

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

E-JASL 1999-2009 (volumes 1-10)

E-JASL: The Electronic Journal of Academic
and Special Librarianship

Fall 2004

Language-Related Open Archives: Impact on Scholarly Communities and Academic Librarianship

Jung-ran Park
Drexel University

Follow this and additional works at: <https://digitalcommons.unl.edu/ejasljournal>



Part of the [Communication Technology and New Media Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

Park, Jung-ran, "Language-Related Open Archives: Impact on Scholarly Communities and Academic Librarianship" (2004). *E-JASL 1999-2009 (volumes 1-10)*. 42.
<https://digitalcommons.unl.edu/ejasljournal/42>

This Article is brought to you for free and open access by the E-JASL: The Electronic Journal of Academic and Special Librarianship at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in E-JASL 1999-2009 (volumes 1-10) by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Language-Related Open Archives: Impact on Scholarly Communities and Academic Librarianship

Jung-ran Park, Assistant Professor
College of Information Science and Technology, Drexel University

The evolution of new forms of scholarly communication since the advent of Web technology has brought unprecedented opportunities for potential global connection among the rapidly growing number of electronic repositories among scholarly communities. Under the open archive infrastructures, scholarly resources that had been invisible to Web search engines and thus afforded limited dissemination and access are now becoming increasingly visible with speedy and wide distribution. This paper addresses the emergent issues and challenges faced by academic librarians: participation in archiving, organization, and preservation of open repositories; integration of Web-based repositories into traditional collections; and mediation and direction of academic users into this new realm of rich resources.

1. Introduction

The advent of Web technology has brought unprecedented opportunities to scholarly communities by providing a dramatically different communication mode from the traditional paper-based one. Web-based infrastructure has provided highly efficient means for the production and dissemination of scholarly resources, while new Web-based communication modes have contributed to overcoming limitations posed by traditional scholarly communication. Among these limitations: high expense for production, storage and dissemination; limited distribution and access; and slow turn-around time from production to dissemination.

Inexpensive mass storage technology allows large resources to be stored in digital form. As well, Web-based communication has transformed the static text-based output of the traditional scholarly publication into that of dynamic multimodal (e.g. integrated sound, texts, transcripts, visual image, etc.) production, thus generating rich resources for scholarly communication. It has also provided efficient means for wide distribution and access to scholarly communities and the public. The discovery of diverse resources can be especially promising in digital production when standardized vocabularies are employed for indexing such resources. In this sense, new digital production modes and dissemination of resources holds the potential for global connection to the rapidly expanding multitude of resources.

However, for Web-based scholarly communication to reach its full potential, standardization of metadata and lexicon (i.e. employment of controlled vocabularies), is a must in the indexing and harvesting of diverse Web resources. Without such a standardized classification scheme for organizing and indexing Web resources, low recall and precision are inevitable in information retrieval. Another prerequisite for global connection is an interoperable technological infrastructure among the rapidly growing number of Web-based repositories.

Recognition of the drawbacks (i.e. low precision and recall) and limitations of Web-search engines in discovering scholarly resources;¹ centralized archive encompassing previously scattered scholarly resources; the necessity for interoperability and standardization among index terms, data formats, and data encoding schemes; the necessity for the archiving and long-term preservation of linguistic and cultural heritage; and the awareness of the rapidly growing number of Web resources and concomitant recognition of the unprecedented potential of Web technology for scholarly communication have together spurred the creation of three initiatives. These are:

- Dublin Core Metadata Initiative (DCMI), founded in 1995 in Dublin, Ohio²
- Open Archives Initiative (OAI), founded in 1999 in Santa Fe³
- Open Language Archives Community (OLAC), founded in 2000 in Philadelphia⁴

This article aims at advancing the involvement of scholarly communities across a range of disciplines, including registration of the special collections of academic libraries with the OLAC. The aim here is to spur the development of a digital library of language-related repositories. As will be discussed later, there are endless disciplines connected to language resources and, in consequence, the value of the OLAC to a variety of scholarly disciplines is potentially enormous. In the following sections, I will outline the value of metadata, a critical component in the foundation of the technical infrastructure used in creating digital archives including the OLAC. I will also briefly touch upon the OAI, on which the technological infrastructure of the OLAC is founded.

This article also aims at addressing the significant impact of Web-based scholarly communication on academic collection development by introducing the archives and open source tools that are currently registered to the OLAC. I will also address the necessity for proactive participation in the archiving, organizing, and preserving of repositories and integration of these digital repositories into the traditional collection. Finally, I will touch on the mediating and directing of academic scholars into this new integrated realm of rich resources.

2. What are Metadata and Why Do We Care?

Metadata, or data about data, is not a novel concept. The library and information communities have employed metadata for organizing and discovering information for centuries. Traditional metadata came from the library card catalog in which a physical object, such as a book, was indexed to a pertinent metadata description. Metadata is also a familiar concept to the general academic community, even though the term *per se* might be unfamiliar.⁵ Thus, a citation of a book consists of metadata: the citation describes information about the book and the metadata within the citation provides access points and aids for users to locate the particular book.

To illustrate, the following citation style contains data about the book *Intellectual Foundation of Information Organization*:

Svenonius, Elaine. (2000). *Intellectual Foundation of Information Organization*. Cambridge: MIT Press.

The above citation includes the following pieces of metadata, in order: date of publication, creator, title, place of publication, and publisher. Library catalogs are much richer in describing given physical objects such as books, videos/DVDs, sound recordings, maps, etc., through provision of refined descriptive metadata such as the table of contents, subject descriptors, summary description, and other pertinent descriptive notes.

As shown, metadata have long been widely employed by library and information professionals and scholarly communities in the organizing of information and the discovery of resources. The Library of Congress Subject Headings (LCSH) is a standardized set of metadata tailored to the description of a physical object. Such standardized metadata have generated enormous power for bibliographic control and, in consequence, for building a centralized union catalog such as the OCLC (Online Computer Library Center) Online Union Catalog. Bibliographic control through the centralized union catalog has significantly contributed to the ability of scholarly communities to discover relevant resources.

As mentioned at the outset, the advent of Web technology has presented unprecedented opportunities to discover and access rapidly growing scholarly resources. Scholarly communication through the Web is significantly different from the traditional mode in its formats and speed of production and dissemination. The advance of multimodal information systems (text, image, sound, etc.) has also contributed to dynamic digital production. The speed of production has also generated rapidly expanding vast resources. Most importantly, through the advent of Web technology, the potential for

the global connections across the diverse and scattered resources of scholarly communities has become realizable.

Recognition of the necessity for creating community-driven standardized index terms tailored to these digital repository resources moved the scholarly community to develop and establish the Dublin Core (DC) metadata set, which was created through a broad interdisciplinary consensus.⁶ The fifteen DC metadata elements, i.e. title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage and rights, are all optional and repeatable.⁷

The salient characteristics of the DC metadata set emerge from its simplicity, flexibility, and interoperability.⁸ The functionalities of the DC metadata set are easy-to-implement owing to this simplicity, which creates high compatibility across multiple repositories. (However, Park has pointed out that there are inevitable hindrances during the process of mapping metadata elements across repositories that employ non-DC metadata schemes such as MARC (Machine Readable Cataloging)).⁹

Inasmuch as the DC metadata set is not tailored to a specific community-driven resource, the extension and refinement of the DC metadata set may be an inevitable step in order to adequately describe community-specific resources. Based on this, in the scheme “DCMI Metadata Terms” (a version of “Dublin Core Qualifier”), “refinements” and “encoding scheme” are allowed.¹⁰ Refinement qualifiers make “the meaning of an element narrower or more specific.”¹¹ Encoding scheme qualifiers “identify schemes that aid in the interpretation of an element value. These schemes include controlled vocabularies and formal notations or parsing rules.”¹² The OLAC metadata set,¹³ which will be discussed later, is an instance of such an extension and refinement of the DC metadata set.

As can be seen, metadata is neither a novel nor a complicated concept. However, employing standardized metadata for Web-based scholarly communication is fundamental to ensuring successful recall and precision of digital resources across rapidly expanding multiple repositories on the Web. Employment of standardized metadata is also critical to realizing the potential of global connection across the multiple repositories of scholarly communities.

3. OAI (Open Archives Initiative) and OLAC (Open Language Archives Community)

Based on the foundation of the Dublin Core metadata standard, the OAI was launched in late 1999 out of a project named “Libraries Without Walls” by the Research Library at Los Alamos National Laboratory.¹⁴ It started as a forum envisioning technical solutions to the transformation of scholarly communication in Santa Fe. The convention

“specified how electronic preprint repositories could share metadata with third parties, to support the establishment of cross-repository discovery services.”¹⁵

The infrastructure of OAI is founded on the DC metadata set and the OAI Protocol for Metadata Harvesting (OAI-PMH¹⁶) to support interoperability across diverse electronic preprint (e-print) repositories. The mechanism for interoperability creates the potential for global connection among individual archives that are scattered and incompatible to a centralized and interoperable integrated block. When realized, this will lead to the wide distribution of scholarly works; through this development, individual scholars stand to reap great benefit in being able to reach a much wider universe of users for their works. As Herbert Van de Sompel states:

Santa Fe recommendations to interoperability at the level of metadata harvesting: 1. The definition of a set of simple metadata elements—the Open Archive Metadata Set (OAMS)—for the sole purpose of enabling coarse granularity document discovery among archives; 2. The agreement to use a common syntax, XML, for representing and transporting both OAMS and archive-specific metadata sets; 3. The definition of a common protocol—the Open Archives Dienst Subset—to enable extraction of OAMS and archive-specific metadata from participating archives.¹⁷

At the beginning, the initiative limited the scope of cross-repository discovery to e-print resources. However, the scope of repositories has been significantly broadened to include digital resources as suggested in the mission statement:

The roots of the OAI lie in the E-Print community, which promotes and maintains web-accessible archives of scholarly papers as a means of increasing access to scholarly research. Initial work in the OAI was motivated by a desire to develop interoperability frameworks for federating E-Print archives. It soon became evident, however, that the concepts in the OAI interoperability framework—exposing multiple forms of metadata through a harvesting protocol—had applications beyond the E-Print community. Therefore, the OAI has adopted a mission statement with broader application: opening up access to a range of digital materials.¹⁸

The OAI defines the usage of the term ‘open’ in the following way: defining and promoting machine interfaces that facilitate the availability of content from a variety of providers.¹⁹

Thus, openness is seen as creating centralized service providers through the OAI metadata harvesting protocol by allowing content from diverse data providers. Openness also signifies reproduction and reuse by third parties as Van de Sompel points out:

an open machine interface that enables third parties to collect data from the archive. ...facilitating the broad dissemination of archive data through third party services is a crucial feature of an e-print archive.²⁰

The impact of OAI in transforming scholarly communication has been enormous. Participating open archives operating within the infrastructure of the OAI currently comprise 192 data providers.²¹ However, because registration is optional the actual number of adopters of the OAI-PMH is unknown. The OAI in effect functions as a springboard for sub-communities in the building of community-specific open archives and repositories.

To illustrate, the OLAC, which mainly comprises language and culture-related resources, was founded on the framework of the OAI infrastructure (i.e. the DC metadata standard and the OAI-PMH [metadata harvesting protocol]) in December 2000 through an NSF-funded workshop on Web-based Language Documentation and Description held at University of Philadelphia. The following is from the statement describing motivations of the workshop:

...lay the foundation of an open, web-based infrastructure for collecting, storing and disseminating the primary materials which document and describe human languages, including wordlists, lexicons, annotated signals, interlinear texts, paradigms, field notes, and linguistic descriptions, as well as the metadata which indexes and classifies these materials. The infrastructure will support the modeling, creation, archiving and access of these materials, using centralized repositories of metadata, data, best practice guidelines, and open software tools.²²

Participants in the workshop comprise a group of approximately 100 language software developers, linguists, and archivists hailing from North America, Europe, Africa, the Middle East, Asia, and Australia. The following is the mission statement of OLAC:

An international partnership of institutions and individuals who are creating a world-wide virtual library of language resources by: (i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resources.²³

As mentioned earlier, because the DC metadata set is not tailored to community-specific resources, special metadata is necessary for language-related resources based on the DC metadata framework to adequately describe language-related resources. The OLAC metadata set²⁴ is the result of this extension of the DC metadata set. The elements of the OLAC metadata set consist of the 15 DC data elements together with special metadata for language resources through the employment of DC qualifiers: attributes and encoding scheme (i.e. controlled vocabularies). The OLAC metadata set employs

three attributes: *refine*, *code* and *lang*. The *refine* attribute identifies element refinements; the *code* attribute is used for “holding metadata values that are taken from a specific encoding scheme;” and the *lang* attribute “specifies the language in which the text in the content of the element is written.”²⁵

Let me briefly touch on the controlled vocabularies for the *code* attribute. For language resource classification, ‘language identification’ (a language that the content of the resource describes) and ‘linguistic type’ (the nature or genre of the content of the resource) are critical components. For ‘language identification,’ the OLAC adopted SIL’s Ethnologue, which is superior to the language identification standard (ISO 639)²⁶ because of its complete scheme of language identifiers.²⁷ Concerning the ‘linguistic type’ resource, four top-level types such as *transcription*, *annotation*, *description* and *lexicon* are distinguished. For each of these top-level types, more specific subtypes can be utilized. For instance, sub-types such as *wordlists*, *wordnets*, *thesauri*, etc., could distinguish the *lexicon* type further.

The primary service provider for the OLAC archives is the Linguist List.²⁸ By virtue of the centralized single gateway to the OLAC archives, i.e. the Linguist List Website, end users benefit in high recall and precision. The standardized metadata and control vocabularies ensure that individual archives are consistently described and make possible federated searching across all language-related archives from a single site. This enables end-users to discard unnecessary steps in the searching of individual repositories that are scattered and incompatible; in turn, the discovery of pertinent language and culture related resources is maximized.

Since the foundation of OLAC in late 2000, the OLAC standards (i.e. metadata set and harvesting protocol) have been applied to the wider academic community. According to the OLAC timeframe, the OLAC standards were further refined based on experience during the pilot phase; the OLAC operational phase began in early 2003.²⁹ Thus, archives planning to register with the OLAC will have a more solid foundation derived from refinements implemented during the pilot phase.

The activities of OLAC have been recognized by a variety of mass media such as *BBC News* (“Digital race to save languages”),³⁰ *Weird News* (“Word Up: Keeping Languages Alive”),³¹ and *Scientific American* (“Saving Dying Languages”).³² Active outreach by OLAC coordinators Gary Simons and Steven Bird, through presentations and articles published in various scholarly journals is noteworthy—this outreach will touch a variety of scholarly communities that are potential data providers for the OLAC.³³

In the following section, I will introduce 30 archives that are currently registered with the OLAC.

4. Language-Related Digital Archives: Impact on Scholarly Communities and Academic Librarianship

Language confers humanity. Acquisition of a mother tongue is one of the most prominent characteristics distinguishing human beings from animals. Our cognitive activities are also closely interlocked with the faculty of acquiring a native language. The fundamental medium for human communication, knowledge organization and discovery, and information delivery across time, space, and generations is language. Moreover, the embodiment and inheritance of human intellectual and cultural heritage is made possible through the core medium of human language expressed through other media such as paper, audio-visual recordings, microform, digital media, etc., together with the advancement of technology and socio-economical change through the passage of time.

In this sense, linguistics, the discipline dealing with language, is a meta-discipline as Susan Hockey pointed out in the workshop which generated founding the OLAC:

This initiative is particularly interesting because linguistics is a meta-discipline. It impacts on almost everything that is done in our daily lives. What is developed as a result of this workshop may have implications throughout the scholarly community and beyond...^{[34](#)}

Steven Bird and Gary Simons enunciate the same meta-disciplinary characteristics in the following way:

The list of disciplines which study some aspect of language is virtually endless: linguistics, phonetics, psychology, anthropology, philosophy, cognitive science, neuroscience, speech science, political science, history, literature, language teaching, literacy, translation, information science, communication studies.^{[35](#)}

The commonality of scholarly communities across various academic disciplines is owing to the role of language, even though there are obviously differences in the depth and breadth of language and language-related resources across different disciplines. In addition, considering the fact that most digital libraries have been built along a single discipline, the foundation of the OLAC, which comprises virtually all scholarly communities, brings an inestimable added value to scholarly communities.

The activities of the OLAC address crucial issues that academic information professionals need to take note of. Scholarly communities, especially linguists, have been archiving, disseminating, and preserving language-related resources comprising secondary sources such as research papers and conference proceedings as well as primary sources such as field notes, transcriptions of spoken corpora, dictionaries, digi-

tized texts, audio and video recordings, and open source tools. As well, linguists have been greatly concerned with and engaged in building a centralized digital library and developing tools for collecting, organizing, and preserving endangered cultures and languages by employing emergent technologies.

Let me now turn to the OLAC archives that are currently registered to the OLAC as of the update of the site on October 29, 2004. The compass of these 30 archives is international in scope in that the archives comprise American, European, Australasian and Panpacific countries. The archives listed below can be accessed at the following OLAC page: <http://www.language-archives.org/archives.php4>.

A description of how OLAC archives are organized provides essential insight into the manner in which these resources can be utilized by various academic and community groups. The large numbers of OLAC archives are composed of various types of resources. However, for the purposes of this paper they can be categorized into three subject domains.

First, there are several archives that concern preservation of indigenous and endangered languages and cultures. The activities of documenting these resources using survey and interview methods in consultation with native speakers and subsequently preserving such resources in digitized form through the utilization of metadata are directly related to the information needs of humanities scholars. The archives function as primary sources for the furthering of research on human heritage across indigenous languages and cultures. These archives are mostly composed of ethnographic resources such as audio-recordings of interviews with text transcriptions, naturally-occurring discourse, ritual speech, songs, etc.

The following are the archives related to this category:

- Survey for California and Other Indian Languages
- Aboriginal Studies Electronic Data Archive (ASEDA)
- Archives of the Indigenous Languages of Latin America
- AIATSIS ASEDA (Australian Institute of Aboriginal and Torres Strait Islander Studies, the Aboriginal Studies Electronic Data Archive)
- LACITO (Langues et Civilisations a Tradition Orale) Archive
- Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)
- Perseus Digital Library
- Tibetan and Himalayan Digital Library
- Virtual Kayardild Archive
- Alaska Native Language Center Archive
- UQ Flint Archive

Second, there are several large-scale OLAC archives that are composed of mostly open source tools dealing with human language technology, covering electronic dictionaries, electronic textual databases and multimedia and multi-modal databases that integrate speech, text and gesture that in turn are linked to audio-visual media and natural language processing software such as a parser and speech recognizer. These archives evince great value for attracting scholars across various academic disciplines such as the humanities, library and information science, engineering and computer science, etc. The large extent of human language technology software such as ontologies and lexicons in turn has laid the foundation for constructing semantic tools toward knowledge representation and information retrieval on the Web.

There are also multilingual open source tools that can be utilized for retrieving information across different language boundaries. Considering the fact that development of semantic tools for cross-lingual and cross-cultural information retrieval has been spurred by advancement of web technologies and globalization trends, such open source tools have a great potential for furthering studies in this area and for providing information needs of scholars from across a variety of disciplines. In addition, open source tools such as parser for processing written and spoken texts, speech annotation, speech recognizer, etc. have a great value for developing spoken language interface and for retrieving multimedia and multimodal resources. Research papers in computational linguistics are also available.

The following are the related archives:

- TalkBank
- The LDC (Linguistic Data Consortium) Corpus Catalog
- The Natural Language Software Registry
- Oxford Text Archive
- ATILF Resources
- Boiste
- CHILDES Data repository
- TRACTOR Test Archive
- European Language Resources Association (ELRA)
- A Digital Archive of Research Papers in Computational Linguistics

Third, archives of documentation of over 8000 languages across the world and of linguistic and ESL (English as Second Language) studies are the following:

- Rosetta Project 1000 Language Archive
- Ethnologue: Languages of the World
- Academia Sinica Balanced Corpus of Modern Chinese
- Academia Sinica Formosan Language Archive

- Academia Sinica Tagged Corpus of Early Mandarin Chinese
- Cornell Language Acquisition Laboratory (CLAL)
- SIL Languages and Culture Archives
- Typological Database Project
- Surrey Morphology Group Databases
- Bielefeld Language Archive

As shown, the activities of OLAC are parallel to ones of information professionals to the extent of collection, resource organization by utilizing human language technology and standardization, distribution and provision of access, preservation of language and culture related resources. In this respect, the demarcation between humanities scholars and information professionals has become blurred. Without engaging these impending issues through proactive involvement in the building of digital archives, the ground for academic librarians stands to become weaker. The following table illustrates how the usage of DC metadata varies among different institutions:³⁶

Table
Variations in DC Element Usage

1

	Subject	Description
Digital Libraries (10 total, 122,719 records)	78%	36%
Museums, historical societies, etc. (6 total, 255,800 records)	93%	93%
Academic libraries (7 total, 235,294 records)	15%	13%

As can be seen, DC metadata participation by academic libraries is significantly lower than other institutions such as museums. The building of open archives by humanities scholars and the report on the usage of the DC metadata shown above suggest that proactive participation by academic libraries in building scholarly digital repositories is a necessity. Academic catalogers have created metadata for physical objects for centuries. It is time for catalogers to organize and provide valuable access points through metadata tailored to digital resources, such as the DC metadata set for the digital repositories.

To academic librarians in the areas of reference and instruction, the OLAC archives are excellent sources to direct and mediate OLAC archives to users. The ever-growing number of digital archives has generated enormous challenges to preservation and

record-keeping due to safety and longevity/permanence concerns.³⁷ Special collections dealing with language and culture will benefit by registering with the OLAC, inasmuch as the collections will be accessible to a much wider audience among scholarly communities. (The following site gives registration instructions on how to become an OLAC data provider: <http://www.language-archives.org/register/archive.html>.) OLAC archives are invaluable resources and should attract attention from collection development librarians so that these archives can be integrated into the traditional collection. Virtual collection development and related management and other issues are eminently necessary.

5. Conclusion

The evolution in scholarly communication since the advent of Web technology has brought unprecedented opportunities for the potential global connection of the rapidly growing multitude of electronic repositories scattered among scholarly communities. Under the open archive infrastructure, scholarly resources, including primary sources that have been invisible to Web search engines and thus have had limited dissemination, and access are now becoming increasingly visible with a speedy and wide distribution.

In addition, the diversity of data formats enables scholarly communities to conduct the richest possible study. For instance, in the study of the lexicon of Middle English, diverse data sources such as digitized texts, images, open source tools for describing the pronunciation of Middle English, secondary papers that are peer-reviewed, etc., have increasingly been accessible to the relevant scholarly community.

Standardized metadata and controlled vocabularies ensure that individual archives are consistently described and enable federated and interoperable searching across archives. Owing to the well-defined infrastructure of the Open Archives Initiative and subcommunities of OAI that are compliant with its infrastructure (e.g. OLAC), the end-users of certain scholarly communities are able to discard unnecessary steps in searching individual and multiple repositories that are scattered and incompatible. In consequence, full exploitation of pertinent resources in research becomes realizable.

This highlights the issues and challenges that academic librarians must tackle in fostering proactive participation in archiving, organizing, and preserving repositories; integrating these Web-based repositories into traditional collections; and mediating and directing academic users into the new realm of these rich resources.

Academic catalogers have created metadata for physical objects for centuries. It is time for catalogers to organize and provide valuable access points to open repositories through metadata tailored to digital resources as exemplified by the DC metadata set. In the areas of reference and instruction, the open repositories are excellent sources by

which to direct and mediate users to sources. Special collections dealing with language and culture will benefit greatly by registering with the OLAC, as the collections will be accessible to a much wider audience among the scholarly communities.

Open archives are an invaluable resource requiring attention from collection development librarians so that they can be integrated into traditional collections. The ever-growing number of digital archives has engendered enormous challenges to preservation and record-keeping librarians owing to safety and longevity concerns related to the digital materials. These issues need to be recognized and tackled in order for academic librarians to stand on the solid ground of gatekeeper and mediator to the proliferating number of scholarly resources.

Dr. Park is currently an assistant professor at the College of Information Science and Technology at Drexel University. Her teaching areas are cataloging and classification, metadata, and information resources in the humanities. Prior to her current position, she held the position of cataloger and subject specialist in languages, literature, and linguistics at Indiana State University.

Notes

1. Steven Bird and Gary Simons, "The OLAC Metadata Set and Controlled Vocabularies," *Proceedings of ACL/EACL Workshop on Sharing Tools and Resources for Research and Education* (2001).
Available: <http://arXiv.org/abs/cs/0105030> October 30, 2004.
2. Dublin Core Metadata Initiative. Available: <http://dublincore.org> (October 30, 2004).
3. Open Archives Initiative. Available: <http://www.openarchives.org> (October 30, 2004).
4. Open Language Archives Community. Available: <http://www.language-archives.org> (October 30, 2004).
5. Jeff Good, "Gentle Introduction to Metadata," *OLAC Notes* (2002).
Available: <http://www.language-archives.org/documents/gentle-intro.html> (October 30, 2004).
6. Bird and Simons, "The OLAC Metadata Set and Controlled Vocabularies," *Proceedings of ACL/EACL Workshop on Sharing Tools and Resources for Research and Education* (2001).
Available: <http://arXiv.org/abs/cs/0105030> (October 30, 2004).

7. Dublin Core Metadata Element Set, Version 1.1: Reference Description (February 2003). Available: <http://dublincore.org/documents/2003/02/04/dces/> (October 30, 2004).
8. Stuart Weibel, "Dublin Core: A Simple Content Description Model for Electronic Resources," *Discovering Online Resources* (archive version, July 2001).
9. Jung-ran Park, "Hindrances in Semantic Mapping among Metadata Schemes: A Linguistic Perspective," *Journal of Internet Cataloging* 5(3) (2002): 59-79.
10. DCMI, "DCMI Metadata Terms," (March 2003). Available: <http://dublincore.org/documents/2003/03/04/dcmi-terms/> (October 30, 2004).
11. Ibid.
12. Ibid.
13. OLAC metadata. Available: <http://www.language-archives.org/OLAC/metadata.html> (October 30, 2004).
14. Los Alamos National Laboratory. Available: <http://lib-www.lanl.gov/lww/> (October 30, 2004).
15. The Open Archives Initiative (2002). Available: <http://lib-www.lanl.gov/lww/oai.htm> (October 30, 2004).
16. Open Archives Initiative Protocol for Metadata Harvesting. Available: <http://lib-www.lanl.gov/lww/oai.htm> (October 30, 2004).
17. Herbert Van de Sompel and Carl Lagoze, "The Santa Fe Convention of the Open Archives Initiative," *D-Lib Magazine* 6 (2) (February 2000).
18. Open Archives Initiative Frequently Asked Questions (FAQ). Available: <http://www.openarchives.org/documents/FAQ.html> (October 30, 2004).
19. Ibid.
20. Van de Sompel, "The Santa Fe Convention of the Open Archives Initiative," *D-Lib Magazine* 6, no. 2 (February 2000).
21. OAI, Registered Data Providers. Available: <http://www.openarchives.org/Register/BrowseSites.pl> (October 30, 2004).

22. *Linguistic Exploration: Workshop on Web-Based Language Documentation and Description* (2000). Available: <http://www ldc.upenn.edu/exploration/expl2000/> (October 30, 2004).
23. OLAC Mission. Available: <http://www.language-archives.org/> (October 30, 2004).
24. Bird and Simons, “OLAC Metadata,” (2002a). Available: <http://www.language-archives.org/OLAC/metadata.html>. (October 30, 2004)
25. Bird and Simons, “The OLAC Metadata Set and Controlled Vocabularies,” *Proceedings of ACL/EACL Workshop on Sharing Tools and Resources for Research and Education* (2001). Available: <http://arXiv.org/abs/cs/0105030> (October 30, 2004).
26. ISO. “ISO 639: Codes for the Representation of Names of Languages” (1998). Available: <http://lcweb.loc.gov/standards/iso639-2/langhome.html> (October 30, 2004).
27. Barbara F. Grimes (editor), *Ethnologue: Languages of the World*. Dallas: Summer Institute of Linguistics, 14th edition (2000). Available: <http://www.sil.org/ethnologue> (October 30, 2004).
28. The Linguist List. Available: <http://linguist.emich.edu/olac> (October 30, 2004).
29. OLAC Timeframe. Available: <http://www.language-archives.org/tools.html> (October 30, 2004).
30. Webster, “Digital race to save languages,” *BBC News* (March 20, 2003). Available: <http://news.bbc.co.uk/2/hi/technology/2857041.stm> (October 30, 2004).
31. Kendra Mayfield, “Word Up: Keeping Languages Alive” *Wired News* (November 4, 2002). Available: <http://wired.com/news/culture/0,1284,54345,00.html> (October 30, 2004).
32. Wayt Gibbs, “Saving Dying Languages,” *Scientific American* (August 2002). Available: <http://www.language-archives.org/docs/sciam.pdf> (October 30, 2004).

33. Bird and Simons, "Building an Open Language Archives Community on the OAI Foundation," *Library Hi Tech* 21, no. 2 (2003). Available: <http://www.arxiv.org/abs/cs.CL/0302021> (October 30, 2004).
Bird and Simons, "Requirements on the Infrastructure for Open Language Archiving," (2000a). Available: <http://www.language-archives.org/docs/requirements.html> (October 30, 2004).
Bird and Simons, "The Seven Pillars of Open Language Archiving: A Vision Statement," (2000b). Available: <http://www.language-archives.org/docs/vision.html> (October 30, 2004).
Bird and Simons, "OLAC Protocol for Metadata Harvesting," (2001). Available: <http://www.language-archives.org/OLAC/protocol.html> (October 30, 2004).
Bird and Simons, "OLAC Overview," (2001). Available: <http://www.language-archives.org/documents/overview.html> (October 30, 2004).
Gary Simons, "Language Identification in Metadata Descriptions of Language Archive Holdings," Paper presented at the workshop on *Web-Based Language Documentation and Description* (2000). Available: <http://www ldc.upenn.edu/exploration/expl2000/papers/simons/simons.htm> (October 30, 2004).
34. Susan Hockey, "Towards a Model for Web-based Language Documentation and Description: Some Contributions from Digital Libraries and Humanities Computing Research," Paper presented at the workshop on *Web-Based Language Documentation and Description* (2000). Available: <http://www ldc.upenn.edu/exploration/expl2000/papers/hockey/hockey.htm> (October 30, 2004).
35. Bird and Simons, "Seven Dimensions of Portability for Language Documentation and Description," *Language* 79 (3) (September 2003): 557-582.
36. Timothy Cole, "Using OAI-PMH to Aggregate Metadata Describing Cultural Heritage Resources," Presented at the *ALA/CLA Annual Conference* (June 2003, Toronto, Canada). Available: http://dli.grainger.uiuc.edu/publications/twcole/ala2003oai/ALA2003_OAI.ppt (October 30, 2004).
37. Bird and Simons, "Seven Dimensions of Portability for Language Documentation and Description," *Language* 79 (3) (September 2003): 557-582.

[Back to Contents](#)

http://southernlibrarianship.icaap.org/content/v05n02/park_j01.htm.