managed to emerge with what we needed to move on to the next steps

- The essence of our progress came from a deep-seated desire to contribute to the common good
International Perspective

- The Web demands a global perspective, hospitable to the diversity of languages and culture - technically demanding
- From the outset, we strove to respect this diversity by holding workshops, tutorials, and conferences wherever we found concentrations of metadata development
- DCMI maintains core documents in multiple languages
Part III: The major impediments

- Data Modeling
- Creeping Elegance
- Business Models
- A rapidly evolving (and complicated) technological environment
- Inherent friction and complexity as a global infrastructure
Data Modeling

- Entity-relationship models define key **concepts** and **things** and the **relationships** among them.

- *Reality is messy*: A model should be designed to **solve a problem**, not emulate **reality** (which means you have to understand the problem very well indeed).

- Identifying the **appropriate level of abstraction** is an art.
FRBR is a prominent example of data modeling

- FRBR is the most prominent example of modeling in the library world

- The world of bibliography, which was large to begin with and is MUCH larger in the digital age

- How long have we been working on FRBR?
The Problem with Models

- Matching the complexity of a model to diverse and evolving problems is difficult and full of compromises.
  - too much complexity leads to failure: *creeping elegance*
  - too little complexity leads to insufficient richness that leaves the model unable to address the problem
- Knowing when to stop is an art
Creeping Elegance
(and it isn’t just true for modeling)

- “We have met the enemy, and he is us”
- There is a tricky balance between simplicity and functionality.
- We did not always find that balance, or even a clearly defined abstraction of the problem
Business Models

- The business model of a library is to make expensive resources look free: to remove the impediment of cost from the user and distribute that cost through a shared burden (usually via taxes).

- The business model of the web is to remove the cost barriers to distributed computing through economies of scale that reduce the burden to a user through government subsidies and modest user fees (your telephone carrier or cable provider bills).
What is the business model?

- I never figured this out
- The products (the standards and consensus process) need to be free, else adoption is hindered
- It is too general-purpose to be adopted by a given organization
- Funding agencies like the NSF will support innovation, but don’t like to pay for maintenance
- Sustaining membership fees are difficult to attract
The Web of the 1990 was a rapidly moving target

- HTML, XML, XML-Schemas, RDF, Linked Data…

- Complicated, highly technical standards that were changing and evolving at high rates

- From Day One, we tried to keep semantics and syntax separate, but inevitably they became entangled, and it caused unending confusion
The Friction of Global Infrastructure

- DC *had* to be a global standard
- DC *had* to be respectful of language and cultural diversity
- Traveling all over the globe on behalf of metadata is costly, both in terms of money and in terms of the personal costs associated with such travel.
Part III
Personal Reflections

- Leading DCMI through its first decade or so was the greatest professional privilege of my life.

- As a group, we made many mistakes, missed opportunities, wandered down unproductive pathways, and failed to understand our priorities clearly enough.
But at the end of the day…

- There was widespread recognition of need for such a standard: not just for search, but resource administration and rights management.

- The people were — are — exceptional in their dedication, in their competence, and in their passionate commitment to a common purpose: to make the Web and their societies work better by making digital information more accessible and more manageable.

- We are still here, more than two decades later
Thank you

- Stuart Weibel (on Facebook)
- stuartweibel (on twitter)
- stuart.weibel@gmail.com