

Public Policy Research in PolicyArchive: Use of DSpace When Metadata Meets Open Access

Buchanan, Sarah

Legislation at the state and national level is shaped through the conduct of research on the existing conditions, infrastructure, and community practice of specific areas of social inquiry. Public policy research, both within the academy and in government, serves a vital role in preparing legislators and policymakers to enact well-informed measures that are grounded in regional studies. Such work is becoming increasingly urgent as social issues and phenomena become more layered, complex, and interdisciplinary, and require more nuanced investigations. The results of this research, while essential at the time of need, have not historically been collected in a systematic manner that would enable access to the data beyond the immediate time period. A clear need has been expressed by researchers in many contexts who seek access to the rich content produced by local organizations, publishers, specialized institutes, and independent bodies of experts.

In the U.S., the ability to access federal information has been established through, among other measures, the Freedom of Information Act (orig. 1966). Access to state and local documents, often seen as a corollary right, is not specifically governed under a similar measure and as such, access rights have been realized through localized practices. State archive and library systems have led efforts to mobilize institutional and regional resources with which to provide access to their holdings.

While government documents are important primary source material, researchers also seek access to publications which have contextualized these legislative texts within a broader research agenda, and analyzed consequent measures passed over time. No piece of legislation is an island. Some organizations have focused their research around certain unifying subjects; perhaps instigated by a specific ruling but later sustained by considering the broader environment in which a condition exists. This research of public governance and policy has produced secondary sources; publications which illustrate the issue in depth and are the results of consultation with many sources of information.

PolicyArchive [<http://www.policyarchive.org>] is an open repository modeled on the principle that access to secondary source material - research reports, data, community analysis - enables more effective and accountable public policymaking. PolicyArchive is the first and

most comprehensive free, searchable archive of public policy research. The archive carries an emphasis on multi-source contributions, equitable access, and continuing outreach to new organizations. PolicyArchive remains established with the DSpace infrastructure, which supports several associated file formats and metadata content. It originated as a collaboration between a nonprofit organization and an academic library.

Publications, particularly those produced by local organizations, have historically faced significant institutional barriers in reaching the Web, and prior efforts have not been all-inclusive. Conversely, PolicyArchive fosters a broad, non-partisan scope of contributed content and allows content submissions directly from publishers and institutional representatives. The archive staff performs several activities which improve the remote user's experience toward selecting the appropriate subject and resources. This presentation will describe the functionality of this digital repository, which is identified by its subject breadth (306 subtopics on social studies) and depth. The archive provides persistent access to over 29,000 publications which are directly accessible through "permalinks" in concert with the Handle System. A preferred file format enables permanent and coherent storage, as well as compatibility with hard copy documents which can be scanned electronically. In addition to the main document described by the metadata, related files - called "bitstreams" - can be added to a document's page; e.g. associated tables, appendices, and bibliographies. Following initial upload of the document, metadata input is verified against established conventions; the author will discuss the respective usage of the DSpace automatic fields and specifically application of DCMI metadata schema, standardized organization name formats, and standardized place names. Adoption of these practices enables PolicyArchive to exchange data in a standardized format with libraries all over the world. The OAI-PMH standard, which allows metadata harvesting of host repository content via XML, is also available for future use. This is a standard that allows harvesting metadata (but not the files) in formats useful for editing and uploading into PolicyArchive.

Although organizations do report greater acceptance of the Web in general, as an information provider, they also observe that the current unregulated nature of information search does not easily lead researchers to their own organization's resources. As noted before, concentrated research may already have been conducted on just the topic of interest to a researcher - but without a central, browseable repository this research would remain hidden to the individual. Over time and with greater awareness of PolicyArchive, this repository will continue its reach into the global society of information seekers. Drawn from over 700 highly regarded national research organizations and publishers, electronic documents which represent

diverse policy perspectives are now available to private and public institutions, elected officials, foundations, and the news media. The structured metadata, described above, allows the reader to place a given document in context of its time of creation, overseeing body, and intended scope as provided by the author.

Another developing area within the archive which draws on the capabilities of DSpace is the assignment of a Special Collection: a portal page highlighted from the home page which serves as a gateway to contributors' research areas. A Special Collection allows an organization, working with PolicyArchive staff, to create a customized selection of works which represent a common theme or are centered around a programmatic goal. This innovation is a new area of development in the digital library community, which is looking at methods to highlight common interest and bring librarians' expertise to the digital realm: how to transform unmediated raw data into navigable, accessible learning resources. The author will discuss the current developments and applicability of this practice to similar digital initiatives across institutions.

American philanthropic foundations alone spend over \$1.5 billion per year on research. Would that even proportionate attention be paid to the sustainability and future accessibility of these research outputs, but PolicyArchive alone has crossed these institutional barriers. Encouragingly, several recent articles in the popular press have drawn attention to the issues of digital preservation and librarians' role in providing access to information - and a new wave of popular support may yet keep libraries and archival repositories connected to citizens' daily life. PolicyArchive has a demonstrable role in all stages of the public policy lifecycle: precedent research, topic coverage, creation, amendment, implementation, and transferability. Additionally it serves not only legislators and policymakers' information needs, but also those of academics who study public policy and administration, and who require access to past research which may not have since been replicated. The internet as it currently exists can deceive researchers who are unaware of deep databases of content which provide in-depth scholarly research. For any researcher conducting even cursory study of a societal issue, fact-finding such as that found through the "surface web" may provide a knowledge base, but it does not encourage the inquiry of new questions or point out gaping areas which are not online, and/or do not fall within the public domain. As others have stated, companies' commercial interests in Search appear to favor the popular opinion of the masses rather than the subject expertise of specialized researchers; this can lead to uninformed and shallow research. One tenet of information literacy is the ability to distinguish scholarly articles from personal opinion pieces, among which the

current Web environment presents no such distinction. To students of public policy at the undergraduate and graduate level, accessing trustworthy reference sources plays a crucial role in developing one's understanding of an issue. The development of metadata schema within the last few decades, and their implementation in PolicyArchive, all represent attempts to define information resources and provide structured access digitally through vocabularies and analysis of the subjects contained within them. Without this mediation, information would remain siloed in physical offices and inaccessible beyond a single geographic location.

Libraries, archives, and digital open repositories such as PolicyArchive, contain an inherent "added benefit" to society through their role as information interpreters and catalogers of the raw data (published or unpublished) which is in demand. While data in the form of keywords will enable retrieval in some situations, users also must be invested in their research well enough to be able to articulate and distinguish their interest area in the face of homonyms and related turns-of-phrase. Librarians and archivists' efforts in the cataloging and classification arenas have for decades enabled the conduct of focused research by applying standards and methods which help define the varied information resources at our disposal, including now digital content. As members of the professional information community, we recognize the benefits of the communal resources to which we contribute our energies, but also the constraints which guide our decisions and intake on a daily basis. A discussion of the continued development of PolicyArchive in context of the current information policy environment shared among the Open Repositories 2010 attendees - including digital preservation concerns, bibliographic control, and item- versus collection-level curation - will also be presented.