

Running head: OPEN ARCHIVES REPOSITORIES

Open Archives Repositories:

Evidence of Citizen-State Interaction

Kristin Yiotis

LIBR 256.01 Archives and Manuscript Management

San Jose State University, School of Library and Information Science

Lori Lindberg, Professor

April 18, 2005

## Abstract

Budgetary shortfalls at libraries nationwide and steadily increasing costs for peer-reviewed scholarly journals have librarians and researchers concerned that current practices in publishing will hamper the dissemination of information and dampen the interchange of ideas. A paradigm shift is in the making brought about by digital technology and the Internet. Open Archives repositories enable a more socially responsible and equitable way of publishing and disseminating scholarly communications. These archives represent points of intersection between societal functions, institutions, and people who come into contact with them and create a clear image of society.

Open Archives Repositories:  
Evidence of Citizen-State Interaction

This paper attempts to answer the question are Open Archives Initiative (OAI) repositories really archives? Are they archives in the traditional, Jenkinsonian sense of the term, or merely called archives as are many electronic records that are no longer current? I will discuss OAI repositories, what they are, where they come from, who invented them, when they first started, and how they work. I will discuss the notion of documents, records, and archives as providing evidence or proof of the interaction between the citizen and the functions and activities of the state (Cook, 1992). I will argue that OAI repositories exhibit the characteristics of archives, have permanent value, and show a clear image of society.

What are Archives?

Jenkinson (1984) and Schellenberg (1956) designate as archival materials those that have three basic elements: recordness—they consists of records created and accumulated to accomplish a purpose, the record's primary purpose; preservation—they are preserved and used for reasons other than those for which they were created, the record's secondary purpose; and provenance—they demonstrate an unbroken line of custody and original arrangement that provides proof of authenticity and reliability.

*Why do archives exist?*

Archives exist in order to establish for posterity a true representation of society recorded as the activities of societal institutions and agencies. Records provide proof that the activities of societal institutions and agencies happened. The accumulation of the whole body of records provides a true image of society as a whole. According to Jenkinson (1984) archival materials come together or accumulate by a neutral process in the natural order of a civilized, literate

society (p. 18). According to Cook (1992, p. 46) archives exist in order to establish for posterity a true representation of society, which archival records provide. According to Duranti (1994, pp. 334-5) the four characteristics of archival records—impartiality (inherently truthful), authenticity (unbroken custody), naturalness (accumulate naturally), and interrelationship (unique place in the whole)—ensure the evidentiary nature of archives.

### What is Self-Archiving?

#### *A Subversive Proposal*

In 1994, Stevan Harnad posted "A Subversive Proposal" to a Virginia Polytechnic Institute list serve. Harnad, Professor of Cognitive Science at Princeton, New Jersey, and University of Southampton, United Kingdom, was for many years a researcher and editor of Behavioral and Brain Sciences. In 1990, he introduced *Psycology* (2004), the first peer-reviewed scientific journal on the Internet, and in 1997, the Cognitive Sciences Eprints Archive (CogPrints, 2004). In 1998, he started the American Scientist Open Access Forum, a high volume discussion list concerned with Open Archives (Harnad, 2005)

Harnad's proposal was the inspiration for self-archiving and the Open Archives Initiative. He suggested that all scholars publish their preprints of unpublished, unrefereed, original work on a globally accessible, local archive, freely available anywhere in the world to all other scholars with network access. If a work gets formally published, then everyone will substitute the published work for the preprint, his point being that scholars need not withdraw preprints from public viewing after refereed versions are accepted for paper publication. Once this becomes common, journal publishers will be forced to restructure their costs for the electronic-only versions to be truer to actual costs, which he estimated to be 25 percent less than the paper page costs (Harnad, 1994). He also suggested that the cost for local archiving be built in to the cost of

research and be paid in advance rather than by the end user. If the current publishers didn't restructure, then a new generation of electronic-only publishers would take over the market.

Harnad's intention was to get around restrictive copyright legally. Self-archiving unrefereed preprints before submitting the paper to a journal enables the author to negotiate to hold, rather than transfer, copyright. If the author holds copyright, she would self-archive the refereed postprint. If the author loses copyright, he would self-archive the corrigenda, the differences between the preprint and the postprint (Harnad, n.d. Resolving). Either way the entire article would be freely available and the author's research impact continue unfettered.

### *Self-archive repositories*

The scientific community had been using electronic files for self-archiving and transferring scientific literature since 1991 when the first centralized archive was begun, arXiv.org, a physics archives out of Los Alamos, New Mexico, now owned and operated by Cornell University (Gustafson & Pitman, 2004). Until standards emerged that allowed for cross archive searching, institutional repositories were not interoperable; hence, archiving did not guarantee research impact, a major reason scholars publish their findings. Self-archiving involves depositing a digital document at a publicly accessible, institutional Web site.

In 1999, the Open Archives Initiative convened in Santa Fe, New Mexico, to work out "a technical and organizational framework to support basic interoperability among archives" (Digital Library Federation, 2001) The framework instituted OAI compliance, enabling interoperability among electronic archives such that all can be harvested, integrated, navigated, and searched seamlessly, as if they were all in one global archive. Interoperability guarantees that any user anywhere in the world can search archives in repositories also located anywhere (Harnad, 2001, May) The user could search, via a cross-archive search engine, a virtual archive

and retrieve documents from university or institutional archives, whether the archive were centrally located or distributed (ARC, 2004, Homepage).

*Eprints software and metadata harvesting*

The technical breakthrough that makes interoperability possible is extensible mark-up language (XML). Interoperability involves a single Web interface where the depositor enters XML metadata tags for date, author-name, title, journal-name, and then attaches the full-text document (OpCit, n.d., Self-archiving). Full text documents can be in different formats and locations, but the XML metadata tags make them interoperable. The interoperable interface was developed by an international organization called the Open Archives Initiative (OAI). The software that supports cross archive interoperability is GNU Eprints developed at Massachusetts Institute of Technology (MIT) and the University of Southampton, Great Britain in 1985 (Free Software Foundation, 2005, GNU). GNU (pronounced “guh-noo”) is an operating system that is compatible with UNIX (Free Software Foundation, 2005, Definitions). Eprints software enables the whole open archives system to work technologically.

The Open Archives Initiative established a registry for OAI-compliant archives using Eprints software (OpCit, n.d., Registry). Any individual or institution running a UNIX operating system can download Eprints software for free, set up a self-archiving repository, and register with OAI, thus making their local archives part of a worldwide, searchable archives (Eprints.org, 2004). Free software means the user is free to run, to study, to distribute, and to improve the program. Each open access journal becomes its own open archive to become a part of a worldwide Open Access Initiative resource of refereed research literature (Suber, 2004, April).

Today the Open Archives Initiative’s Protocol for Metadata Harvesting (OAI-PMH) (2004) provides “an application-independent interoperability framework based on metadata

harvesting” (Introduction section). Harvesting refers to the process of collecting metadata from repositories. A repository is an online server running Eprints software that can process OAI-PMH requests. Interoperability requires that repositories use Dublin Core metadata. Metadata is assigned a unique identifier that identifies an item within a repository. The framework allows for two types of participation: At the input end, data providers use the protocol as a means for compiling metadata in a repository. At the output end, service providers use the protocol to harvest metadata in response to a search (2004, Introduction section). The protocol relies on the unique identifiers to extract metadata from the item (2004, Definitions and concepts section).

*OAI service providers and data providers*

One OAI service provider at the output end is ARC: A Cross Archive Search Service maintained by Old Dominion University Digital Library Research Group (2004). ARC is used to harvest OAI compliant repositories by making them accessible through a unified search interface. ARC is not a production service but searches other repositories. Users enter a search for a document in ARC by seeking keyword matches on any bibliographic field: author, title, type, language, archive, subject, accession date, discovery date, or abstract fields (ARC, n.d., Help). Then ARC goes out to search all worldwide OAI compliant data providers listed in its registry called DP9: An OAI Gateway Service for Web Crawlers, a gateway to the deep Web (ARC, n.d., DP9).

For example: DP9 lists the Haverford College Senior Thesis Archive as repository number 310 (ARC, n.d., DP9, Repository name: Haverford College). The hyperlink brings you to lists of documents in its repository. If you click on record “oai:HaverfordCollegeThesis.OAI2:44,” you will find the Dublin Core metadata and abstract for Diana Postemsky’s (2003) BA thesis, “Through the Looking-Glass: Reading and Reflecting from Wide Sargasso Sea to

Jane Eyre.” Metadata tags—title, creator, subject, description, contributor, publisher, date, type, format, identifier, source, language, and rights—are completed by the depositor, as explained in the Help page at Haverford College Senior Thesis Archive (n.d., User documentation, Depositing records section). The thesis is available in full text as a PDF document at the Haverford College Library Senior Thesis Archive.

OAIster is OAI service provider from the University of Michigan Digital Library Production Service (2004). Its goal is to create a collection of freely available, previously difficult to access, academically oriented, digital resources that are easily searchable by anyone. As of March 2005, OAIster holds 740,751 records from 390 institutions’ open archive repositories (University of Michigan, 2004). Perseus Digital Library Lookup Tool, out of Tufts University (2004), is another cross archive search service.

OAI data providers administer local repositories, such as the Haverford College Senior Thesis Archive, repository number 301 in ARC, and Vanderbilt University Electronic Thesis and Dissertation Archive, repository number 167 (ARC, n.d., DP9, Repository name: Vanderbilt Electronic Thesis and Dissertation Archive). The document “Life Cycle Extension Strategies For Legacy Systems” by Autumn Sellars is available at the Vanderbilt University Archive as record “oai:VANDERBILTETD:etd-07062004-151044.”

Open Archives service providers, like ARC, OAIster, and Perseus Lookup Tool, harvest, index, and search the metadata of the data providers. OAI metadata protocol enables participants in the OAI community to share information when they register as data and service providers at the ARC Community Web site (OAI, n.d., Community). The service providers uses Eprints software as a means for collecting metadata about the contents in data providers.

*Citation search engines*

All OAI Eprints archives are citation-interlinked so that the research literature can be navigated by citation. CiteBase Search (Eprints.org, n.d.) was developed by the Open Citation Project (OpCit, 2004) as a search and citation analysis tool with data from August 1999 to the present. Users search by metadata tags, citations, or OAI identifiers (Eprints.org, n.d.). The OpCit Project (2004) was a three-year collaboration between Southampton University, Cornell University, and arXiv.org that ended in 2002. OpCit developed products and services that support OAI. One of its accomplishments was to explore the relationship between usage and impact for free online papers. The issue of impact is a serious one that has been studied extensively. Impact is a major reason why scholars publish research findings. Scholars fear that self-archiving and online publishing will limit the impact of their research.

OpCit developed CiteBase as a tool to measure the usage and impact of OAI published papers. OpCit (2004, October) reported on studies tracking the effect of open archives. Results show that OAI publishing increases impact. Antelman (2004) reports, “across all four disciplines...philosophy, political science, electrical and electronic engineering and mathematics...freely available articles do have a greater research impact.”

OAI advocates calculate that the research impact of Eprints repositories will increase as researchers in all disciplines self-archive their research, both pre-refereeing preprints and refereed postprints. The Core Metalist of OAI Eprint Archives presents the scale of author self-archiving (Open Citation Project, 2003). The list is organized into the following categories: OAI archives, Lists of institutional archives, Eprints.org archives, Gateways (indexes), Open access journal archives, Disciplinary archives. Under Library and Information Science, for instance, are E-LIS and Dlist. The total number of journals and local repositories is enormous.

### Do OAI repositories exhibit the characteristics of an archive?

According to the International Council on Archives (ICA), “archives are records made or received and accumulated by a person or organization in the course of affairs and preserved because of their continuing value” (Lindberg, n.d. Introduction 1 presentation). What is a record? Lindberg defines records as “a special class of documents...that provide permanent evidence of...an event...or some type of activity,...usually evidence of a transaction.” What is a document? Lindberg (2003) defines documents as “items containing information organized in way that makes sense and can be used as some type of reference.” Documents can exist as “discrete, independent, identifiable units of text (Lindberg, 2003). An example is the Vanderbilt University Electronic Thesis and Dissertation Archive. It holds the document “Life Cycle Extension Strategies For Legacy Systems” by Autumn Sellars, a form of document known as a thesis (ARC, n.d., DP9). In this archive, “oai:VANDERBILTETD:etd-07062004-151044” is the record for the Sellars document, and as a record, it contains evidence of a transaction in the form of metadata. The information compiled in the thesis is organized by the author and represents her knowledge. The university’s archive consists of records that have been organized by attaching metadata tags for permanent storage and retrieval.

### *Do OAI repositories provide evidence?*

The records in OAI repositories show qualities of recordness because metadata tags record evidence of transactions that ensure the authenticity and uniqueness of this document. Metadata tags that make that document record unique also make it retrievable: title, creator, subject, description, contributor, publisher, date, type, format, identifier, source, language, and rights. The metadata tags give this record a unique place within the context of all other records in the collection. “All records are documents, but a record needs context to have meaning”

(Lindberg, 2003), which in this case means usefulness to a user. Meaning can be derived from indexing categories as well as physical location in relationship to the whole body of records (Lindberg, 2003). Recordness provides an additional kind of knowledge that is derived from the context of the records and exists beyond the knowledge within the documents themselves.

The knowledge derived through qualities of recordness establishes evidence of the context of its creation as a record, evidence of its relationship to other records not only in that database but, through the interoperability of metadata tags, to all other databases. “An archival collection of records shows evidence of continued activity over time and provides knowledge...that can lead a researcher to even more related evidence existent in other collections of records” (Lindberg, 2003), in this case through metadata tags.

OAI documents begin as research articles, a type of text that scholars use to record and publish their knowledge and communicate it to others. The documents act as vehicles for conveying and communicating information and knowledge to others. When these articles are published, the knowledge they contain gets wide exposure and impacts the research of others. A publication’s wider dissemination ensures that knowledge is transferred more widely, which in turn ensure the growth, progress, and expansion of knowledge for society as a whole. Benefits for the individual scholar follow a Darwinian model: the more one publishes the greater one’s exposure; the greater one’s exposure the stronger one’s impact, the stronger one’s impact the greater one’s influence, the greater one’s influence the higher one’s status in the field, the higher one’s status the more opportunity for faculty appointments.

Are OAI repositories really archives? Do they display the three basic characteristics of archives: recordness, preservation, and provenance? OAI records are archived in OAI compliant repositories to ensure a permanent and persistent URL for long-term access and preservation

(Suber, 2004, SPARC, Case for OAI section). OAI repositories qualify as archives because they display the characteristics of archives and show a clear image of society.

Do OAI repositories provide a clear image of society?

Cook (1994) considers the computer as causing a “fundamental revolution affecting the... nature of society’s collective memory” (p. 301). “For the first time, we have business officers and professionals creating and storing their own records...(Cook, p. 302). Archivists are “trying to understand and preserve logical and virtual patterns that give electronic information its structure, content, context, and thus its meaning as a ‘record’” (p. 302). “We are beginning to shift our professional attention from archives [as place] to archiving [as management]” (p. 302). “Archivists and information managers can contribute to this...revolution [moving from place to context]...by maximizing the power of creating and using records in context, that is, provenance” (p. 303). This shift in emphasis from physical records to conceptual management, from looking after physical objects to focusing on the functional context in which the record-creating activity takes place” (p. 303) will give archivists greater opportunity in the history of the profession.

One way that this shift in emphasis from physical records to conceptual management is being expressed is that people are doing their own archiving. My argument is that OAI repositories represent the interaction between societal function, societal programs and activities, and the citizen/clients. Cook’s (1992) macro-appraisal model entails analyzing societal functions to determine the programs and activities where the citizen-state interactions are clearest. The model moves archives from being a place to being a functional context. The three dimensions, structure, function, and client, must be represented in the citizen-state interaction if the records of that interaction are to create an image of society worth preserving for future generations (Cook,

1992, p. 57). Functions are mores, traditions, and purposes established by a culture. This abstract concept is made concrete through structures. Structures are societal institutions that carry out societal functions. Clients are people in general (Cook, 1992, p. 50).

Cook doesn't plan to preserve all records because electronic records make this impossible, only the records that contain the dialectic, the interaction between the citizen and the state in circumstances showing citizens' "direct input to the agency...where the citizen consciously interacts with the agency (structure) and programme (function) and has influence on the decisions made" (p. 52). Cook says that those records with enduring value will be those exhibiting revealing or information laden interactions, or points of intersection between societal functions, institutions, and people who come into contact with them (p. 53). This point of contact represents the activity between the institution (structure), the individual (citizen-client), and the purpose (the function) this institution is designed to fulfill (p. 53). My argument is that OAI repositories exhibit characteristics of citizen-state interaction making them potentially "hot spots" or highly relevant archives where the image of society is most direct (Cook, 1992, p. 58).

What is the traditional system of scholarly communication?

Scholarly journals have been disseminating knowledge and stimulating intellectual thought since the invention of the printing press. The printing press as a technology enabled the development of knowledge and the exchange of information leading to growth and expansion in scholarship and invention. I consider the traditional system of scholarly publications a powerful and influential societal structure that intersects with citizens/clients, those scientists who create research articles, taxpayers who fund research articles, scholars who search for and read the research articles, and the librarians who retrieve research articles.

The entire process of scholarship depends on the free exchange of information, from

disseminating the latest research findings to preserving them for future use. Built into this traditional system are standards for evaluating research by way of editorial boards made up of scholars who referee or determine who and what gets published based on merit alone (ARL, ACRL, SPARC, 2000, Scholars lost). Also built into this system is the willingness of the creators of this research, the scientist and scholar writers, to submit their articles free of charge to professional journals, and the willingness of editors and referees to peer-review the articles also free of charge. Costs come about in the printing and publishing of the journals, not in the payment of royalties or fees to the writers or editors (ARL, ACRL, SPARC, 2000, Scholars lost). Scholars willingly publish their research in professional association journals not for financial, but for professional, gain. The more work one publishes, the greater one's chances for tenured faculty positions and research grants. Publishing exposes one's ideas to as wide an audience as possible and produces impact and professional recognition.

The system of scholarly communications that has existed for hundreds of years consists of research and other scholarly writings created free of charge, edited or peer-reviewed also free of charge, printed and published at cost, and sold to libraries and research institutions for dissemination (Harnad, 2001, The (refereed) literature). Functionally, the system can be viewed as having six parts: creation, quality control, production, distribution, consumption, and support (ARL, ACRL, SPARC, 2000, Scholars Lost). Creation, the key function of the model, is the domain of scholars. Quality control, the editorial process through rigorous peer review, is also the responsibility of scholars. Production is the job of the publishers. Distribution is handled by both libraries and publishers, with libraries disseminating the finished publication to most readers. Consumption is also the domain of scholars along with students and nonscholars. Underlying the system is the support provided by institutions, such as universities, governments,

granting institutions, and taxpayers (ARL, ACRL, SPARC, 2000, Scholars lost).

*The system gets distorted as more research is published*

An explosion in scholarly research and information began taking place in the 1960s and 1970s, brought about by increased federal spending in support of higher education, and resulting in great advances in science and technology (ARL, ACRL, SPARC, 2003, Create change). A bonanza of federally funded research dollars flowed into educational institutions during the Cold War and its aftermath, and university faculties increased because of expanded enrollments. Many universities aspired to and won research status. The pressure on faculty to publish and to get research support through grants grew. And with that growth, the quantity of research grew beyond the capacity of the scholarly publication system, which was then still dominated by scholarly societies.

Enterprising commercial publishers became interested in the potential profits to be made from publishing in the context of a well-established creative source and an equally well-established pattern of consumption. Scholarly societies off loaded journals to be published and managed by commercial publishers. Existing journals expanded and new journals were formed until a majority of the market was in the hands of commercial publishers. Eventually, scholarly communication became a multi-billion dollar, international business. By the end of the twentieth century, “ownership of information—ownership of content created by scholars and then given away by them—had become the linchpin of a huge and profitable industry”(ARL, ACRL, SPARC, 2000, Scholars under siege). Initially faculty benefited by “hav[ing] an increasing pool of publications that enhanced their chances to win tenure and promotion both through publication and through other opportunities, such as appointment to editorial boards” (ARL, ACRL, SPARC, 2000, Scholars lost).

By the mid 1980s, commercial publishers began a process of consolidation and mergers that enabled them to gain monopolies on all publishing in their designated fields, particularly the journals in science, technology and medicine (STM). Along with these monopolies came huge subscription hikes beyond inflation and cost of living increases, and profits margins of 20 to 40 percent per year, the levels of luxury market goods such as fine jewelry, yachts, and expensive cars. Association of Research Library (ARL) 1999 data show Science Publishing Division, Reed Elsevier, operating with a 35 to 40 percent profit margin (ARL, ACRL, SPARC, 2000, Scholars lost). “During the period 1990-2000, the average journal subscription price increased by more than 10 percent per year for a ten-year total increase of nearly 170 percent (ARL, ACRL, SPARC, 2000, Scholars under siege).”

The price hikes impacted library budgets. For libraries to keep a constant level of serial subscriptions while prices are rising, they have to cut back in other areas. ARL statistics, 1997 to 1998, show an annual average increase in serial costs ranging from 8 to 10 percent in 1994 to 1998 (Kyrillidou, 1999). While the average ARL library managed to keep constant the level of serial subscriptions, they had to make changes elsewhere in the libraries’ collections, mostly by reducing purchases of monographs. But by 2000 and beyond, libraries could no longer maintain the constant level of serials subscriptions and began a steady cut back in subscriptions in addition to paying more for the journals they continued getting (Kyrillidou, 1999).

#### A new paradigm: Open Access Archives

Scientists, scholars, and librarians began fighting back, from the grassroots level of individual scholars making decisions about publishing, to worldwide organizations concerned with issues such as equal access to information and ownership of knowledge. Harnad’s (1994) subversive proposal instigated the idea of a new paradigm emerging in the post-Gutenberg era,

enabled by Internet technology, with the aim of shifting the control of knowledge resources away from commercial publishers and back to scholars if they reclaimed ownership and copyright of their work.

In 1998, the Association of Research Libraries (ARL) launched SPARC, the Scholarly Publishing and Academic Resources Coalition, an alliance of university research libraries and organizations. SPARC addresses the high cost of scholarly journals by supporting competitive and open access repositories. SPARC's goal is to expand information dissemination using a networked digital environment by helping existing journals adopt an open access format and by forming partnerships with new journals to get them started (ACRL, n.d., SPARC). SPARC's agenda focuses on three strategic pursuits: incubation, advocacy, and education, for which it has developed an arsenal of resources (ACRL, 2004, About SPARC). Four programs are at work: The Partners Program sets up partnerships for incubating new journals and converting existing ones to open access. Scientific Community partnerships develop nonprofit information portals that serve specific scientific communities, such as the Directory of Open Access Journals (DOAJ) (Lund University, 2004). Alternative partnerships set up competitive alternatives to existing high-priced titles, especially in science, technology, and medicine (STM). Leading Edge partnerships represent a paradigm shift by presenting scholars with new outlets for their papers. Examples are the Public Library of Science (PLoS), the New Journal of Physics, and BioMed Central (PLoS, n.d.). This array of journals is reflected in the OAI Core Metalist categories discussed earlier.

In 1999, SPARC started the Create Change campaign as its advocacy and education arm. Create Change works with university faculty and librarians to build momentum to further the cause of Open Access by providing information and resources at its Web site, such as an FAQ,

Advocacy Kit, brochures, graphs and charts that display the history of book and journal costs from 1986 to 2002 (ARL, ACRL, SPARC, 2003, Create change).

SPARC also publishes a newsletter called SPARC Open Access Newsletter (SOAN), written by Peter Suber, Professor of Philosophy at Earlham College in Indiana, and an advocate of open access. Suber began his own newsletter in 2001, the Free Online Scholarship (FOS) Newsletter, and then in 2002, when he combined forces with others in the movement, changed the name to SOAN to comply with the Budapest Open Access Initiative. In 2002, Suber added a daily Web log to which others can contribute (ACRL, 2004, About SPARC). Suber has articulated, in the well-reasoned manner of the philosopher, the position of the Budapest Open Access Initiative, of which he is one of the original signatories, and his voice is unmistakable in much of the literature. His FOS Web site has virtually unlimited pages, an excellent frequently asked questions (FAQ), a glossary of terms, a time line of the history of the movement, articles on where the movement stands today, the FOS and SOAN newsletter archives, and links to conferences and other sites involved in the movement (Suber, 2004, SPARC newsletter). While he clearly understands the technology, his main concern is in the sociological, legal, and philosophical aspects of the movement, bringing in people from the humanities, social sciences, and arts, to what has been an endeavor mainly of science, medicine, and technology.

*An image of society that matches reality*

In a 2003 article in C&RL News, Suber argues the major thesis of the Open Access Initiative: Both the serials pricing and permission crises can be solved by open access, first because it is free of charge to users, and second because copyright holders consent in advance to unrestricted reading, downloading, copying, sharing, storing, printing, searching, linking and crawling of open access articles (Removing barriers). If scholars retain copyright to their work,

then they consent to give users open access to research articles for which they expect no payment. If scholars transfer the copyright to the traditional publisher, then the publisher will erect price and permission barriers to prevent open access (Suber, 2003, Removing barriers). The operative factor is that scholars never expected to make any profit from their work anyway. Open access is ideal for this unique form of literature—refereed, scholarly articles—and works only because scholars retain copyright. Scholars are most interested in the impact and exposure that unrestricted publishing provides; therefore scholars benefit by consenting in advance to unrestricted and unlimited copying of their work. Publishers have an entirely different motive, which is profit.

Suber's 2003 article was followed by a White Paper published by Association of College and Research Libraries (ACRL), "Principles and Strategies for the Reform of Scholarly Communication," which spearheaded ACRL's Scholarly Communication Initiative. This initiative became the central focus of ACRL Issues and Advocacy Forum and remains so today, evidenced by Scholarly Communications 101, ACRL's preconference workshop at the 2004 American Library Association (ALA) Annual Conference. The purpose of the ACRL initiative is to work in partnership with other library and higher education organizations to encourage reform of the system and broaden involvement by academic libraries (ARC, 2004, Issues).

#### *International support for Open Access*

The Budapest Open Access Initiative (BOAI) of 2002 was an important milestone in freeing scholarly communication because it unified, in a single statement and under a common name and purpose, the different terms that the many groups used for the same idea: Free Online Scholarship (FOS), Scholarly Communication Initiative, Immediate Free Web Access (IFWA), Refereed Literature Liberation Movement, Intellectual Property Conservancy (IPC). The BOAI

advocates open access for scholarly journal articles and elicits signatures from individuals and institutions at its Web site (Open Society Institute, n.d.). Its parent organization, the Open Society Institute founded by George Soros, the Hungarian financier, is active in persuading foundations and other organizations to donate resources, lead by a donation of \$3 million from Soros (Open Society Institute, 2002, BOAI). The Initiative endorses two strategies for achieving the goal of open access to scholarly journal literature: institutional repositories that use the metadata tagging standards created by the Open Archives Initiative, and open access journals, both the creation of a new generation of journals committed to open access and helping existing journals make the transition to open access (Open Society Institute, 2002, Read).

Another major player is the World Summit on the Information Society (WSIS), a United Nations sponsored Working Group concerned with closing the gap in access to scientific knowledge and information between have and have not nations. Open access is a small portion of WSIS. WSIS advocates open access as a move toward removing access barriers comparable to removing political and economic boundaries, such as the European Union (EU). Open access signifies the democratization of knowledge and supports a socially responsible way to distribute knowledge. Open access makes the same knowledge and information available to scholars in wealthy first-world nations, in developing ex-communist, second-world nations, and in underdeveloped third-world nations. WSIS's Open Access campaign and the existence of Open Archives repositories is the outcome of growing desires, worldwide, for an equitable and democratic distribution of information resources between have and have not nations, and here within the United States, for a more sensible allocation of limited tax dollars, such that those who fund the research get the benefits of the research freely and without additional payment. Open Access archives are evidence that the world is moving in the direction of democratization of

information and knowledge by removing access restrictions in the form of copyright protection or fee-based dissemination policies.

### Conclusion

If, according to Cook's (1992) macro-appraisal model, the records of permanent value are those displaying distortions, abnormalities, or effecting change to the existing societal structures and functions, then OAI self-archiving repositories represent "hot spots" or places where the image of society is most direct (p. 58). Harnad called his proposal subversive because its purpose was to change a powerful, existing societal structure. To figure out where the image of society is clearest archivists ask, "What are the reasons for and the nature of the communication between the citizen and the state?" For the purposes of this argument the state equals the intellectual community in general. Citizens are self-archiving directly into repositories without involving commercial interests embodied in commercially published journals—powerful, existing societal structures. I argue that OAI repositories contain records that are created in contexts that involve conflicting elements of society, thus providing more evidence and a better image of society to future generations. The act of self-archiving allows for a more direct and clearer image of contemporary society. Therefore OAI repositories qualify as archives that have permanent value and should be preserved.

## REFERENCES

- American Library Association (ALA).(2004). *Preconferences*. Special events. Retrieved Jan 25, 2005, from <http://www.ala.org/ala/eventsandconferencesb/annual/an2004/preconferences.htm>.
- Antelman, K. (2004, September). Do open-access articles have greater research impact? *College and Research Libraries* 65(5) 372-82. Retrieved December 5, 2004, from [http://www.lib.ncsu.edu/staff/kantelman/do\\_open\\_access\\_CRL.pdf](http://www.lib.ncsu.edu/staff/kantelman/do_open_access_CRL.pdf).
- ARC-A cross archives search service. (2004). *Homepage*. Retrieved January 11, 2005, from <http://arc.cs.odu.edu/>.
- ARC. (2004). *Community*. Retrieved December 5, 2004, from <http://www.openarchives.org/community/index.html>.
- ARC. (2004). *DP9- An OAI gateway service for web crawlers*. Retrieved December 4, 2004, from <http://arc.cs.odu.edu:8080/dp9/index.jsp> .
- ARC. (n.d.). *Archive groups*. Retrieved January 11, 2005, from <http://128.82.7.99:8080/oai/results.jsp>.
- ARC. (n.d.). *Help*. Retrieved December 4, 2004, from [http://128.82.7.99:8080/oai/service\\_help.html#specific](http://128.82.7.99:8080/oai/service_help.html#specific).
- ARC. (n.d.). *Repository name: Haverford College senior thesis archive*. DP9. Retrieved April 16, 2005, from <http://arc.cs.odu.edu:8080/dp9/listidentifiers/HaverfordCollegeThesis.OAI2/response0.html>.
- ARL, ACRL, SPARC. (2003). *Create change*. Retrieved Jan 25, 2005, from <http://www.arl.org/create/home.html>.

- ARL, ACRL, SPARC. (2000). *Scholars have lost control*. Create change. Retrieved Jan 25, 2005, from <http://www.createchange.org/faculty/issues/scholars.html>.
- ARL, ACRL, SPARC. (2000). *Scholars under siege*. Create change. Retrieved Jan 25, 2005, from <http://www.createchange.org/faculty/issues/silent.html>.
- Association of College and Research Libraries (ACRL). (2004). *About SPARC*. Retrieved Jan 25, 2005, from <http://www.arl.org/sparc/about/index.html>.
- Association of College and Research Libraries (ACRL). (2004). *Principles and strategies for the reform of scholarly communication*. White papers and reports. Retrieved Jan 26, 2005, from <http://www.ala.org/ala/acrl/acrlpubs/whitepapers/principlesstrategies.htm>.
- Association of College and Research Libraries (ACRL). (2005). *Scholarly communication*. Issues and advocacy. Retrieved Jan 25, 2005, from <http://www.ala.org/ala/acrl/acrlissues/scholarlycomm/scholcomminitiative.htm>.
- Association of College and Research Libraries (ACRL). (n.d.). *SPARC-scholarly publishing and academic research coalition*. Retrieved Jan 25, 2005, from <http://www.arl.org/sparc/>.
- Association of Research Libraries (ARL). (2005). *About SPARC open access newsletter*. SPARC. Retrieved Jan 25, 2005, from <http://www.arl.org/sparc/soa/index.html>.
- Association of Research Libraries (ARL). (2004). *Issues in scholarly communication*. Retrieved Jan 25, 2005, from <http://www.arl.org/scomm/index.html>.
- CogPrints: Cognitive Sciences Eprints Archive. (2004). *Welcome to Cogprints*. Retrieved December 5, 2004, from <http://cogprints.org/>.
- Cook, T. (1994, November). Electronic records, paper minds: The revolution in information management and archives in the post-custodial and post-modernist era. *Archives and Manuscripts* 22, 300-328.

Cook, T. (1992). Mind over matter: Towards a new theory of archival appraisal. In B. L. Craig (Ed.), *The Archivist imagination: Essays in honour of Hugh A. Taylor*, (pp. 38-70).

Ottawa: Association of Canadian Archivists.

Digital Library Federation. (n.d.). *Open Archives Initiative (OAI)*. Retrieved Jan 25, 2005, from <http://www.openarchives.org/>.

Digital Library Federation. (2001). *The Santa Fe Convention for the Open Archives Initiative*.

Open Archives Initiative. Retrieved Jan 25, 2005, from

[http://www.openarchives.org/meetings/SantaFe1999/sfc\\_entry.htm](http://www.openarchives.org/meetings/SantaFe1999/sfc_entry.htm).

Duranti, L. (1994). The concept of appraisal and archival theory. *American Archivist*, 57(2), 328-344.

Eprints.org. (2004). *Self-archiving and Open Access (OA) Eprint archives*. Retrieved December 5, 2004, from <http://www.eprints.org/>.

Eprints.org. (n.d.). *Citebase search*. Retrieved December 5, 2004, from

<http://citebase.eprints.org/cgi-bin/search?type=metadata>.

Eprints.org. (n.d.) *Eprints.org FAQ*. Retrieved April 24, 2004, from <http://www.eprints.org/self-faq/#serch-archiving>.

Eprints.org. (n.d.). *Institutional archives registry homepage*. Retrieved January 11, 2005, from

<http://archives.eprints.org/>.

Free Software Foundation. (2005). *Free software definition*. GNU Project. Retrieved Jan 25,

2005, from <http://www.gnu.org/philosophy/free-sw.html>.

Free Software Foundation. (2005). *GNU operating system*. GNU Project. Retrieved Jan 25, 2005,

from <http://www.gnu.org/>.

Gustafson, E. & Pitman, J. (2004, November). IMS journals on arXiv. *IMS Bulletin*. Retrieved Jan 25, 2005, from [http://stat-www.berkeley.edu/users/pitman/arxiv\\_article.html](http://stat-www.berkeley.edu/users/pitman/arxiv_article.html).

Harnad, S. (2001, May). The self-archiving initiative. *Nature Web Debates*. Retrieved December 4, 2004, from <http://www.nature.com/nature/debates/e-access/Articles/harnad.html>.

Harnad, S. (2001). The (refereed) literature-liberation movement. *The New Scientist*. Retrieved April 24, 2004 from <http://www.ecs.soton.ac.uk/~harnad/Temp/newscientist.htm>.

Harnad, S. (2000). *Re: Eprints Open Archive software*. Retrieved December 4, 2004, from <http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/1057.html>.

Harnad, S. (1994). Scholarly journals at the crossroads: A subversive proposal for electronic publishing. *ARL Issues in Scholarly Communication*. Retrieved April 30, 2004, from <http://www.arl.org/scomm/subversive/sub01.html>.

Harnad, S. (n.d.). *Resolving the anomaly*. Retrieved December 5, 2004, from <http://www.ecs.soton.ac.uk/~harnad/Tp/2-Resolving-the-Anomaly/sld001.htm>.

Harnad, S. (n.d.). *How to get around restrictive copyright legally*. Retrieved December 5, 2004, from <http://www.ecs.soton.ac.uk/~harnad/Tp/2-Resolving-the-Anomaly/sld007.htm>.

*Haverford College Senior Thesis Archive*. (2004). Retrieved December 4, 2004, from <http://thesis.haverford.edu/view/department/English.html>.

Haverford College Senior Thesis Archive. (n.d.). *User archive: Depositing records*. Retrieved January 11, 2005, from <http://thesis.haverford.edu/help/#Depositing>.

Hitchcock, S. (2005, January). *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies*. OpCit: The open citation project. Retrieved December 5, 2004, from <http://opcit.eprints.org/oacitation-biblio.html>.

- Hitchcock, S. (2003). *Core metalist of Open Access Eprint archives*. OpCit: The open citation project. Retrieved January 11, 2005, from <http://opcit.eprints.org/exploearchives.shtml>.
- Jenkinson, H. (1984). Reflections of an archivist. In M. Daniels & T. Walch (Eds.), *A modern archives reader: Basic readings on archival theory and practice* (pp. 15-21). Washington, D.C.: National Archives and Records Service. (Original work published 1944).
- Kyrillidou, M. (1999, June). Spending more for less.... *ARL Bimonthly Report on Research Library Issues and Actions 204*. Retrieved Mar 25, 2005, from <http://www.arl.org/newsltr/204/spending.html>.
- Lindberg, L. (2003). *Definitions assignment*. Unpublished manuscript. University of California, Los Angeles (UCLA), CA.
- Lindberg, L. (n.d.). Introduction 1 PowerPoint presentation. Retrieved Feb 1, 2005, from LIBR 256.01. Blackboard Web site. SLIS, San Jose State University.
- Lund University Libraries. (2004). *DOAJ-directory of open access journals*. Retrieved Jan 25, 2005, from <http://www.doaj.org/>.
- OpCit: The Open Citation Project. (2004). *Homepage*. Retrieved December 5, 2004, from <http://opcit.eprints.org/>.
- OpCit: The Open Citation Project. (2003). *Core metalist of open access eprints archives*. Retrieved April 16, 2005, from <http://opcit.eprints.org/exploearchives.shtml#eprintsorg>.
- OpCit: The Open Citation Project. (n.d.). *Registry of institutional archives home*. Eprints.org, Retrieved Jan 25, 2005, from <http://archives.eprints.org/>.
- OpCit: The Open Citation Project. (n.d.). *Self archiving FAQ*. Eprints.org, Retrieved Jan 26, 2005, from <http://www.eprints.org/self-faq/#serch-archiving>.

Open Archives Initiative (OAI). (2004). *OAI homepage*. Retrieved April 28, 2004, from <http://www.openarchives.org/index.html>.

Open Archives Initiative (OAI). (2004). *Protocol for metadata harvesting [OAI-PMH]*. Retrieved December 4, 2004, from <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

Open Archives Initiative(OAI). (2004). *OAI-PMH: Introduction*. Retrieved December 4, 2004, from <http://www.openarchives.org/OAI/openarchivesprotocol.html#Introduction>.

Open Archives Initiative (OAI). (2004). *OAI-PMH: Definitions and concepts*. Retrieved December 4, 2004, from <http://www.openarchives.org/OAI/openarchivesprotocol.html#DefinitionsConcepts>.

Open Archives Initiative (OAI). (n.d.). *Community*. Retrieved January 11, 2005, from <http://www.openarchives.org/community/index.html>.

Open Society Institute. (n.d.). *Budapest Open Access Initiative (BOAI)*. Retrieved Jan 25, 2005, from <http://www.soros.org/openaccess/index.shtml>.

Open Society Institute. (2002). *Budapest Open Access Initiative supported by the Open Society Institute's information program*. BOAI. Retrieved Mar 26, 2005, from <http://www.soros.org/openaccess/commitment.shtml>.

Open Society Institute. (2002). *Read the initiative*. Budapest Open Access Initiative (BOAI). Retrieved Jan 25, 2005, from <http://www.soros.org/openaccess/read.shtml>.

PLoS-Public Library of Science. (n.d.). *Homepage*. Retrieved Jan 25, 2005, from <http://www.plos.org/index.html>.

Postemsky, D. (2003). *Through the looking-glass: Reading and reflecting from wide Sargasso Sea to Jane Eyre*. Haverford College Senior Thesis Archive. Retrieved December 4, 2004, from <http://thesis.haverford.edu/archive/00000044/01/2003PotemskyD.doc.pdf>.

Psycholoquy. (2004). *Homepage*. Retrieved Dec. 5, 2004, from <http://psycprints.ecs.soton.ac.uk/>.

Schellenberg, T. R. (1956). *Modern archives: Principles and techniques*. Chicago: University of Chicago Press.

Suber, P. (2004, May 3). Two distractions. *SPARC Open Access Newsletter 73*. Retrieved Apr 15, 2005, from <http://www.earlham.edu/~peters/fos/newsletter/05-03-04.htm#oai-google>.

Suber, P. (2003, February 1) Removing barriers to research: An introduction to open access for librarians. *C&RL News 64*. In Back Issues 2003 [database online]. Retrieved Mar 25, 2005, from <http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2003/february1/removingbarriers.htm>.

Suber, P. (2004, November). FY05 omnibus appropriations conference report, NIH, Office of Director. *Mailing List SPARC-OAForum@arl.org Message #1317*. Retrieved December 4, 2004, from <https://mx2.arl.org/Lists/SPARC-OAForum/Message/1317.html>.

Suber, P. (2004) *The SPARC Open Access Newsletter*. Retrieved Mar 26, 2005, from <http://www.earlham.edu/~peters/fos/index.htm>.

Suber, P. (2002, April). Where does the free online scholarship movement stand today? *ARC Bimonthly Report*. Retrieved Mar 26, 2005, from <http://www.earlham.edu/~peters/writing/cortex.htm>.

Stallman, R. (2004). *Linux and the GNU project*. GNU Project. Retrieved December 4, 2004, from <http://www.gnu.org/gnu/linux-and-gnu.html>.

Tufts University Perseus Digital Library. (2004). *Perseus lookup tool*. Retrieved December 5, 2004, from <http://www.perseus.tufts.edu/cgi-bin/vor>.

United Nations (UN). (n.d.). *World summit on the information society (WSIS)*. Retrieved Jan 26, 2005, from <http://www.wsis-si.org/>.

University of Michigan Digital Library Production Service. (2004). *OAIster homepage*. Retrieved December 4, 2004, from <http://oaister.umdl.umich.edu/o/oaister/>.

University of Southampton. *CogPrints: Cognitive Sciences Eprints Archive*, [n.d.; cited 26 Jan. 2005]. Available from <http://cogprints.org/>.