educational technology®

the magazine for managers of change in education

CONTENTS

Special Issue on Opening Educational Resources

Regular Features

educational technology ®

About This Issue

A special issue
on opening
educational
resources; plus
regular features

Open Content in Open Context

Sarah Whitcher Kansa
Eric C. Kansa
The Alexandria Archive Institute
Open Context Project

This article presents the challenges and rewards of sharing research content through a discussion of Open Context, a new open access data publication system for field sciences and museum collections. Open Context is the first data repository of its kind, allowing self-publication of research data, community commentary through tagging, and clear citation and stable hyperlinks, and Creative Commons licenses that make reusing content legal and easy.

The Challenges of Sharing Data

Advances in digital technology are transforming research and education. Field researchers, from archaeologists to environmental scientists, collect vast and growing amounts of multimedia content. Open Context is our initiative to make Internet publication of this research easy, open access, and free of charge. Our goal is to see this digital material shared globally, without barriers and, ideally, in real time. With Open Context and other emerging technologies, researchers, educators, and students at all levels can enjoy unprecedented levels of access to ongoing research and, perhaps more importantly, to access and communicate with the researchers themselves.

Why are we focusing on field research, and what does this have to do with education? The answer is simple: Field research gives students a window into other worlds. These other worlds may be remote time periods and ancient cultures, or they may be endangered habitats that may disappear before today's youth

ever get a chance to explore them first-hand. Finding cost-effective strategies for publishing field research is critical if we want to give today's generation and future generations a chance to learn about our history and environment. Field researchers collect a vast amount of digital information but currently lack good ways to publish and preserve these irreplaceable observations. In too many cases, our knowledge of the past and the natural world is one hard-drive crash away from irrevocable loss.

Fortunately, the Internet now provides the capacity to publish all of this vulnerable and irreplaceable content. However, given the complexity of the field sciences, online data publication is easier said than done. Tools and incentives are needed to rescue field research from loss. The Alexandria Archive Institute (AAI) was formed to address these problems so that researchers, educators, and students can explore and understand our natural and cultural heritage. The AAI builds free, open source tools to publish field science research on the Internet. The AAI also works to understand and structure incentives so that researchers will be willing to participate in a culture of openness, and share their materials in ways that enable others to build upon their work. More specifically, our work addresses the technical and cultural challenges of open access to research content.

Technical Challenges to Data Sharing

One of the greatest hurdles to sharing data online is the great diversity of content generated in most disciplines. Databases tend to be project-specific, posing a challenge to data sharing, even within a narrow field. Large project databases often include tens of thousands of individual records created by multidisciplinary teams, all in complex relationships. If a dataset needs to be downloaded and deployed on specific software, it may be difficult to use even with adequate documentation. Once deployed, users will have to familiarize themselves with a project's database organization and interface. The steps involved in downloading and deploying such databases require too much time and expertise for the average user.

To attract and keep users, Internet-based data sharing systems should be designed so that they appear immediately familiar and accessible to first-time users. They should be easy to navigate and not require any additional steps for use (such as downloading plug-ins or creating password-protected accounts). They should also provide apparent and immediate rewards for data contributors, such as clear authorship, editing functions, and search engine exposure. Many well-funded projects have custom designed Websites that provide access to all or part of their research. However, customization and database design are time-consuming and costly. A generalized solution for sharing and

---

Sarah Whitcher Kansa and **Eric C. Kansa** are co-founders of the Alexandria Archive Institute, a California-based nonprofit working to create tools for scholars to share open content. With support from The William and Flora Hewlett Foundation and individual donors, they have created Open Context, an open access, Web-based system for publishing cultural heritage content from the field sciences and museum collections (e-mails: skansa@alexandriaarchive.org; ekansa@alexandriaarchive.org; Websites: *www.alexandriaarchive.org*; *www.opencontext.org*).

exploring data across multiple projects would be much more effective than customized, project-specific systems.

Even if data sharing systems are simple to use, many researchers are reluctant to publish data that they perceive as too small, too incomplete, or too "messy" to share with the world. Smaller projects, in particular, suffer from information loss because they have little capacity to develop customized solutions on their own. They may generate rich bodies of documentation, but without Internet dissemination, much of this material will never see publication because paper simply isn't up to the task. Data sharing systems must be flexible to accommodate all content generated in projects large and small, complete and incomplete.

Scholars sharing their research via the Internet need assurance that it will be preserved in its original form. Too much research is vulnerable to loss through accidents or neglect. Researchers need easy tools to publish their digital materials in open file formats (that can be read by free software and are more likely to stand the test of time). They also need help migrating their materials to digital archives. Thus, open content systems should include simple migration and archiving options.

Cultural Barriers to Data Sharing

In our experience creating Open Context, the most frequently voiced concerns over data sharing are the perceived loss of professionalism (peer review) and authorship. "Going online" does not mean abandoning peer-review and adopting the radical egalitarianism of the *Wikipedia*. Peer review can be built into open digital dissemination and is already successfully in place in many open access journals (Harnad & Brody, 2004).

Professionalism also requires proper attribution of an author's work, and many scholars worry that open access will cause their content to be "scooped" by someone else. Clear and recognized forms of citation will make researchers more comfortable with online data publication. Systems should have clearly marked authorship and citation information for every piece of content they contain. Other features such as time-stamps on contributions, logos, and search engine indexing, make original contributors easy to recognize and such exposure would deter misuse.

Open Context: An Open Access Data Publishing Solution

In 2005, with a series of grants from the William and Flora Hewlett Foundation, we set out to create a system that faced these challenges. The most important requirements for this system were the following:

1. It had to be easy to navigate for the "average" computer-literate user.

2. It had to accommodate diverse and non-standardized projects.

3. It had to be scalable (generalized to apply to many different projects without requiring expensive customization).

4. It had to be academically robust (peer-reviewed and citable).

The resulting system, called Open Context (*www.opencontext.org*), is an online, open access publishing system, developed to support enhanced scholarly communication in archaeology and related disciplines. Open Context enables researchers to publish structured data along with textual narratives and media (images, maps, drawings, videos) on the Web. This new system provides a cost-effective and scalable solution to the current problems of data loss and limitations to data sharing.

How Open Context Works

Open Context is especially effective for publishing large bodies of complex documentation, where diverse datasets can be linked together in an integrated data publication system. It is also designed to accommodate idiosyncratic data from small or incomplete projects, thereby giving a home to content that might otherwise never see publication. Open Context is a cost-effective and powerful means to share large collections of rich media and complex data (see Figures 1 and 2). Contributors to Open Context use the system's Web-based data import application (called Penelope) to upload and edit their own data, before submitting it for peer-review and online publication. No one but the contributor can edit an original item; however, anyone can make comments and links to items through tagging. In this way, the user community draws linkages between items across different projects. With this streamlined data publication process, a dataset that takes months (or even years) to develop can be published in a matter of minutes in a system that gives clear credit to the author. Because Open Context draws on powerful open source technologies, it can be easily adopted by others and maintained by anyone with standard Web development skills. That is, users can either explore Open Context as it is, or they can adopt the open source code to use for their specific data sharing needs.

Open Context currently features the following tools for scholars and students to easily share, preserve, and reuse educational materials:

- **Web-Based Data Publication:** Penelope, Open Context's Web-based publishing tool (see Figure 3), enables individuals to publish their own datasets and make them interoperable with other projects and collections in Open Context. With Penelope, data contributors upload and edit all content within the application, without having to

Figure 1. Results of a search for "carnelian," showing items from multiple projects.

Figure 2. An image linked with its small-finds registry record and context.

Figure 3. A view of Penelope, Open Context's Web application for data publication.

download any special software or viewers.

- **Flexible Design:** Open Context acts as a (near) "universal recipient" for the vast variety of information generated by field research. The system can accept and integrate diverse content from a huge project such as ten years of excavation drawings, photos, small-finds databases, maps, and specialist analyses from an archaeological site. It can also house very small projects that might consist of only a few field notes and photographs, projects that are at risk of loss without a simple and cost-effective means of publication.

- **Simple Tools for Use:** Open Context's interface offers straightforward browse, search, and analysis functions. Users have a variety of options to find materials in Open Context, including simple, "Google-like" text searches to more sophisticated, advanced searches that use complex query logic. Simple charting tools help with data visualization, and a selected dataset (potentially drawing records from multiple projects) can be exported into common formats, such as *Excel.*

- **Data to Knowledge:** By offering comprehensive access to highly structured research and collections data, Open Context supports reanalysis and reinterpretation of research results. Students can use Open Context as a primary source to develop important analytical skills by exploring primary data.

- **Citation and Easy Retrieval:** Stable URL links are attached to *each and every piece of content*, so items can be referenced in paper or e-publications and retrieved on the Web. Automatic generation of citations for each item enables scholarly use and promotes good scholarly practice (see Figure 4).

- **Peer Review:** Open Context, like most open access publication systems, retains professional editorial control over content. It only accepts content from permitted research and recognized collections, and all content is subject to professional editorial review.

- **Community Tagging:** Open Context offers a "folksonomy" tool to let others annotate and identify materials of specific interest (Bearman & Trant 2005). Users can tag items either individually or collectively (i.e., users can assign a tag to items in a query result set). Tagging can serve research interests, or may identify a set of material useful for a specific course, problem set, or even grade level. The folksonomy tool is a very simple way to help make complex databases easy to use for many different communities of users.

- **Comments and Community:** Open Context pro-

Figure 4. Automatically-generated citation for an Open Context item.

vides feedback tools, such as a "ping-back," which keeps track of external links to the data in the system. For example, if someone links to a set of items in Open Context, the system would automatically recognize the link and, upon administrative approval, set up a reciprocal link. In this way, the ping-back tool is useful for linking Open Context content with external analyses, interpretations, and comments.

- **Sharing Updates:** Open Context broadcasts a number of RSS ("Really Simple Syndication") feeds for content contributors to share news and updates across the Internet. For example, if a project director adds a series of photos to an existing project, the system will automatically generate an announcement of the update.
- **Preservation:** The Internet Archive, a leading digital repository, helps safeguard Open Context materials. The Internet Archive offers this invaluable service for free, so long as copyright permissions are granted for others to use and reuse the content (copyright is discussed further below).

Continuing Development and the Power of Peer Production

Future development plans for Open Context include customization tools for data contributors to "brand" their content with logos and styles, further clarifying authorship. Customization will also help users or groups of users to target information specific to their research or instructional interests, thereby organizing the system into overlapping, user-defined sub-groups. These tools would allow teachers to customize modules for students to work with during the course of a semester. Modules would only draw from specific Open Context content selected by the teacher. Students would both gain first-hand experience working with a focused body of primary data and contribute to the research process by tagging items or publishing the results of their reuse of the data in Open Context.

Researchers publish, review, and edit materials as part of their academic activities. Open Context streamlines all of these processes and distributes the job of data publication through "peer production." Peer production is how the *Wikipedia* and open source software are generated. Participants in peer production, motivated by social returns, contribute small pieces to a larger whole, often resulting in very sophisticated and high-quality outputs (Benkler, 2006). Leveraging peer production and distributing the work of data publication and software updates is an important way to keep costs down for Open Context and open educational resources more generally.

What Makes Open Context "Open"?

"Openness" is much more than simply not charging for access. Even if a resource is free to view, the legal default of "all rights reserved" copyright makes it very difficult for anyone to do anything with the content. Copyright law forbids nearly all copying and use of

material to make new "derivative" works. This runs counter to the mission of research, which is to build upon the works of others. It also runs counter to many exciting instructional opportunities, including having students act as active participants in knowledge creation. Thus, to make Open Context content valuable to the community, we must provide needed copyright permissions. Each item in Open Context is licensed with a Creative Commons copyright license, which gives explicit permissions to copy and use the material so long as users properly attribute the source.

Creative Commons licenses also offer certain optional conditions, such as restrictions against commercial uses. Unlike the public domain, creators still retain their copyright when they use Creative Commons. However, creators are able to grant permissions to encourage research, education, and many forms of creativity while retaining their right to be attributed for their contributions. Creative Commons licenses now see wide implementation in many leading scholarly Web resources. The prestigious journals published by the Public Library of Science (with impact factors rivaling *Nature*, *Science*, and *Cell*) use Creative Commons licenses to clarify terms and permissions for each of their scientific articles. This openness and flexibility ensures that Open Context content is of maximum value for reuse in both instructional and research applications.

Open Access for Educational Reuse

Open Context and related open access systems can forge new paths to understanding by making research easier to find and use and giving that research more impact and significance. As the user community grows, their contributions (in the form of published open content, tagged sets, and other links) will work to expedite searches and facilitate reuse of content. Open access to this digital content means that students and scholars who use search engines to do research will be more likely to find primary, scholarly data and relevant syntheses linked to those data. Thus, the entire research process is made transparent, from primary data collected in the field to published syntheses and finally to further re-mixes by others (including students). This will create a virtuous cycle of creative reuse of high-quality research material and help draw students into the process as active participants, rather than passive observers. ☐

References

Bearman, David, and Jennifer Trant. (2005). Social Terminology Enhancement Through Vernacular Engagement: Exploring Collaborative Annotation to Encourage Interaction with Museum Collections. *D-Lib Magazine 11*(9); <http://www.dlib.org/dlib/september05/bearman/09bearman.html > .

Benkler, Yochai. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven, CT; full text at: *http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf>* .

Harnad, Stevan, and Tim Brody. (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine 10*(6); *<http://dlib.org/dlib/june04/harnad/06harnad.html>* .

Author Guidelines for Magazine Articles

In preparing an article for *Educational Technology* Magazine the primary fact to keep in mind is that this magazine is not a formal research journal. It is, as the name implies, a magazine. The Editors are looking generally for articles which interpret research and/or practical applications of scientific knowledge in education and training environments.

Thus, your article should not be cast in the form of a traditional research report. The facts of your research, or that of others, should be stated succinctly. Then you should go on to explain the implications of this research, how it can be applied in actual practice, and what suggestions can be made to school administrators, trainers, designers, and others.

The style of writing should be on the informal side—an essay—since once again this is a magazine and not a formal academic journal. Authors are free to state their opinions, as long as the opinions are clearly identified as such. The use of specialized jargon should be kept to a minimum, since this magazine has a very wide interdisciplinary audience.

There are no minimum and maximum length restrictions. Make your article as short as possible to do the job you intend. As a general rule, most articles are about 3,000 words. Include graphics as appropriate.

Note too that this magazine is read in more than 100 countries, by persons holding prominent and influential positions. They expect a very high level of discourse, and it is our goal to provide major articles of excellence and lasting significance.