

Stuart L. Weibel

OCLC Online Computer Library Center, Inc.

A Personal History of the Dublin Core Metadata Initiative

Note: This is a pre-print version of a paper forthcoming in *Encyclopedia of Library and Information Science, Third Edition*. Please cite the published version; a suggested citation appears below. Correspondence about the article may be sent to the author at stuart.weibel@gmail.com.

© 2009 OCLC Online Computer Library, Inc.
6565 Kilgour Place, Dublin, Ohio 43017-3395 USA
<http://www.oclc.org/>

Reproduction of substantial portions of this publication must contain the OCLC copyright notice.

Suggested citation:

Weibel, Stuart L. 2009. "Dublin Core Metadata Initiative: A Personal History." In *Encyclopedia of Library and Information Science, Third Edition*, ed. Marcia J. Bates and Mary Niles Maack. Boca Raton, Fla.: CRC Press. Pre-print available online at: <http://www.oclc.org/research/publications/library/2009/weibel-elis.pdf>.

IN THE BEGINNING

In 1994, Internet search engines were virtually unknown, *resource description* meant library catalogers in technical services departments, and the idea of harvesting and indexing billions of pages of electronic resources was too fantastic for most people to contemplate. Information assets, if they were cataloged at all, were described using a collection of library content and format standards known as MARC cataloging, and even the description of traditional paper resources were typically backlogged in overburdened cataloging departments.

Web-scale harvesting and indexing has long since carried the burden of open online search, but many electronic resources (images, services, or applications, for example) have no suitable harvestable surrogates. Metadata – information about other information resources – is intended to fill this gap, providing explicit descriptions to support many administrative, commercial and management needs. Governments use metadata to connect legislative mandates and citizens, and target services to those who need them. Electronic commerce requires metadata to match search queries with products. Publishers need metadata to create product catalogs and manage rights. Corporations use metadata to organize and integrate documentation, business processes, reporting requirements, and products. Thus, descriptive metadata remains a key aspect of resource management, on the open Web as well in organizational intranets. As the importance of images grows in online media, so too does the metadata that describes such images. Normally-sighted people need image metadata to search for images. For the visually impaired, an image without a description is worse than no image at all... an unfulfilled promise of inaccessible content.

As the Web exploded on our desktops, it was evident that MARC cataloging of electronic resources would be too complex and costly, and might not be suitable for many electronic resources in any case. Might there be a simpler alternative for describing Web assets that does not require the experience of expert catalogers?

That was the question posed over coffee between sessions at the Second International World Wide Web meeting in Chicago in October of 1994. Joseph Hardin, Yuri Rubinski, Eric Miller, Stuart Weibel, and Terry Noreault were bemoaning the fact that it was difficult to find a given item among the nearly 500,000 addressable resources on the World Wide Web [1]. Never mind how many single sites of that size there are now... it seemed like a lot, and of course there were no search engines. Why not convene a group to look at the question systematically? Eric Miller and I spent the next few months organizing such a meeting, driven by the timeline of the next Web conference to be held the following April in Darmstadt, Germany.

THE OCLC-NCSA METADATA WORKSHOP

OCLC and NCSA (the National Center for Super Computer Applications at the University of Illinois) cosponsored this first meeting. We invited people from three general categories: content specialists, Internet technologists, and librarians (later we thought of these groups affectionately as the “*freaks, the geeks, and the people with sensible shoes*”). Fifty-two people attended that first two and a half day workshop in March of 1995. We started with the notion of *document-like object* (DLOs) on the Web – something familiar to give us a mental image without being too precise about what we might want to describe. This vagueness was comforting in our early, unbounded expectations, but sometimes confounded later efforts to bring description standards to bear in actual applications. We stopped talking much about DLOs and often struggled with the differences among web resources and how they should be characterized to enhance their discoverability.

What we hoped to do was to characterize conventions for resource description – to give people (authors, web site managers, content providers) a set of common descriptive metadata elements for describing electronic assets on the Web. We hoped that a relatively simple set of descriptors could be widely applied so the task of traversing silos of information would become easier on the Web than it has been in the physical world.

Dublin, Ohio, (OCLC’s headquarters, and location for the first workshop), gave us the name for the original set of 13 metadata elements – attributes of our document-like-objects that would help others locate and categorize them. We called this set of elements the *Dublin Core* or simply *DC* for short. Our objective, naïve in some respects, was to balance practicality and complexity, aiming for descriptions that untrained users might construct reliably, and which would make resources more findable. We created something that had some traction in the loose, fast-moving community of the early Web, something that attracted consensus from content specialists, web technologists, and librarians. Bottom-up consensus building was, and remains, an essential attribute of what evolved into the Dublin Core Metadata Initiative (DCMI).

The name of the element set, in addition to reflecting its geographic patrimony, captures the notion that DC is a starting place (a core). It has always been an expectation that DC would be a kernel around which other metadata would grow. That first meeting generated basic principles that continue to guide DC development: Elements would be:

- Intrinsic (describe one thing)
- Extensible (modular, easily expanded)
- Syntax Independent (keep syntax and semantics separate)
- Optional (use only the elements you need)
- Repeatable (as many instances of a given attribute as appropriate)
- Modifiable (element *refinements* could be used to sharpen descriptions)

We wrote up the report [2], and published a synopsis in the first issue of *DLib Magazine* [3]. Eric and I felt like we’d launched something of interest, but neither of us knew what to do next. We hoped a path would reveal itself, and in fact it did, in the person of

Lorcan Dempsey, then leader of the United Kingdom Office of Library Networking (UKOLN) in Bath, UK.

THE WARWICK METADATA WORKSHOP [4]

Lorcan suggested we might extend the reach of the original effort through a follow-on event in the UK. Thus did our one-off effort acquire a trajectory. The University of Warwick had a perfect conference venue for the sort of meeting we envisioned: good meeting rooms, all accommodations onsite, and a great bar that stayed open late.

Each of the early workshops had a tight-wire feel about it, a sense that what we were doing was fragile, and to emerge from a workshop without a major step forward would be fatal. The major result of the Warwick meeting was what came to be known as the *Warwick Framework* [5], a conceptual architecture for metadata that recognized the requirement for modular, extensible metadata. The ideas articulated in the Warwick Framework helped us decide what should be part of the Dublin Core and what might be borrowed from metadata element sets designed by others. The infrastructure scarcely existed to support deployment in those days, but the idea survived the intervening years of change, slowly morphing into what we today refer to as *application profiles* – hybrid schemas comprised of elements drawn from some number of metadata sets [6]. These ideas hold up well in the current world of metadata, as long as the element sets can be declared within a common data model, a proviso that has been a sticking point for interoperability from the very beginning.

DC-3: OCLC-CNI WORKSHOP ON IMAGE DESCRIPTION [7]

At this point in 1996, we still didn't have a strong sense of where we were going. In fact the third 'DC workshop' wasn't chartered as such at all. It was co-sponsored by the Coalition for Networked Information (CNI), one of the last major efforts of Paul Evan Peters prior to his untimely death. Eric Miller did most of the organizational work, building on our experiences with the first two DC workshops. We didn't announce it as part of the series, but the subtext of the meeting was basically: Can a simple DC element set be used effectively to describe images? The consensus of the group was yes, though there was a strong sense that the 13 elements needed to be augmented with two additions: *Rights* and *Description*. We declared, somewhat belatedly, that this workshop was the third in the DC series.

The group felt fairly strongly that without the means to determine intellectual property rights (IPR) associated with an image, just finding it was inadequate. History tells us otherwise. The use of the *Rights* element remains marginal at best, largely because it is a difficult problem, with many infrastructural and legal dimensions that are not easily dealt with in a simple declarative metadata element. Nonetheless, it at least has provided for simple declarations about IPR, and a convenient hook for future deployment.

The *Description* element was brought into being because conferees felt that the *Subject* element did not encompass the notion of an image description sufficiently well, and that a

separate element was necessary. One may argue such distinctions forever – I would probably resist this element more strongly today. It is easy enough to condense these two elements as ‘aboutness’. But DC was then, and continues, to be a consensus activity: a roughly-right solution that attracts usage is to be preferred over perfection that never arrives.

At this point it was clear we had a workshop series on our hands, and the acronym DCMI – Dublin Core Metadata Initiative – came into use. One memorable moment, which reinforced this sense of impact, came at a metadata interoperability workshop held near Bonn, Germany, late in 1996 [8]. I attended this small workshop to learn that a group in Osnabrück, Germany had launched a digital library effort that harvested metadata for mathematical preprints. Judith Plümer reported on work she and Roland Schwänzl had led using Dublin Core metadata. It was the first time I encountered people who had not been active in the DC effort, but had simply recognized it as something of use to them and deployed it. Roland later became an enthusiastic DCMI contributor until his death in 2004.

DC-4: SURFACING DOWN UNDER [9]

We certainly thought of ourselves as working on a problem of international scope, and we wanted the venues of the workshop series to reflect that. Australia was actively shaping a government metadata standard based on Dublin Core. Renato Iannella, then of the Distributed Systems Technology Centre in Brisbane, Australia and Warwick Cathro, of the National Library of Australia in Canberra, had attended the Warwick Workshop (DC-2, if you will), and were enthusiastic about broadening the geographic scope of the effort. The National Library of Australia offered to host the meeting. We wondered how we could manage to get people (especially public sector people, who rarely benefit from international travel support) to attend. Australia is a long way from both Europe and the Americas. But it is the *World-Wide Web*, not the US Web or the European Web, and we felt it was important to reflect that global perspective in our efforts. This approach remains central to choosing locations for DC conferences, rotating venues among Europe, the Americas, and Asia.

We enlisted the support of the National Science Foundation, attracting small amounts of funding that we shared with applicants to help make their trips look (financially) more like domestic travel. No one got a free ride, but rather these funds helped to keep travel costs manageable. *DC-Down Under* helped consolidate the work that we’d done to date, and resulted, in part, in the first adoption of DC as a national information standard. This precedent continues today – adoption and localization. Australia was then, and continues to be, an important locale for the advancement of information description standards, and their hosting of DC-4 was a convincing testament to the international character of the effort.

Discussions at DC-4 raised the specter of *creeping elegance* within the community. We all want simple... but exactly the right degree of simple. The functional requirements differ from application to application and domain to domain. For the first time we tried

to instantiate notions articulated in the Warwick Framework – modular and extensible metadata. The so-called *Canberra qualifiers* emerged -- a first attempt to provide the extensibility that developers wanted to meet the local requirements in their applications. One may cavil about complexity and simplicity, and where the right balance lies, but that balance point is found in applications rather than theory.

Thus were element refinements born. We agreed that the basic fifteen *buckets* could be refined in order to sharpen the description. A date might be a date of publication, a date of copyright, a date of availability, or something else. It is permissible to refine or qualify the basic date field so as to make explicit the character of the instance.

But what if people use different refinements? The *dumb-down principle* emerged as (an unfortunately named) rule of thumb, decreeing that applications that lacked knowledge of a given refinement could simply ignore them. Some information may be lost, but the user will be no worse off than having had unrefined elements in the first place.

The *one to one principle* emerged in Canberra as well: in Erik Jul's paraphrase of Ranganathan: *to each resource, its [own] description*. Why is this not obvious? Once again, the tension between alluring simplicity and tempting complexity arises. In the ideal world, all descriptions should be as discrete as their referents. *Stuart Weibel* is the author of this history, but his email address has nothing more than an accidental relationship to it. Still, I might want to have it in the document. Is it metadata? About what? Certainly not the document, though readers may find it handy to have it. It is metadata (contact information) pertaining to a *person*, not an *information asset*. *One to one*, then, is a many-headed-snake, and it has bitten us often over the years.

There was one other advance at DC-4: the workshop T-shirt. Lynn Marko, of the University of Michigan, arrived with her husband, Ken, then of Ford Motor Company. Unfortunately, Ken's luggage didn't make it. We gave him a couple of T-shirts to tide him over. Ken repaid the kindness many times, sponsoring several future workshop T-shirts.

DATA MODELS: WHENCE THE DEVIL RESIDES

While it is useful to recount DCMI history in the rhythm of its annual workshops, a good deal of activity took place outside as well. Workshops begat working groups, tasked with specific objectives, and laden with sunset provisions (working groups were chartered with specific objectives and expired at the following conference). Among the most difficult of objectives was the evolution of a data model for the Dublin Core. No experience of my professional life has been as contentious as these meetings. They were so singularly disruptive that we changed the names to protect the guilty and innocent alike. Initially we had a Data Model group, and later an Architecture group, and the name of the outcome evolved from data model to abstract model. (More on this topic follows).

The early DC workshops were about broad consensus, and they were fiery experiences. Those early ‘bun fights’ revealed a pattern of rising expectations, followed by a rapid slide into a trough of contentious argumentation before we ‘bounced’ to find a plateau of consensus that moved us forward. In Canberra we had to push beyond the surface of simple descriptive metadata to the rather more complicated underpinnings necessary to achieve interoperable metadata systems. This realization paralleled similar recognition elsewhere in the Web. Protocols were evolving, new standards, both formal and informal, were emerging. DC at this stage was far from a standard, but certainly there was recognition that it needed to be if it were to take root.

Somewhere in this interval, Bill Arms, then of the Coalition for National Research Initiatives (CNRI), arranged a meeting that would help jumpstart the formulation of the Resource Description Framework (RDF) working group at the World Wide Web Consortium (W3C). Political pressure to support filtering of content so as to avoid the inadvertent exposure of minors to pornography had led to something called PICS (Platform for Internet Content Selection). PICS was a bit of *ad hocery* born of expedience without great forethought, and did little but provide a sense that someone was at least working on the problem. Recognizing that PICS was a rather-too-specialized solution, W3C folks had it in mind to generalize it as a platform for formalized semantics – metadata. Tim Berners-Lee is credited with originating the notion of the Semantic Web, and RDF would become a key enabling technology to support it.

Jim Miller, then of the W3C, Dan Connolly, also and still of the W3C, Bill Arms, and myself met to discuss the idea of an open community working group to address the issue. I was there largely as a prospective consumer of the results (that is, as leader of a community that could both use such a technology and provide practical functional requirements). Thus began DCMI’s tightly coupled relationship to the beginnings of the Semantic Web. Key DCMI participants have been intimately involved in semantic web activities in the W3C and the wider community since that time. Indeed, my colleague at OCLC, Eric Miller, would eventually move to the W3C and become the Semantic Web Lead.

THE UNFINISHED CACOPHONY: DC-5 -- HELSINKI AND THE FINNISH FINISH [10]

Early DC workshops were hosted by the players: the organizations and people with a stake in the effort. Juha Hakala of the National Library of Finland stepped to the line for DC-5, hosting the meeting in Helsinki. We stayed in a wonderful hotel on the waterfront, within easy striking distance of the herring boats and an urban architecture evocative of Tolstoy.

RDF was a hot topic in Helsinki, held high by some participants as the answer to the infrastructure impediments holding back metadata applications. Others grumbled, eyes narrowed with skepticism, concerned that RDF was immature and over-hyped. But everyone was eager to find solutions, and knots of conferees could be found huddled over laptops in the lobby at 2 am or with laptops open on restaurant tables sorting through

problems of syntax and semantics. The stamina, good will, and collective work done at these workshops remain for me the epitome of professional dedication, as well as foundations of lifelong friendships.

As for elements and refinements, our slogan was the *Finnish finish*... the equivalent of the *Mission Accomplished* banner on the aircraft carrier of metadata. While it is true that no elements have been added since DC-2, and a number of element qualifiers were in use, there were yet plenty of arguments to be heard.

Among the DC geezers, people still talk of the tar ice cream served at the conference dinner in Helsinki... a crystalline, ice-white concoction that, paradoxically, evoked tar and freshness at once. The flavor of pine tar and peppered strawberries defines my memory of this great northern city. The T-shirt was great, too... we had a Dublin Core World Tour T-shirt with the element set in three languages on one side, and the stops and dates of the global tour on the other. The listing embodied at least three variant syntaxes for dates, and some of us imagined mischievous intent in Juha's grin. T-shirts, and now, too, a song. The workshop came to a close with a group sing-along, a characterization of workshop achievements superior in brevity and eloquence to the workshop report. I confess having been taken somewhat aback at this development. At a followup meeting soon after, I asked Roland Schwänzl, our German Mathematician-cum-metadata wonk... "Is this sort of thing common in Europe... a song of summary after a workshop?" Roland seemed surprised at my question... "Ja Ja!"

GROWING PAINS AND CONSOLIDATION: DC-6 (DC:DC) [11]

Our premature declaration of success from Helsinki was not long un-tattered. The DC Tour traveled next to the Library of Congress in Washington, DC. The meeting included representatives of the INDECS project (a metadata activity focused on rights declarations), who felt we'd gotten a few things wrong. Of course, we had. The tension between INDECS and DC is classically a top-down/bottom-up conflict. DC evolved by consensus and the judgment of practitioners, INDECS by deductive inference. Each model has strengths, and each has weaknesses. To quote Marshall Rose:

The distance between theory and practice is always smaller in theory, than in practice

We had hopes of finding the means to bring the two efforts closer together, and I think both activities would have benefited. DC and INDECS supporters collaborated on a paper in an effort to promote this convergence [12], but in the long run our differences were too great. One proposal motivated in part by this hope of consolidation, and in part by an ongoing recognition that simplification of the set would be a benefit in the long run, became known as the '*secret agent proposal*'. The notion of consolidating Publisher, Contributor, and Creator into a single element (*Agent*, or *Names* perhaps) was floated quietly among some DC participants. The idea never got to the point of serious open consideration, as there was strong opposition based on breaking legacy applications. This was perhaps the first time we were aware that DC *was* a legacy system. We weren't

engaging in an intellectual exercise in resource description: people cared, and systems depended on stability. The outcome of this attempt at reconciliation was almost certainly impaired as well by a sense of hostile intent between the two camps. This was, I believe, an unfortunate interpretation of the very natural inclination to cling to one's own ideas. The clarity of hindsight leaves me with regret that we did not accomplish this simplification of model and application.

In this same vein, the RDF proponents lobbied hard to make this syntax the preferred encoding of DC metadata instances. There was strong pushback by others, who judged that simpler encodings (embedded HTML meta-tags for example) were being deprecated in favor of what they judged to be an over-elaborated and unproven technology that mere mortals found hard to grasp. The practitioners reined in the wild-eyed theoreticians with a "Now just hold on!" Those of us who believed in the value of RDF (I count myself among them) had to acknowledge that we weren't marching inexorably toward a single, coherent deployment strategy, but that various paths would persist, progressing at their own speeds, indefinitely. DCMI needed to support each of these paths. It still does.

Hindsight may be 20/20 vision, but I am still blind on this one. On the one hand, a single deployment path would have greatly simplified the onerous task of syntax support, and consolidated more effort around reference code. This would have been helpful (still would be, for all that). On the other hand, the Web teaches us that technology deployment in the early 21st century is riverine and decentralized. Attempts to channel and coerce are rarely successful. As I write this, nearly a decade later, RDF struggles on, seemingly gaining momentum, but the tools and techniques we expected to catch fire have mostly smoldered. Timing is everything, and we may have been ahead of the curve, or perhaps off in the weeds. The last chapters are yet unwritten.

The conference dinner for DC:DC was held in what was once the library for the Smithsonian Institution – a large and lovely hall in the Smithsonian's 'castle', its first building. On the way into the dinner, there were Buddhist monks engaged in the composition of a sand Mandala. These meticulously wrought sand paintings take many hours on the part of several monks, and at the end, the visual esthetic, created as an act of prayer, is swept away. Standards creation is a bit like this, a kind of hopefulness that is inevitably short lived, but may nonetheless result in an abstraction of intent that reverberates in many ways through its practitioners and community.

That dinner was a great one, and fomented another workshop sing-along. I do not know all the details of its genesis, but I am given to understand that Diann Rusch-Feja and Godfrey Rust of the INDECS project were co-conspirators. The lyrics, sung to the tune of *Both Sides Now*, were, as in Helsinki, a far more succinct summary of the workshop than any reports could have captured. I hope someone still has the lyrics.

THE BIG VOTE

The instability of consensus regarding qualifiers led to a formal voting procedure in 2000 that marked an inflection point for DCMI, creating for the first time a structured (if yet

messy) procedure for decision-making. We subsequently borrowed governance methods where we could find them – the Internet Engineering Task Force (IETF), MARBI (the community that looks after the MARC standards), the W3C, and others. From these, we crafted procedures to make evident a commitment to open, public processes to manage the element sets in our care.

The DC Advisory Committee, made up of working group chairs and invited experts, agreed to vote on proposed qualifiers. The results of this vote comprised the first formal approval of terms beyond the basic 15, and what we learned in the process informed the next stage of formal procedures, which remain largely intact today. This vote, contentious though it was in certain respects, marks the maturing of the Dublin Core from a group of metadata enthusiasts to a formal standards community. The character of the meetings changed at about this time as well, evolving away from contentious passions to a more business-like collegiality.

DC-7 FRANKFURT

Die Deutsche Bibliothek hosted the next workshop in their splendid facility in Frankfurt, Germany in 1999. The tenor and scope of the workshops evolved towards a stronger educational and informational tone, leaving behind us the sense of *we'll-die-if-we-fail* and strengthening the recognition that DC had a strong foothold that wasn't going to evaporate. There were, of course, myriad details of contention and ambiguity that required attention. But the approach was workman-like and deliberate, with a confidence born of more stable ground beneath our feet.

DC-8, OTTAWA, AND LOM [13]

The following year we returned to North America for DC-8 at the National Library of Canada in Ottawa. One of the important developments that emerged at this meeting was a closer alignment between the Learning Object Metadata (LOM) community and the DC community. LOM embraces both descriptive metadata for resources, based on Dublin Core semantics, as well as a variety of other classes of metadata intended to support management and use of educational resources.

Unfortunately, the two initiatives evolved in different directions structurally, a problem that plagues the technical implementation of metadata to this day. In Ottawa, DCMI and LOM agreed on a Memorandum of Understanding that has helped to guide a slow process of convergence. One of the deliverables of that agreement remains an important articulation of common principles that should be useful to many communities: *Metadata Principles and Practicalities*, co-authored by principals of both initiatives [14].

The enduring lesson of the long history between DCMI and LOM is that without a common underlying data model, sharing metadata becomes difficult, imprecise, and (perhaps most crippling) expensive. I've drawn an analogy elsewhere to the incompatibilities between railway gages [15]. Different underlying 'rail gages' force the unpacking and reloading of cargo, whether bales of cotton or metadata. The lack of a

common underlying metadata model remains, in my estimation, a major impediment to broad cross-disciplinary metadata interoperability.

DCMI COMES OF AGE

In 2001, the DCMI community transitioned from workshops to conferences. In addition to the working group meetings, we offered a tutorial track and a refereed conference paper track. The National Institute for Information (NII) in Tokyo hosted the meeting in their beautiful facility a short walk from the Imperial Palace. We marked this transition in the names of the conferences, using years to distinguish them. The transition from a standards creation activity to a standards maintenance activity was taking place as well, and much more of the effort involved sharing of practice, through tutorials, plenary lectures, conference papers, and poster sessions. The locations of the conferences, listed in Table 1, are testament to the ongoing commitment of DCMI to its international foundations.

Table 1: The Dublin Core Conference Series

DC-2001	Tokyo, Japan
DC-2002	Florence, Italy
DC-2003	Seattle, Washington (USA)
DC-2004	Shanghai, China
DC-2005	Madrid, Spain
DC-2006	Colima, Mexico
DC-2007	Singapore
DC-2008	Berlin, Germany

Behind the public face of the Initiative, much work was still underway to strengthen the procedures necessary to support an evolving standard. We reorganized the structure and procedures of the initiative, to include:

- A Board of Trustees: distinguished representatives of sectors and regions with experience and commitment to the advancement of the Dublin Core, and metadata in general.
- The DCMI Directorate: administrative staff to manage the activities of the Initiative.
- A Usage Board: recognized experts in the field who would review and guide proposed additions or changes in the standard.
- An Affiliate Program: Institutional sponsors for DCMI in various countries around the world to promote and support DCMI activities
- An Advisory Board: (successor to the original Advisory Committee) Working group chairs that have been involved in the evolution of DCMI and other invited experts who provide consultation and advice.

Building an organizational structure for the initiative also exposed issues of ownership and sponsorship. OCLC generously paid most of DCMI's bills for more than the first decade of its evolution, but certainly never owned or directed its development. OCLC's role as host was a side effect of the organizational involvement of myself and Eric Miller

and the willingness of OCLC management to support our efforts. Subsequently, the formal link between OCLC and DCMI has been (amicably) severed, DCMI having been established as a limited liability company under the auspices of the National Library of Singapore; DCMI's formal organizational structure now reflects what was always operationally true – a genuine operational independence befitting a community-based initiative.

As the need for managerial capacity evolved, Makx Dekkers and Tom Baker were enlisted as members of the DCMI Directorate. Tom had been an active contributor to DC since the second workshop, and has authored many papers and DCMI documents that are milestones of DC development. His knowledge of languages and linguistics have been particularly valuable. I have been privileged to have shared with him many wonderful brainstorming sessions about metadata. Two particularly memorable experiences – developing notions of metadata grammar atop a sunlit tower in Lucca, overlooking the Tuscan hills, and a middle of the night Japanese *onsen* (hot springs bath) -- rank among the most pleasurable of work experiences. Makx, also a long time DC participant, brought to the initiative strong project managerial skills and experiences in the European standards world.

My own role in the Dublin Core began to decline during this period. Start-up and sustenance require different skill sets. Separating from guidance and management of DCMI was both difficult and right (for the Initiative and for myself). But surely surviving its founders is an important measure of success for any organization!

STANDARDIZATION STRATEGIES

For most countries and many large organizations, information management practices must be built around formal standards. Our approach was to transition DC from simpler to more formal efforts. We started with an IETF RFC (request for comments). The IETF route had two advantages: many of our anticipated adopters came from the Internet culture, and establishing an RFC was relatively simple. An informational RFC is little more than a public declaration of an approach, useful for informing others of what has been done and why it is the right thing to do. It requires a minimum of external review or persuasion. Publication of RFC 2413 in 1998 [16] constituted the initial stake in the ground for standardization of DC metadata (this RFC was updated as RFC 5013 in 2007 [17]).

Subsequently DCMI embarked upon standardization in the National Information Standards Organization (NISO) in the US. NISO approval was seen as a step to international standardization in ISO, as NISO is a US organizational member of ISO. NISO standardization proved challenging, both because DCMI encountered political opposition in various quarters, and because we were the initial use case for a 'fast-track' approval process intended to reduce the latency of standardization activities in a rapidly changing Web environment. Pat Harris, then director of NISO, skillfully negotiated the minefields, and helped DCMI achieve this important imprimatur, and DC metadata was

standardized as ANSI/NISO Z39.85 in 2001. Once DC metadata became a NISO standard, ratification as ISO 15836 followed relatively easily.

THE DCMI ABSTRACT MODEL AND INTEROPERABILITY

As alluded to earlier, data modeling has from the start been problematic for DCMI. In pre-web days, it was less important to have a formal model for metadata. MARC standards, for example, have never benefited from an underlying data model, and yet have enjoyed wide adoption and usefulness. As infrastructural boundaries between domains have dissolved in the universal solvent of Web protocols, the need for agreements about syntax and structure increased. While humans readily distinguish nuances of meaning and structure, machines do not.

The formal emergence of a data model for DC took a decade. The final breakthrough was spearheaded by Andy Powell and Mikael Nilsson, following on Tom Baker's exposition of DC grammar [18] (and a decade of vociferous argumentation from many quarters). The DCMI Abstract Model (DCAM) [19] abstracts the basic metadata entities used in DC metadata and specifies their expression in RDF.

Sharing metadata requires agreements on a variety of levels. Some of the salient issues:

- What is the meaning conveyed in metadata assertions? Meaning, of course, resides in the minds of people, not machines. The focus of the Dublin Core was initially on promoting common meanings and make them sharable on the Web.
- How are metadata assertions packed so that one machine can send them to another without changing their meaning?
- How can metadata be structured so as to assure consistent and unambiguous interpretation? Reliable syntax encoding requires consistent data structures. The sorts of things that must be specified in a well-structured metadata assertion include:
 - The boundaries of a set of assertions (what constitutes a record)
 - Cardinality: can an element be repeated, and if so, is there a limit on the number of repetitions?
 - How is a name structured? What is the delimiter separating elements of a compound name (*Prince* and *Bono* excepted, most names are compound structures, many with surprising and confounding complexity).
 - At what granularity are descriptions associated with resources? How is nesting managed for hierarchical resources?
 - How are dates encoded? YYYY-MM-DD? DD-MM-YYYY? MM-DD-YYYY?
 - How does one identify an encoding scheme that specifies the above question?
 - How are character encodings declared?

- How does one identify a value-encoding scheme (eg. LCSH, MeSH, Dewey...) from which metadata values can be chosen? Are such schemes required or optional?
- Are metadata values specified by reference (URI) or by value (literal strings)?

It is in these details that the success or failure of interoperability is written; understanding and declaration of these details remains incomplete.

Current thinking among DCMI metadata architects suggests it is useful to distinguish 4 levels of interoperability, each characterized by a set of choices, costs, and benefits [20]:

Level 1: Informal Interoperability: A vocabulary of concepts expressed in natural languages that support sharing of meaning among people. Metadata terms can be described as informally interoperable with Dublin Core if their natural language representations can be mapped to the 15 legacy DC elements. There is no requirement for DCAM conformance, and values for metadata are expressed as human-readable strings. Prospects for machine-to-machine interoperability at this level are low.

Level 2: Semantic Interoperability: Metadata that can be accurately and completely expressed using RDF formalisms, but which is not conformant with additional structures elaborated in the DCAM.

Level 3: Syntactic interoperability based on the DCMI Abstract Model: RDF-compatible metadata that conforms to all elements of the DCAM (in particular, *bounded descriptions* – a collection of metadata assertions bound together and identified as what is thought of as a *record* in traditional bibliographic practice).

Level 4: Singapore Framework Interoperability: conformance to a complete specification of functional requirements, a domain model, and description set profile [21].

Levels 1 & 2 roughly define the divide between human-sharable and machine-processable metadata strategies of the first decade of DC's existence. Level 3 promises a higher level of interoperability through conformance to abstract metadata structures that are defined in the DCAM (in particular, bounded records which can be identified using URIs). The benefits of these additional structures accrue to managing and validating metadata collections.

If Level 4 seems vague, it is perhaps a reflection of yet-poorly understood requirements for bringing together large aggregations of heterogeneously managed metadata collections. This remains an area of active exploration.

At this writing, metadata modeling experts do not agree on the finality of the DCAM, or its general usefulness for all metadata. It is currently the closest approximation of a generalized data model for metadata in a Web environment, and is one of the most

important contributions of the Dublin Core to the larger metadata community, the result of many years of grappling with a difficult problem.

IN SUMMARY

The protocols and standards that underpin our digital lives are, in the long term, as ephemeral as the sand mandelas of the monks. The understanding that emerges from their creation, however, is part of a more persistent trajectory. The Dublin Core marks its 15th year in 2009, arising and co-evolving with the Web itself. The conference series remains the preeminent venue for defining and deploying structured description on the Web. Its practitioners are found in 50 or more countries, and its descriptions are quietly at work in applications that serve literally billions of people over a range of domains that includes the latest media formats, government publications, scientific data, Internet syndication feeds, image collections, and many more.

The architecture and details of DCMI metadata have evolved over its lifetime. Its practitioners have labored together in meetings in 30 countries or more. Notable venues for DC events include the rolling hills of Kataminga, outside Melbourne, Australia (the first DC-Education meeting), the magnificent National Museum of Art in Canada (DC-8 conference dinner, Ottawa), the splendid conference hall in the National Institute for Information in Tokyo (DC-2001), the Shanghai Library (DC-2004) and perhaps most spectacular of all, the Hall of the 500 in the Palazzo Vecchio in Florence (DC-2002 conference dinner). The privilege of working and socializing in such places (and so many others, too numerous to mention), have left rich imprints on participants and the work alike.

The principles and practices elaborated in DCMI reverberate throughout the metadata community and provide a useful framework for metadata now and into the future. That this is so is a credit to the many people who have contributed their time, their expertise, and their passion to its evolution. These efforts are an important and lasting contribution to the digital future.

NOTES & BIBLIOGRAPHY

[1] Joseph Hardin was leader of the software development team at the National Center for Super Computer Applications (NCSA) that developed Mosaic, the first cross-platform Web browser. The late Yuri Rubinski founded SoftQuad, and is credited with popularizing SGML, the precursor to XML. Terry Noreault was then director of the OCLC Office of Research, and Eric and I were colleagues in the Office of Research.

[2] OCLC/NCSA Metadata Workshop Report.
Stuart Weibel, Jean Godby, Eric Miller and Ron Daniel. 1995.
<http://dublincore.org/workshops/dc1/report.shtml>

[3] Metadata: The Foundations of Resource Description.
Stuart Weibel. *D-Lib Magazine*, July 1995.
<http://www.dlib.org/dlib/July95/07weibel.html>

[4] The Warwick Metadata Workshop:
A Framework for the Deployment of Resource Description
Lorcan Dempsey and Stuart L. Weibel. *D-Lib Magazine*, July/August 1996.
ISSN 1082-9873.
<http://www.dlib.org/dlib/july96/07weibel.html>

[5] The Warwick Framework: A Container Architecture for Diverse Sets of Metadata.
Carl Lagoze. *D-Lib Magazine*, July/August 1996. ISSN 1082-9873
<http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>

[6] Application profiles: mixing and matching metadata schemas.
Rachel Heery and Manjula Patel. *Ariadne*. Issue 25, September 2000.
<http://www.ariadne.ac.uk/issue25/app-profiles/>

[7] Image Description on the Internet: A Summary of the CNI/OCLC Image Metadata
Workshop
Stuart Weibel and Eric Miller. *D-Lib Magazine*, January 1997. ISSN 1082-9873.
<http://www.dlib.org/dlib/january97/oclc/01weibel.html>

[8] The second DELOS workshop: Metadata and Interoperability
GMD, Bonn, Germany, 7-8 October 1996
<http://www.iei.pi.cnr.it/ErcimDL/second-DELOS-workshop/hall.html>

[9] The 4th Dublin Core Metadata Workshop Report.
Stuart Weibel, Renato Iannella, and Warwick Cathro. *D-Lib Magazine*, June 1997. ISSN
1082-9873.
<http://www.dlib.org/dlib/june97/metadata/06weibel.html>

[10] DC-5: The Helsinki Metadata Workshop: A Report on the Workshop and
Subsequent Developments.
Stuart Weibel and Juha Hakala. *D-Lib Magazine*, February 1998. ISSN 1082-9873.
<http://www.dlib.org/dlib/february98/02weibel.html>

[11] The State of the Dublin Core Metadata Initiative.
Stuart Weibel. *D-Lib Magazine*, April 1999, Volume 5 Number 4, ISSN 1082-9873.
<http://www.dlib.org/dlib/april99/04weibel.html>

[12] A Common Model to Support Interoperable Metadata: Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities.

David Bearman, Eric Miller, Godfrey Rust, Jennifer Trant, and Stuart Weibel. *D-Lib Magazine* January 1999, Volume 5 Number 1, ISSN 1082-9873.

<http://www.dlib.org/dlib/january99/bearman/01bearman.html>

[13] The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions.

Stuart L. Weibel and Traugott Koch. *D-Lib Magazine*, December 2000, Volume 6 Number 12, ISSN 1082-9873.

<http://www.dlib.org/dlib/december00/weibel/12weibel.html>

[14] Metadata Principles and Practicalities.

Erik Duval, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel. *D-Lib Magazine*, April 2002, Volume 8 Number 4. ISSN 1082-9873.

<http://www.dlib.org/dlib/april02/weibel/04weibel.html>

[15] Border Crossings: Reflections on a Decade of Metadata Consensus Building.

Stuart L. Weibel. *D-Lib Magazine*, July/August 2005, Volume 11 Number 7/8. ISSN 1082-9873.

<http://www.dlib.org/dlib/july05/weibel/07weibel.htm>

[16] Dublin Core Metadata for Resource Discovery.

S. Weibel, J. Kunze, C. Lagoze, and M. Wolf. IETF Informational Request for Comments: 2413. September 1998

<http://www.ietf.org/rfc/rfc2413.txt>

[17] The Dublin Core Metadata Element Set.

J. Kunze and T. Baker. IETF Informational Request for Comments: 5013. August 2007

<http://www.ietf.org/rfc/rfc5013.txt>

[18] A Grammar of Dublin Core.

Thomas Baker. *D-Lib Magazine*, October 2000, Volume 6 Number 10, ISSN 1082-9873.

<http://www.dlib.org/dlib/october00/baker/10baker.html>

[19] DCMI Abstract Model.

Andy Powell, Mikael Nilsson, Ambjörn Naeve, Pete Johnston, Thomas Baker. 2007-06-04 (DCMI Web site).

<http://dublincore.org/documents/2007/06/04/abstract-model/>

[20] DCMI Architecture Wiki: Interoperability Levels.

Mikael Nilsson, Thomas Baker, Pete Johnston. DCMI Architecture Working Group Wiki.

<http://dublincore.org/architecturewiki/InteroperabilityLevels>

[21] The Singapore Framework for Dublin Core Application Profiles.
Mikael Nilsson, Thomas Baker, and Pete Johnston.
2008-01-14, DCMI Web Site.
<http://dublincore.org/documents/singapore-framework/>