The Making of the Open Archives Initiative Protocol for Metadata Harvesting

Carl Lagoze (lagoze@cs.cornell.edu)**,** Computing and Information Science, Cornell University
Herbert Van de Sompel (herbertv@lanl.gov)**,** Research Library, Los Alamos National Laboratory

Abstract

The authors, who jointly serve as Open Archives Initiative (OAI) Executive, reflect on the three-year history of the OAI. Three years of technical work recently culminated in the release of a stable production version 2 of the OAI Protocol for Metadata Harvesting (OAI-PMH). This technical product, the work that led up to it, and the process that made it possible have attracted some favor from the digital library and information community. The paper explores a number of factors in the history of the OAI that the authors believe have contributed to this positive response. The factors include focus on a defined problem statement, an operational model in which strong leadership is balanced with solicited participation, a healthy dose of community building and support, and sensible technical decisions.

Introduction

June 14th 2002 marked the official release of version 2 of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [38]. It was the most recent milestone in a process that has been cited by a number of observers for its important contribution to digital information infrastructure. For example, at a recent meeting of Principle Investigators of NSF-funded Digital Library projects [43], Hector Garcia-Molina of Stanford University cited the OAI-PMH as one of the outstanding outcomes of the last eight years of NSF-funded digital library projects. And, when announcing version 2 of the protocol to the CNI-announce list [41], Clifford Lynch praised both the technology and the process that led to its creation when writing the following:

"I believe that this is going to be a vital component of the digital information infrastructure, (…). I think that this project has been a superb model of how to rapidly develop a robust and stable protocol."

What is it about the context, process, and product of the Open Archives Initiative that contributed to this apparent success? This lead-off article to an issue on applications of the OAI-PMH provides the authors (who share the role of OAI Executive) the opportunity to reflect on these factors. We do so by focusing on four aspects that we believe contributed to its effectiveness and recognition: the definition of the problem addressed, process leading to a solution of the problem, community building and support during the process, and technical decisions that were made.

It would be presumptuous to suggest that the Open Archives Initiative provides some globally applicable template for successful infrastructure development. Nevertheless, we do propose that within the short history of the OAI there lie a number of important lessons that deserve attention from other organizations involved in similar endeavors.

Introducing the OAI – purpose and history

Readers who have stumbled upon this special issue with little or no knowledge of why the Open Archives Initiative exists or what it has been doing for the past three years need a little context; thus, the brief summary and history in this section. Readers who want more detail are directed to the following sources:

- The complete history of the Open Archives Initiative is well documented in a series of earlier documents [36, 48, 51] corresponding with each milestone.
- Complete details on the technical work of the OAI are available in the current OAI-PMH specification [38] and user guidelines document suite [37].
- The most up-to-date information on the activities of the OAI (with links to all relevant other documentation) is at the OAI website [15].

The purpose of the Open Archives Initiative is cogently stated in the OAI mission statement as follows:

> *The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. Continued support of this work remains a cornerstone of the Open Archives program. The fundamental technological framework and standards that are developing to support this work are, however, independent of the both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials.*

The primary focus of the OAI has been technical and the nature of its work so far is an *application-independent* specification for *metadata harvesting* known as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The OAI-PMH is based on a simple and subtly powerful model whereby repositories (*data providers*) make metadata (structured information) about resources available via a well-defined protocol. The exposure of the metadata allows other organizations (*service providers*) to harvest it and then aggregate it, post-process it, and refine it with the goal of developing services that add value to the metadata. Some examples of these external services are cross-repository searching, current-awareness, and reference linking.

The roots of the OAI lie in a vision to stimulate the growth of open ePrint repositories. The proposed vehicle for doing this was technology to federate and add-value to the information in those individual repositories via cross-repository services. This concept was realized in the UPS (Universal PrePrint Service) prototype [47], initiated by Herbert Van de Sompel. The prototype gathered metadata from major ePrint repositories – including arXiv [4], NCSTRL [39], CogPrints [31], and RePEc [32] – and normalized it. A search interface based on the metadata was created, linking search results to the ePrint content that remained in the originating repositories. Also, an experimental SFX linking server [46] dynamically linked the ePrint metadata to related information available in established scholarly resources such as citation databases and abstracting and indexing

databases. The prototype effectively demonstrated the advantage of services across heterogeneous repositories and successfully bridged the divide between 'free-for-all' (ePrints) and 'pay-for-all' (journals) publishing domains.

The technology demonstrated by the prototype and the recommendations brought forward by the team that created it [47], were a valuable inspiration to the UPS meeting held in Santa Fe in October, 1999 [30]. The goal of this meeting was to bring together a group of experts to identify technologies "to stimulate the adoption of the concept of author self-archived systems in scholarly communication". The meeting was funded through support from CLIR (Council on Library and Information Resources), Digital Library Federation (DLF), Scholarly Publishing and Academic Resources Coalition (SPARC), and Los Alamos National Laboratory Research Library. The presence at the meeting of a number of influential community leaders such as Deanna Marcum (CLIR), Don Waters (DLF), Clifford Lynch (CNI), and Paul Ginsparg (LANL-arXiv) lent considerable credibility to this kick-off meeting.

The effect of the UPS meeting was twofold. First, it brought together the core participants in the Open Archives Initiative leadership and community. Second, it led to the first milestone in the history of the OAI; a technical agreement for metadata harvesting known as the *Santa Fe Convention* [51]. A number of the meeting participants subsequently implemented this first experimental protocol, establishing the first members of a growing list of data providers.

In the few months after the release of the Santa Fe Convention it became apparent that the simple metadata harvesting idea had appeal to a broader reach of communities than that engaged in ePrint publishing. Members of the research library, museum, publishing, and other communities expressed interest in this low-barrier infrastructure. The DLF was especially instrumental in spreading interest in the research library community through its sponsorship of meetings at Harvard to explore the relevance of metadata harvesting as a simple vehicle for information sharing among research libraries.

This broad interest made it clear that a number of the ePrint dependencies and specializations specified in the Santa Fe Convention needed generalization for cross-community applicability. It also made it apparent that moving the OAI beyond its relatively modest base required a more established organizational model. In response, the OAI Steering Committee (OAI-SC) [16] was formed with membership reflecting the cross-community relevance of the harvesting technology. The OAI-SC established the OAI Executive consisting of Carl Lagoze and Herbert Van de Sompel. In addition, the OAI technical committee was established and then met [2] with the goal of defining a new version of the protocol that would match the broader community base.

The outcome of the work of the OAI technical committee marks the second major milestone in the history of the OAI – the release of version 1.0 of the OAI-PMH [49] in January 2001, following an extensive alpha and beta testing period. This release, accompanied by both US and European open meetings [20, 21], was explicitly labeled as *experimental.* This initiated a sixteen month experimental period during which

implementers could exchange experiences with protocol details and with the notion of metadata harvesting in general. The main vehicles for exchange were the OAI-general [11] and OAI-implementers [13] archived mail lists. In order to provide a consistent environment for experimentation, the protocol was kept as stable as possible. In fact, only one minor change was made in July 2002 [50] in response to changes in the W3C XML Schema specification.

While the OAI community was evaluating the protocol through implementations, select members of this implementers community were formed into a new OAI Technical Committee [9] with the goal of examining the experimental results and finalizing a stable version of the protocol. Following an extended period of evaluation and alpha and beta test, version 2.0 [38] of the OAI-PMH was released in June 2002 [18], marking the third and final (to date) milestone in the history of the OAI. Unlike the previous releases, OAI-PMH v2.0 is explicitly *not* experimental, and thus meant to form the basis for production software.

Measuring and evaluating "success"

Before examining some of the aspects of the OAI that underlie its professed success, it is worthwhile to describe some of the outward evidence of its impact on the digital information community. The most direct evidence is the number of repositories that support the protocol. Unfortunately, an exact count is impossible due to the fact that there is no required registration and an increasing number of applications seem to be emerging in closed intranets. The only suggestion of the actual number is in the purely voluntary OAI data provider registry [23]. Figure 1 shows the growth of registered OAI data providers since the release of the 1.0 version of the protocol in January 2001.

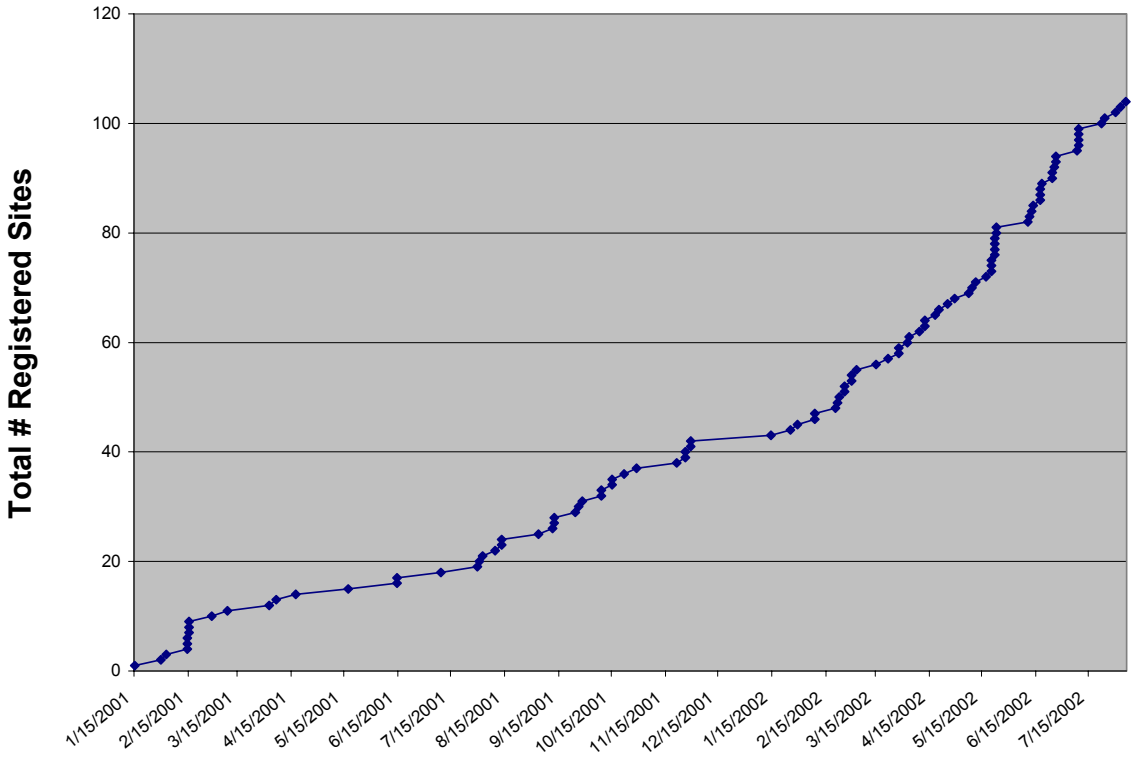

Figure 1 OAI data provider registration

Another important indicator of impact is the level of attention from funding agencies. A few notable examples of funded projects and programs that promote or are based on the work of the Open Archives Initiative are:

- **eprints.org –** is self-archiving software developed by the Electronics and Computer Science Department at the University of Southampton (UK) [6] as basic infrastructure in support of a general model of "author self-archiving". Institutions such as California Institute of Technology [5] have exploited the OAI-PMH support built-in to the software to build a federated ePrint system.
- **FAIR** – is a Joint Information Systems Committee (UK) program to fund projects that explore dissemination of institutional assets or create services, amongst others via the OAI-PMH [7].
- **Metadata Harvesting Initiative of the Mellon Foundation** – is a set of seven grants totaling $1.5M to fund development of services on top of the OAI-PMH infrastructure [52].
- **NSF National Science Digital Library (NSDL)** – is a $25M(US) five year program to build what will be perhaps the largest distributed digital library. The base technology builds on OAI-PMH [33].
- **Open Language Archives Community (OLAC)** - is a distributed, federated archive of language resources that builds on and extents the OAI-PMH [26].

Finally, and somewhat informal, are various indicators of importance in the digital library research and web communities. Both major digital library conferences in 2002, JCDL [8]

and ECDL [1], dedicated paper sessions to OAI. On October 1st 2002, Google indicated that over 3300 sites on the web linked to the OAI home page. Finally, OAI has been cited in various reports about scholarly publishing alternatives in the academic and popular press [28, 29, 44, 53].

With proper context and history established, we now turn attention in the remaining sections to the four factors that we, who have led the process, believe have contributed to its effectiveness. Do not expect a formal, scholarly analysis here. These observations are by nature subjective, yet we hope not controversial.

Focusing on a well-scoped problem

Fred Brooks in his seminal work on software engineering [27] described the "second system effect" (also known as "creeping featurism") whereby features are heaped on initially well-defined designs or systems in an uncontrolled manner. Such feature bloat has been the bane of a variety of standards and systems efforts as original realizable and important goals are lost in pursuit of more impressive targets.

We believe that the ultimate effectiveness of the OAI has benefited significantly by maintaining focus on completing a single technical task over the past three years. This task was succinctly defined at the first UPS meeting [30]; establish a low-barrier infrastructure for federating information based on the principle of metadata harvesting. While some of the details of the protocol have changed through the years, a comparison of the three major protocol versions [38, 50, 51] demonstrates that the core functional features remain constant. For example,

- The number and semantics of the protocol requests has remained nearly the same, with only the addition of the "Identify" request with OAI-PMH version 1.0.
- Key concepts such as sets, unique identifiers, multiple metadata formats, and selective harvesting (by date and set) have been consistent since the initial Santa Fe Convention.
- The main technical underpinnings such as HTTP embedding and the use of XML remain constant.

Consistency at the cost of stagnation is certainly not an admirable quality, and indeed the OAI has refined its problem statement over the course of the years. The initial Santa Fe Convention addressed interoperability among ePrint systems. Some of the initial technical details reflected this ePrint orientation, especially the bias of the mandatory oams metadata set. The goal of the next release of the protocol was generalized to metadata harvesting from content repositories in general. A number of technical changes were made to reflect this, most notably the decision to accept the de-facto standard of Dublin Core as the mandatory metadata set. Finally, with version 2.0 the problem statement was broadened somewhat more to accommodate a more general view of a *repository*. Earlier revisions of the protocol implied that repositories exposed through the OAI-PMH metadata about *contained* content. The version 1.x experimental period demonstrated that this model was too restrictive and didn't, for example, provide for repositories that might contain only metadata aggregated from other OAI-PMH data providers or metadata about non-digital objects . Rather than creating new functional

requirements, this latest refinement of the problem statement required greater generality in the data models underlying the metadata harvesting specification.

As the saying goes "the devil's in the details". To some, three years to get to a rather simple protocol with modest functionality, metadata harvesting, may seem painfully slow. What about broader interoperability problems? How about dissemination of general content? Why not handle rights management? Why not allow general search semantics? These are all valid questions. Yet we believe that the focus of the OAI on a single problem statement, albeit with minor refinements, and incremental technical enhancements has been necessary to iron out those devil's details. The result is a stable product OAI-PMH version 2.0 that effectively defines relatively simple, deployable, and, indeed, powerful technology.

Defining a clear leadership and participatory process model

What is the best process model for an organization? We make no claims to be experts in this area and suspect that the goals of an organization determine the best organizational model. In the case of the OAI, our goals, as described above, have always been primarily technical. Drafting and perfecting a useful and usable technical specification is arguably among the most exacting of tasks. Over the past three years and through three milestones, the OAI successfully delivered technical specifications at self-imposed deadlines though a process model with strong leadership, invited participation, and vigorous prototyping and testing.

The justification for strong leadership lies in the need for focus that was described in the previous section. The opportunities for an organization to drift between conflicting and expanding goals abound. We as executives of the OAI claim some credit for preventing drift from our primary mission through rather dogged insistence on focus and taking tasks to completion. The precedent for this was set at the original UPS meeting. The first day of the meeting proceeded in the manner that many workshops on "interoperability" have. The discussion wandered over a broad spectrum and for a while it looked like the incipient OAI might get lost in an attempt to solve every digital information problem. At the beginning of the second day Clifford Lynch stepped in and declared that we needed to come out of the meeting with a practical result, and then steadfastly guided the group towards unity on the metadata harvesting solution that had been recommended by the UPS prototype team. Over the past three years we have found that judicious use of "we will do this" and "we will not do that" from the OAI leadership has been instrumental to getting results on time and on target.

A recent discussion with one of the members of the OAI Steering Committee revealed interesting side-effects of a two person executive that are worth noting. Obviously having two people distributes the work and prevents overload. More important, apparently the legitimacy of decisions by the OAI Executive was reinforced by the perception that different alternatives had been explored thoroughly between Van de Sompel and Lagoze, a sort of executive "peer review".

We are not arguing for an autocratic leadership model. In fact, we believe that the effectiveness of the OAI has depended on a careful blend of strong leadership and participatory decision-making. Our model of participation has consistently been multi-tiered.

- The actual writing of the specification has been done in a small committed group consisting of Carl Lagoze, Michael Nelson, Simeon Warner, and Herbert Van de Sompel.
- The task of developing and refining the technical agenda, building the technical foundations for the next milestone, has occurred within a relatively small (less than 15 member) invitation-only technical group. The precedent for this was established in the invitation-only UPS meeting [30]. This precedent was then followed by the appointment of the two iterations of the OAI technical committee [2, 9], which were instrumental in the specification of the major releases of the OAI-PMH. Being an active implementer of the protocol has consistently been an important criterion for membership on the technical committee. This provided an important incentive for the integrity of the product coming from the technical committee since it had to meet their immediate needs. Communications amongst this technical group have taken place mainly in closed email lists and conference calls with one face-to-face meeting only when it was deemed necessary. The justifications for closed deliberations within the group lie in our belief that inviting public comment on partially-formed ideas creates an environment for loss of scope.
- This invitation-only model was followed during protocol testing. Beginning informally with the Santa Fe Convention and then followed formally during the 1.0 and 2.0 OAI-PMH releases, there were well-orchestrated periods of alpha and beta testing of the protocol. Alpha and beta testers were hand-selected from the ranks of the OAI technical committee and other committed community members [17, 19]. Testers agreed to write implementations to the developing specification and their feedback, via mail lists, was perhaps the greatest contributor to the timely and relatively bug-free release of the protocol specifications.
- The public forum has mainly been the two publicly archived and open membership email groups [11], in addition to a few select public meetings. By and large, open public comment has been elicited only after technical development is reasonably advanced and refined by the more controlled mechanisms enumerated above. This public comment has been invaluable in understanding the broader implications of technical decisions.

Building the OAI community

Simply put, digital information infrastructure is about as exciting as the plumbing in a house. The initial success of the web did not come from interest in HTTP and HTML per se, but from the vision of the early creators and adaptors who saw the potential for exciting applications built on that infrastructure.

What is the vision of OAI? It is, as mentioned earlier, a network of rich services built from harvested structured information (metadata). The best illustration of this vision lies

in the original UPS prototype [47], which when shown at the initial Santa Fe meeting served to powerfully demonstrate the potential of harvesting data from federated archives. The influence of this demonstration cannot be downplayed especially considering the level of expertise and community influence among the invited participants. Furthermore, while the mission of the OAI has extended beyond its ePrints roots, the initial enthusiasm of this vibrant community was an important kick-start for the OAI and it maintains a strong presence in the general OAI community.

The model of inspiring and involving "key players" has been a core aspect of the OAI strategy of building a broad-based community. One manifestation of this strategy was the formation of the OAI Steering Committee [16] in the middle of 2000. The OAI-SC brings together recognized experts from a variety of communities to work with the OAI Executive in establishing organizational priorities. The existence of the OAI-SC has been instrumental and its composition has effectively legitimized the OAI "brand"; a factor that should not be underestimated in dissemination into the institutional framework like libraries, publishers, and museums.

There have a number of other factors that have contributed to the formation of an OAI community. The importance of a well maintained and up-to-date website cannot be dismissed. Members of the OAI Executive, technical committee and steering committee have been invited as speakers at a variety of conferences. And, although the protocol has been promoted as low-barrier, presentation of tutorials on its use at a number of conferences has been an effective means of building community. But fundamental to the promotion of the work, have been the imaginative projects that built on the OAI-PMH, and the Tools made available by its implementers. Some have already been mentioned earlier in this paper, and it is impossible to give an adequate or complete list of such activities. Nevertheless, the work of Old Dominion University [3, 40, 42], Virginia Tech [45], and OCLC [10, 12] require special mention.

Finally (certainly not in importance) a contributory factor to the effectiveness of the OAI has been funding. As mentioned earlier, the initial UPS meeting was made possible by CLIR, DLF, and SPARC support. The Los Alamos Research Library, at which Van de Sompel was then a visiting researcher, also lent considerable support for that first meeting. The ability for the OAI-PMH to administratively exist over the past two years – in terms of maintaining a web site, registry, and conducting meetings – is due to generous support of DLF and CNI. In addition, grant funding from the National Science Foundation through Project Prism [25] – a research project investigating digital library infrastructure – has made much of the work on protocol development and testing possible. Support for implementers at both the data and service provider level has come from the DLF, Institute of Museum and Library Services (IMLS), National Science Foundation, Andrew W. Mellon Foundation [52], and the EU-IST [14].

Making good technical decisions

The primary goal of the OAI is the production and dissemination of technical specifications. As such, its success depends largely on making good technical decisions and presenting them in a manner that is useful to a broad community. We know of no

metrics for "goodness" of technical decisions, but believe that acceptance of the OAI-PMH has been due to a number of fundamental technical decisions.

In our opinion the most important technical feature of the OAI-PMH has been the clean distinction between *core* protocol and implementation-specific features. Admittedly this has evolved somewhat over time, with complete realization in the separation of the OAI-PMH version 2.0 protocol document [38] from the implementation guidelines [37]. Nevertheless, the core vs. community-dependent distinction has been built into the protocol since version 1.0 at three levels of functionality; flexibility of metadata formats, 'description' containers for collection description in the Identify verb, and 'about' containers for information about metadata. The combination of these facilities accommodates an ample amount of community and individual decision making in use of the protocol while maintaining robust cross-application interoperability.

The importance of the core vs. extension distinction stems from the fact that it makes it possible for technical work on the OAI to proceed by NOT solving certain problems (and, therefore, leave solutions in certain areas up to individual implementations). One example of this is the issue of rights statements for metadata. Rather than decide upon a (possibly unreachable) single solution to this problem, the protocol provides an "about" container, which can be carried with a disseminated metadata record, and might contain a variety of rights statements. The positive side-effect of this is that it has encouraged other projects to examine the rights management issue in the context of requirements of specific communities and uses (see Project RoMEO [22]). The same approach applies to issues like metadata provenance and certification for which finding a single "global solution" might be impractical.

This purposeful decision to not handle certain functionality is also evident in the basic data vs. service provider distinction. Take for example the notion of adding to the protocol the ability to specify Boolean filters on harvesting requests (e.g., request only metadata records by a certain author and specific keywords in the title). By partitioning this off as functionality added by a specific service provider (for perhaps a specific community or need) we avoided adding it at the core protocol level where a single solution may indeed have been difficult to find.

There are a number of other technical decisions made during the development of the OAI-PMH that are worthy of reflection.

The decision of participants at the initial Santa Fe Meeting decided to build the first protocol [51] as a subset of the functionally richer Dienst protocol [34] was important in two ways. First, by not immediately re-inventing the wheel it was possible to produce relatively quick technical turn-around from the first meeting and demonstrate the validity of the harvesting concept. Second, scaling down the functionality of the Dienst protocol (keeping scope, as mentioned earlier) significantly reduced the cost of implementation; a factor that has played a continuing role in the protocol specification.

Clean and consistent models have also played an important role in the technical architecture. Foremost of these is the data and service provider distinction, which has been sufficiently flexible to stand up to a variety of applications of the OAI-PMH. Other basic modeling concepts in the protocol such as records, items, unique identifiers, and sets have similarly proven to be robust in a number of contexts. The evolution of these concepts over the course of the three protocol milestones reflects refinement rather than change in models in response to a broader definition of data providers (especially to include metadata aggregators).

The decision to use HTTP as the transport layer of the OAI-PMH was essential to keeping barriers to implementations low. The stateless nature of HTTP certainly places some limitations on applications built on top of it. In comparison with the highly functional session semantics of Z39.50 [24], the state maintenance mechanisms of OAI-PMH, using resumption tokens, are rather crude. On the other hand, creating and maintaining an OAI-PMH server in tandem with an existing web server (e.g., Apache) is a relatively simple task.

Finding the proper middle-ground on metadata is not an easy task. Members of the information community have experienced many 'metadata mud fights' over the nature and number of metadata fields. On reflection, it appears that it was a sound decision to require one base metadata format while accommodating and promoting multiple community-specific metadata formats (modeled on the Warwick Framework [35] in the Dublin Core community). Furthermore, the decision to move from an ePrint specific core metadata format to de-facto standard unqualified Dublin Core has made the protocol accessible to a broader domain.

XML and XML Schema have proven to be effective and flexible tools for structuring data in the OAI-PMH. The use of XML Schema for validation of responses has made it possible to build conformance mechanisms, for example the repository explorer [45] and OAI registry [23].

Looking towards the future

What is the future of the Open Archives Initiative? Do the processes and products of the past three years provide some foundation for further work? These are questions that at the time of writing of this paper (October 2002) remain unanswered. We close then with two speculations rather then definite statements.

First there is the issue of whether to pursue formal standardization. We have explored both between ourselves as the OAI Executive and with selected colleagues the plausibility of officially standardizing the protocol, using one of the available standards organizations such as NISO, IETF, or the W3C. There are certainly benefits from standardization including a greater chance of adoption by vendors and organizations (e.g., governments sometimes require official standards status). However, the process of standardization is non-trivial and its costs in human effort need to be weighed carefully against these benefits. Our inclination is to continue the current de-facto standards

process (i.e., standardization through widespread adoption) until the real need for official status becomes obvious.

Second there is the issue of where to go next technically. As stated throughout this paper we are both keen on well-focused tasks with clear deliverables. Furthermore, we are convinced of the distinction between interoperability at the core level and interoperability at the individual community level. Therein sits some of the architectural bases of the OAI-PMH. Thus, we are wary of future work in extending the core functionality of the protocol, for example in the area of rights management of complex content handling.

We are attracted to exploring applications of the core protocol that meet requirements for specific communities. Since this effort began with the goal of providing foundation technology to promote ePrint archives, we are inclined to return to those roots. We are currently developing a set of requirements of the ePrint community and determining how to best serve those requirements layered on top of the core functionality of the OAI-PMH. This work is still in its initial stages and will proceed only if there is the sense that a number of key technical solutions can be developed. We do believe that work in this selected area might provide a model for other community-specific technical layers on top of the OAI-PMH.

Acknowledgements

Support for Open Archives Initiative activities comes from the Digital Library Federation, the Coalition for Networked Information, and from National Science Foundation Grant No. IIS-9817416. The authors acknowledge the fantastic support this work has gotten from the OAI Steering Committee, OAI Technical Committee, and the OAI community as a whole.

References

[1] *6th European Conference on Research and Advanced Technology for Digital Libraries*, 2002 http://ecdl2002.iei.pi.cnr.it.

[2] *Agenda for the Cornell meeting of the Technical Committee of the Open Archives Initiative, September 7-8 2000*, 2000 http://www.openarchives.org/meetings/tech-Cornell/cornell-agenda.htm.

[3] *ARC harvester and search engine*, 2002 http://arc.cs.odu.edu/.

[4] *arxiv.org e-Print archive*, http://www.arxiv.org.

[5] *Caltech Library System Papers and Publications*, 2002 http://caltechlib.library.caltech.edu/.

[6] *eprints.org*, 2002 http://www.eprints.org.

[7] *FAIR Programme (Focus on Access to Institutional Assets)*, 2002 http://www.jisc.ac.uk/events/02/fair.html.

[8] *Joint Conference on Digital Libraries 2002*, 2002 http://www.ohsu.edu/jcdl/.

[9] *OAI Technical Committee*, 2001 http://www.openarchives.org/organization/tech.comm.html.

[10] *OAICat*, 2002 http://www.oclc.org/research/software/oai/cat.shtm.

[11] *OAI-general list and archives*, 2002 http://www.openarchives.org/mailman/listinfo/OAI-general.

[12] *OAIHarvester*, 2002 http://www.oclc.org/research/software/oai/harvester.shtm.
[13] *OAI-implementers list and archive*, http://www.openarchives.org/mailman/listinfo/OAI-implementers.
[14] *Open Archives Forum*, 2002 http://www.oaforum.org/.
[15] *The Open Archives Initiative*, 2002 http://www.openarchives.org.
[16] *Open Archives Initiative Appoints Steering Committee*, 2000 http://www.openarchives.org/news/oaiscpress000825.html.
[17] *Open Archives Initiative launches alpha test of new specifications*, 2000 http://www.openarchives.org/meetings/steering/alpha-testing-press-release.htm.
[18] *Open Archives Initiative Release Version 2.0 of the Protocol for Metadata Harvesting*, 2002 http://www.openarchives.org/news/oaiv2press020614.html.
[19] *Open Archives Protocol for Metadata Harvesting v.2.0 scheduled for release*, 2002 http://www.openarchives.org/news/OAI-2.0A-PR.html.
[20] *Open Meeting Renaissance Hotel, Washington DC, January 23, 2001*, 2001 http://www.openarchives.org/meetings/DC2001/OpenMeeting.html.
[21] *Open Meeting Staatsbibliothek zu Berlin, Germany, February 25, 2001*, 2001 http://www.openarchives.org/meetings/Berlin2001/OpenMeeting-agenda.html.
[22] *Project RoMEO*, 2002 http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html.
[23] *Registered Data Providers*, 2002 http://www.openarchives.org/Register/BrowseSites.pl.
[24] *The Z39.50 Document*, http://lcweb.loc.gov/z3950/agency/document.html.
[25] . *Cornell Project Prism*, http://www.prism.cornell.edu/main.htm.
[26] S. Bird and G. Simons, *OLAC Protocol for Metadata Harvesting*, 2001 http://www.language-archives.org/OLAC/protocol.html.
[27] F. P. Brooks, *The mythical man-month: essays on software engineering*, Anniversary ed. Reading, Mass.: Addison-Wesley Pub. Co., 1995.
[28] D. Butler, "The future of electronic scientific literature," *Nature*, September 6, September 6, 2001.
[29] R. Crow, *The Case for Institutional Repositories: A SPARC Position Paper*, 2002 http://www.arl.org/sparc/IR/ir.html.
[30] P. Ginsparg, R. Luce, and H. Van de Sompel, *The Open Archives Initiative aimed at the further promotion of author self-archived solutions*, 1999 http://www.openarchives.org/meetings/SantaFe1999/ups-invitation-ori.htm.
[31] S. Harnad, *CogPrints Project Page*, 2000 http://www.ukoln.ac.uk/services/elib/projects/cogprints/.
[32] T. Krichel, *Access to Scientific Literature on the WWW: the RePEc concept*, 1998 http://openlib.org/home/krichel/osborne.html.
[33] C. Lagoze, W. Arms, S. Gan, D. Hillmann, C. Ingram, D. Krafft, R. Marisa, J. Phipps, J. Saylor, C. Terrizzi, W. Hoehn, D. Millman, J. Allan, S. Guzman-Lara, and T. Kalt, "Core Services in the Architecture of the National Digital Library for Science Education (NSDL)," Cornell University, Ithaca, arXiv Report cs.DL/0201025, January 29 2002. http://arxiv.org/abs/cs.DL/0201025.
[34] C. Lagoze and J. R. Davis, "Dienst - An Architecture for Distributed Document Libraries," *Communications of the ACM*, 38 (4), pp. 47, 1995.

[35] C. Lagoze, C. A. Lynch, and R. D. Jr., "The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata," Cornell University Computer Science, Technical Report TR96-1593, June 1996. http://cs-tr.cs.cornell.edu:80/Dienst/UI/2.0/Describe/ncstrl.cornell/TR96-1593?abstract=.

[36] C. Lagoze and H. Van de Sompel, "The Open Archives Initiative: Building a low-barrier interoperability framework," presented at Joint Conference on Digital Libraries, Roanoke, VA, 2001.

[37] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, *Implementation Guildelines for the Open Archvies Initiative for Metadata Harvesting*, 2002 http://www.openarchives.org/OAI/2.0/guidelines.htm.

[38] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, *The Open Archives Initiative Protocol for Metadata Harvesting - Version 2.0*, 2002 http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html.

[39] B. M. Leiner, "The NCSTRL Approach to Open Architecture for the Confederated Digital Library," *D-Lib Magazine*, December, 1998.

[40] X. Liu, "DP9 Service Provider for Web Crawlers," *D-Lib Magazine*, 7 (12), 2001.

[41] C. Lynch, *Open Archives Metadata Harvesting Protocol Version 2*, 2002 http://www.cni.org/Hforums/cni-announce/2002/0029.html.

[42] K. Maly, M. Zubair, and X. Liu, "Kepler - An OAI Data/Service Provider for the Individual," *D-Lib Magazine*, 7 (4), 2001.

[43] National Science Foundation, *DLI2 - Digital Libraries Initiative Phase 2*, 2002 http://dli2.nsf.gov/.

[44] B. Nicholls, "Open Meta Tools," *BYTE*, 2002.

[45] H. Suleman, "Enforcing Interoperability with the Open Archives Initiative Repository Explorer," presented at Joint Conference on Digital Libraries, Roanoke, 2001.

[46] H. Van de Sompel, "Reference Linking in a Hybrid Library Environment, Part 1: Frameworks for Linking," *D-Lib Magazine*, 5 (4), 1999.

[47] H. Van de Sompel, T. Krichel, and M. L. Nelson, "The UPS Prototype: an experimental end-user service across e-print archives," *D-Lib Magazine*, 6 (2), February, 2000.

[48] H. Van de Sompel and C. Lagoze, "Notes from the Interoperability Front: A Progress Report on the Open Archives Initiative," presented at Joint Conference on Digital Libraries, Portland, OR, 2002.

[49] H. Van de Sompel and C. Lagoze, *The Open Archives Initiative Protocol for Metadata Harvesting - Version 1.0*, 2001 http://www.openarchives.org/OAI/1.0/openarchivesprotocol.htm.

[50] H. Van de Sompel and C. Lagoze, *The Open Archives Initiative Protocol for Metadata Harvesting, Version 1.1*, 2001 http://www.openarchives.org/OAI/1.1/openarchivesprotocol.htm.

[51] H. Van de Sompel and C. Lagoze, "The Santa Fe Convention of the Open Archives Initiative," *D-Lib Magazine*, 6 (2), 2000.

[52] D. Waters, "The Metadata Harvesting Initiative of the Mellon Foundation," ARL, ARL Bimonthly Report 217, August 2001. http://www.arl.org/newsltr/217/waters.html.

[53] J. Young, "'Superarchives' Could Hold All Scholarly Output," *The Chronicle of Higher Education*, July 5, 2002.