2020 Hindsight: Reflections on 25 years of Metadata

DCMI Virtual Conference

Stuart Weibel   **Seattle, Washington, USA**   **September 2020**

Welcome to the first DCMI Virtual Conference

…In the midst of COVID…

- Some early history

- The conditions for success

- The metaphors in the foundation

- Some major challenges

- What we have achieved

- Hopes for the future

Who I am

A fortunate person who was in the right place at the right time

- 25 years in OCLC Research

- Managed DCMI for the first decade

- Please don't call me 'Dad'

- Now, I look after a wooden boat

In a normal year…
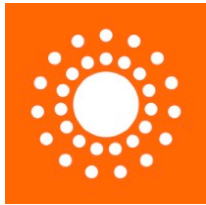
I'd be at the Port Townsend Wooden Boat Festival today

Mark Twain (may have) said:

I am always embarrassed by praise…
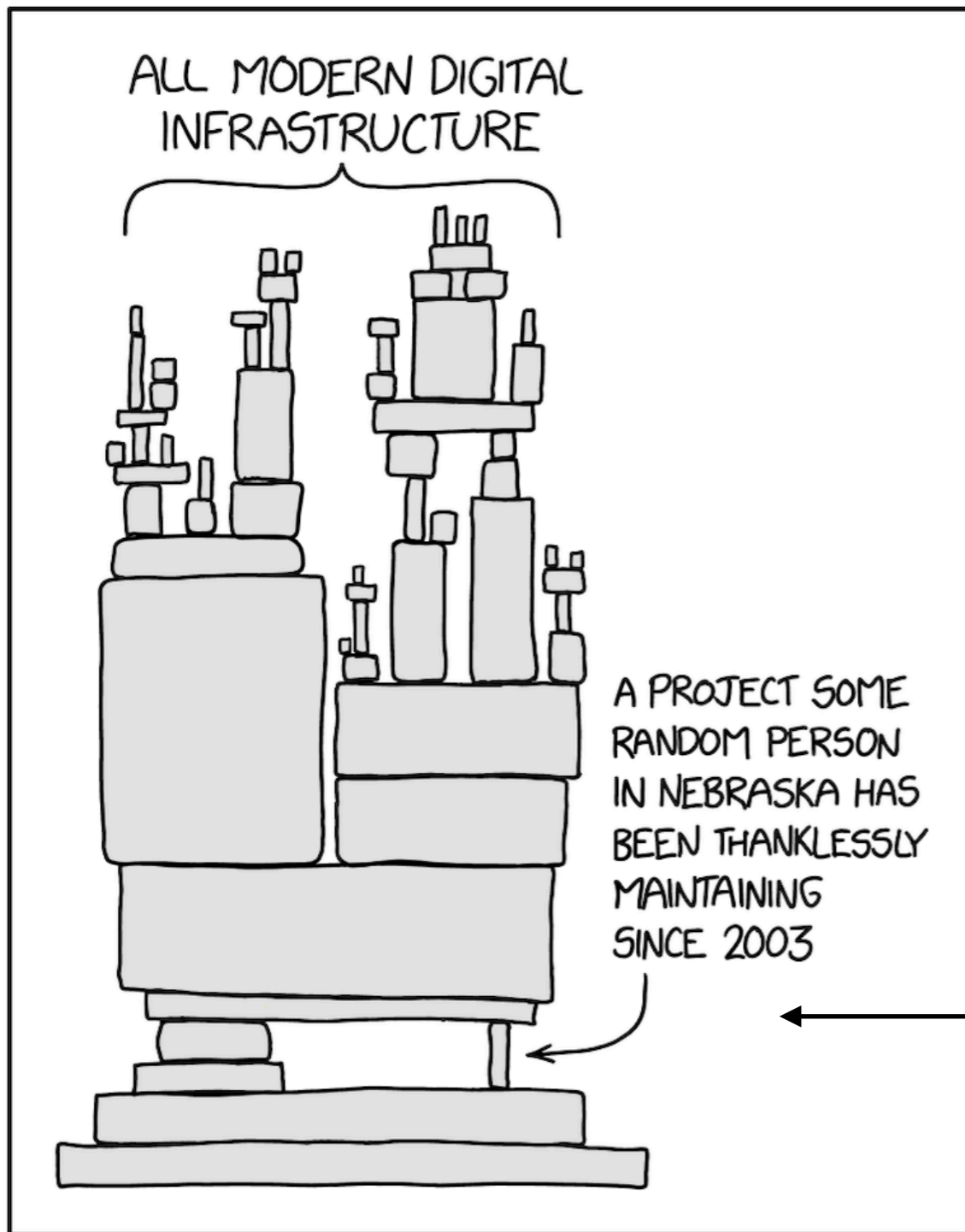
I never feel like they've said enough.

DC-4, Canberra, 1997

The XKCD layered model of digital infrastructure

https://xkcd.com/2347

Dublin Core Metadata standards are among the longest-lasting Web technologies

How the Dublin Core got started

- The Second International World Wide Web Conference in Chicago, October, 1994

- Hall-conversation between sessions

  - Yuri Rubinski (SoftQuad)

  - Joseph Hardin (NCSA)

  - Terry Noreault (OCLC)

  - Eric Miller (OCLC)

  - Me (OCLC

Conditions for success: 1

OCLC was the Right Place at the Right Time

- Global, not-for-profit library technology company

- Broad international trust

- Terry Noreault, Director of Research, supported DCMI

- Management supported it

- Jay Jordan (CEO) was an enthusiastic supporter

Conditions for Success: 2

A problem set of global proportions

- Chicago conference (Mosaic and the Web): the entire Web was on the order of 100,000 resources

- The Web was the Wild West — the hottest ticket in town

- Exponential growth

- Digital 'stuff' was already hard to find

- Metadata for description, discovery, and management of digital assets was becoming mission critical

Conditions for Success: 3

Passionate people trying to make the world work better

- Lots of people out there with expertise and the passion to make the Web work better

- Governments, labs, businesses recognized the importance of the Web, and the missing pieces

How we did our work

- Invitational workshops, then…

- Open workshops, then…

- Conferences

- International locales:

  - US, UK, Australia, Finland, Canada, Japan, Germany, Korea, Singapore, Italy, Mexico, Spain, Portugal…

- Email, email, email…

OCLC/NCSA Metadata Workshop

March 1-3, 1995

- 13 elements (later 15)

  - Intrinsic

  - Extensible

  - Optional

  - Repeatable

  - Modifiable

- Syntax independent

- 52 librarians, Web technologists, domain specialists agreed on the essential features of resource metadata

- These principles have stood the test of time

- We had a brand: Dublin Core

The Cardinal Rule:
Thou shalt not discuss Semantics and Syntax in the same room

- Conflating semantics and syntax was dangerous and confusing

- We (rightly) did not want to limit the expression of metadata to any given syntactic expression (the great mistake of MARC standards)

The First Dublin Core Publication

d-Lib magazine

Metadata: The Foundations of Resource Description

Stuart Weibel
Office of Research, OCLC Online Computer Library Center, Inc.
weibel@oclc.org

- **Subject:** The topic addressed by the work
- **Title:** The name of the object
- **Author:** The person(s) primarily responsible for the intellectual content of the object
- **Publisher:** The agent or agency responsible for making the object available
- **OtherAgent:** The person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work
- **Date:** The date of publication
- **ObjectType:** The genre of the object, such as novel, poem, or dictionary
- **Form:** The physical manifestation of the object, such as Postscript file or Windows executable file
- **Identifier:** String or number used to uniquely identify the object
- **Relation:** Relationship to other objects
- **Source:** Objects, either print or electronic, from which this object is derived, if applicable
- **Language:** Language of the intellectual content
- **Coverage:** The spatial locations and temporal durations characteristic of the object

We had some silly thoughts, too

- We imagined that authors would be happy to create and include metadata in their documents, and we tried to make it easy.

- We were dead wrong

- Author-created metadata has never been a significant contribution

Ever since…

- The basic semantics were (mostly) fine

- The syntax for declaring and sharing was a moving target

  HTML… XML…RDF

- The structure… the architecture… of metadata has always been the slippery part

- Communicating what we were doing to others has been a large part of the challenge

Web Infrastructure

- HTML, XML, RDF, managing schemas

- Open Data

- Ontologies

- We didn't invent them, but we had to accommodate them

- A little like designing and building an airplane while in the air. New features, new models

Data Model Working Group

The Architecture Working Group

- Always the most contentious working group

2007

1998

Metaphors

- Explaining ourselves…

    to ourselves

    to the rest of the digital world

- A picture is worth a thousand words

- A good metaphor is worth whole chapters

Lego

An interoperability and extensibility metaphor

- Precisely engineered toys

- Designed to be future-proof

  - the old ones work with new ones because of the engineering

- The 'semantics' is extensible

- If you do it right, people will build semantics and structures you haven't even thought of
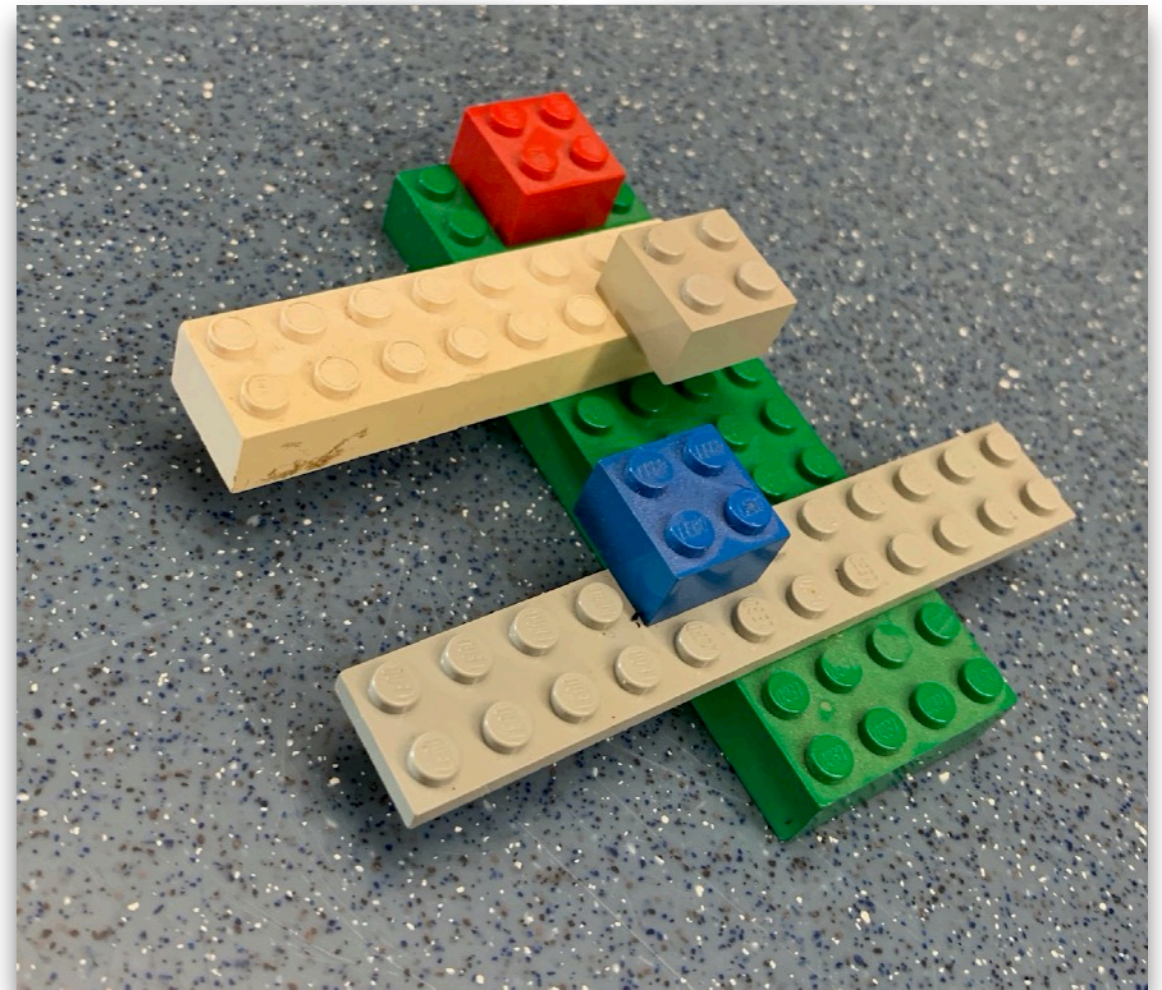
Ukraninan Nesting Dolls

A metaphor about the hierarchical structure of metadata (and the world)

- Information resources nest

- Metadata structures must accommodate hierarchy

Railroad Gages

An Interoperability metaphor: exchange across communities

- The border between China and Mongolia: railroad gage is different (by design)

- For communities to share metadata, there must be common infrastructure to support data interchange

Diagramming Sentences

A metaphor about metadata structure

Elements are repeatable, extensible, modifiable (qualifiers)

More a *model* than a metaphor, really

A Grammar of Dublin Core

Thomas Baker
GMD -- German National Research Center for Information Technology
Scientific Library and Publication Services
Schloss Birlinghoven
D-53754 Sankt Augustin, Germany
thomas.baker@gmd.de
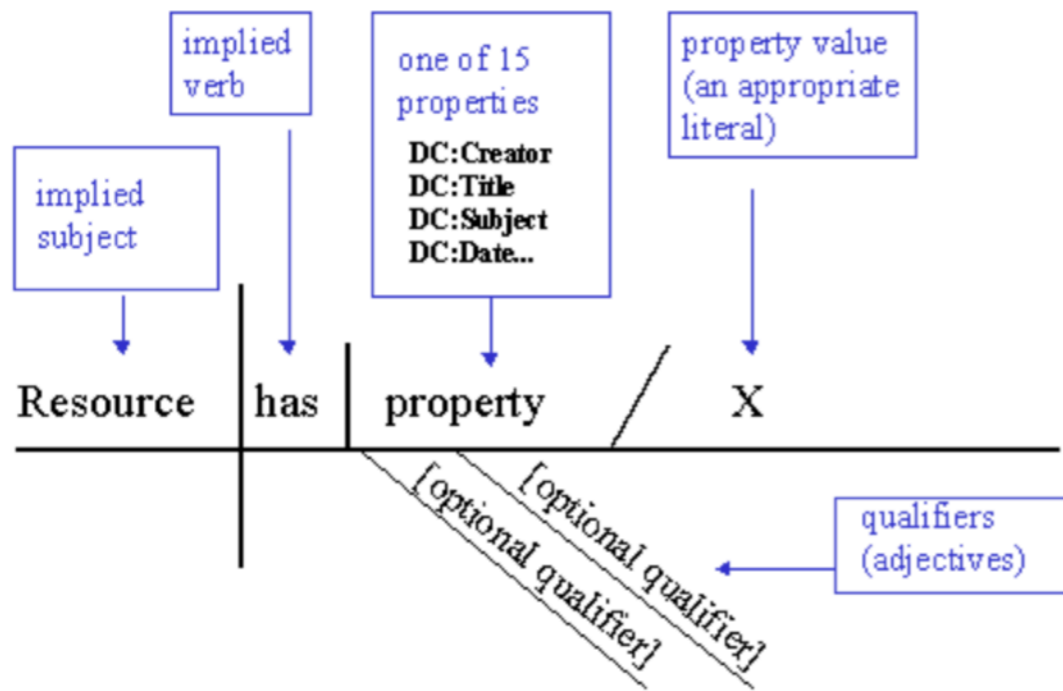
A way to talk about how qualifiers work
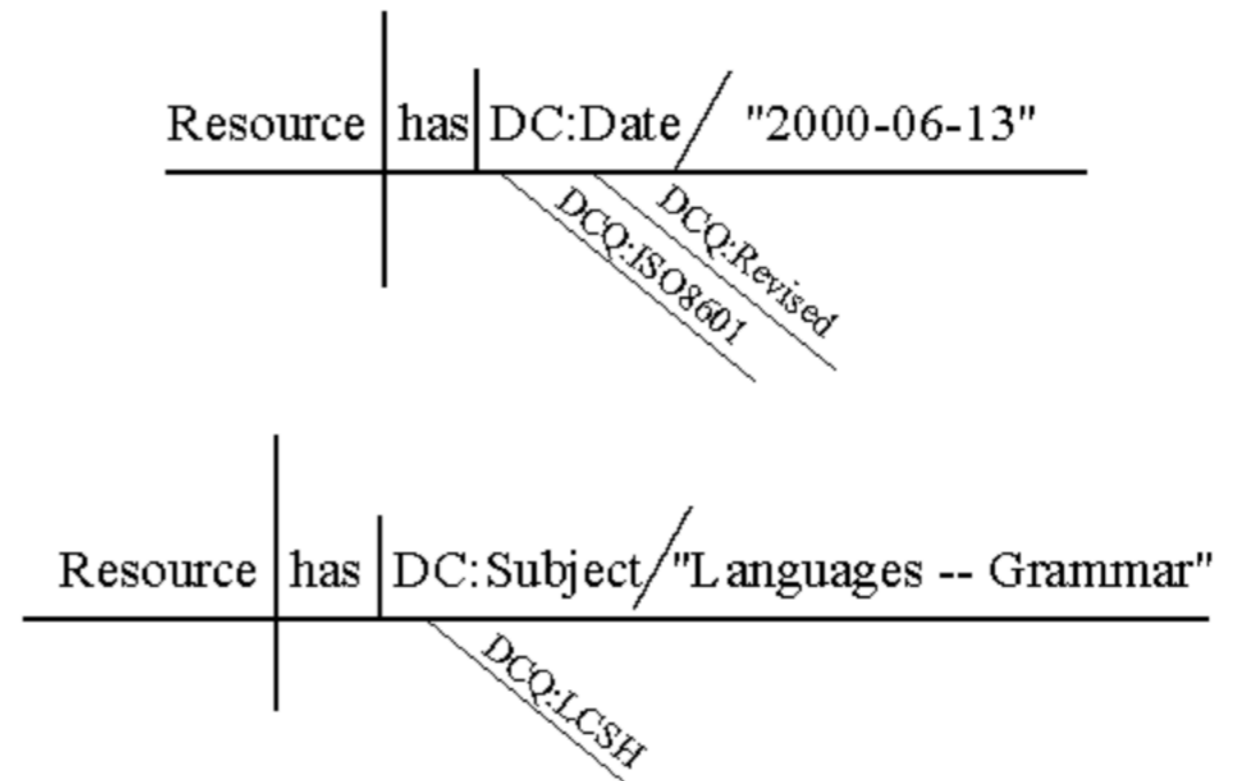

Figure 1: The general pattern

Figure 2: Some statements, diagrammed

A further virtue

Tom's Grammar of Dublin Core maps comfortably into the idiom of RDF
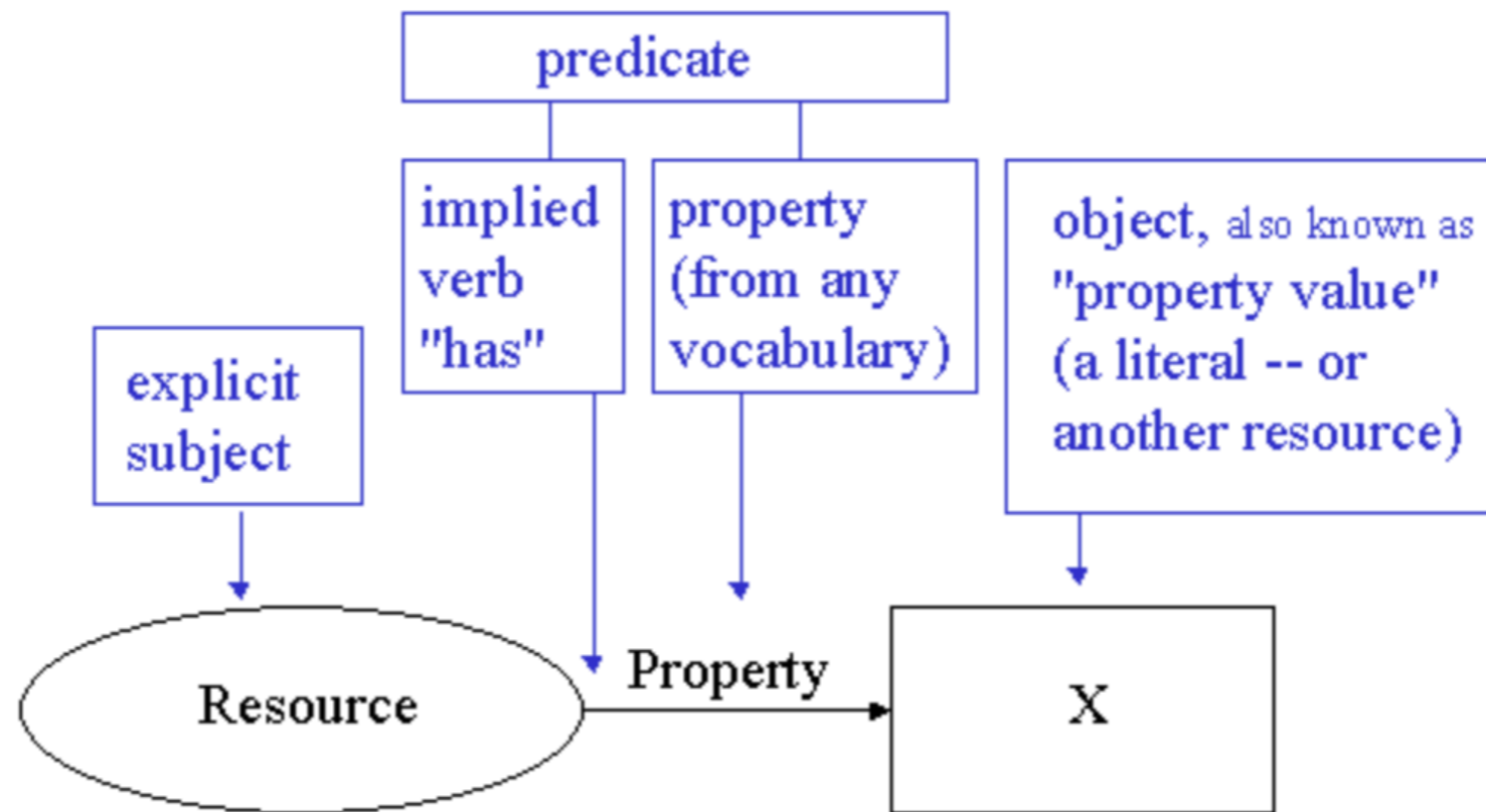

Figure 3: The general pattern of RDF statements ("triples")

One little glitch

- Every American child
learned how to diagram sentences

- Very few *other* places in the world do sentence
diagramming

- Still, the conceptual model of a grammar that is graphical
in nature lives comfortably within the "knowledge graph"
idiom that predominates on the Web

Pidgin languages and the path to multilinguality

- DCMI's contribution to multi-linguality on the Web has few peers

- Multi-linguality is both essential and an impediment

- Adopting DC metadata expressed in English can be likened to its use as metadata pidgin language

- Ricky Erway invoked the idea of a "Digital Tourist"
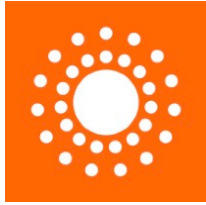
Assigning URIs to metadata terms

- DCMI pioneered the assignment of identifiers to metadata terms, conventions now in wide use on the Web

- Necessary for effective deployment of metadata terms across languages

- Also critical for machine-readable metadata, and in fact, machine readable everything

DCMI Metadata Terms Release History

A Trajectory of two Decades

- Release: 2020-01-20 (Current Version)
  - Definitions of properties and classes in the `/terms/` namespace follow ISO 15836-2:2019 as announced in January 2020, with minor differences of house style.
  - Usage comments for Language and Date in the `/elements/1.1/` namespace updated as per corresponding properties in the `/terms/` namespace
  - Two new properties added: `dcam:domainIncludes` and `dcam:rangeIncludes`.
  - `rdfs:range` changed to `dcam:rangeIncludes` for: `dct:accessRights`, `dct:accrualMethod`, `dct:accrualPeriodicity`, `dct:accrualPolicy`, `dct:audience`, `dct:conformsTo`, `dct:contributor`, `dct:coverage`, `dct:creator`, `dct:educationLevel`, `dct:extent`, `dct:format`, `dct:instructionalMethod`, `dct:language`, `dct:license`, `dct:mediator`, `dct:medium`, `dct:provenance`, `dct:publisher`, `dct:rights`, `dct:rightsHolder`, `dct:spatial`, and `dct:temporal`.
  - `rdfs:domain` changed to `dcam:domainIncludes` for: `dct:medium`.
  - Clarifications of wording for definitions and usage comments.
  - Additional usage examples.
  - Erratum 2020-03-11: Fixed reference URL for dcterms:ISO3166.
- Release: 2012-06-14
- Release: 2010-10-11
- Release: 2008-01-14
- Release: 2006-12-18
- Release: 2006-08-28
- Release: 2005-06-13
- Release: 2005-01-10
- Release: 2004-12-20 An excellent example of
- Release: 2004-09-20
- Release: 2004-06-14
- Release: 2003-11-19
- Release: 2003-03-04
- Release: 2003-02-12
- Release: 2002-10-06
-

An open, transparent standard
useful not only to the Dublin Core,
but to many other communities

It isn't *just* the idea (which is huge),
but also the *implementation* the
transparency, and the *maintenance*

What do we have for 25 years of DCMI?

Infrastructure

- A core 'semantic layer' in global use

    (the original elements): pidgin metadata

- Lego-block extensibility that supports enrichment of the core

- A widely-accepted model for assigning stable global identifiers to metadata terms (ours and other's)

- A hard-won understanding of metadata architecture that has spread widely in and outside of Dublin Core

Dan Brickley, Lead, schema.org

"DCMI's thinking and practice helped shape the formal standards for Linked Data, ontologies and the Semantic Web, as well as emerging approaches to data shapes and the efforts at schema.org."

DCMI Achievments (continued)

Social engineering

- A governance model based on global collaboration that supports **evolution** and **maintenance** of metadata

- Globally inclusive, robust **multilingual** infrastructure

- A global *community* of research and development

The community itself is the treasure

"The core schemas of Dublin Core remain at the heart of countless projects, systems and initiatives but should also be treasured as the practical talking point that brought us all together…"

Dan Brickley, Schema.org

Whither DCMI?

I'm a midwife, not an oracle

- Your guess is better than mine… Really

- A perusal of the talks in this virtual conference gives us a good feel for the short term

- The long term?

  - Its about the people

  - Its about the goals

  - Yes, it is increasingly about the business models

The Dublin Core Community

**In our isolation, we must not lose
the soul of collaboration**

- Communities are greater than the sum of their parts

- Communities have

  - shared purpose,

  - standards (both technical and social),

  - values

- Communities have momentum, they have trajectory, they have gravitational attraction

Look for these things among your peers

Nurture them

Fictive Myth

"We believe in a particular order not because it is objectively true, but because believing in it enables us to cooperate effectively and forge a better society."

Sapiens
Yuval Noah Harari

1998 Data Model Meeting (Crete)

Thank You!

Stuart.Weibel@gmail.com

- I have had no honor greater than working with The Dublin Core

- Thank you for your attention!

- Thanks also to Tom Baker and Dan Brickley for ideas about these slides