ISQ INFORMATION STANDARDS QUARTERLY

SUMMER 2010 | VOL 22 | ISSUE 3 | ISSN 1041-0031

SPECIAL ISSUE: ENHANCED JOURNAL ARTICLES

ARTICLE-LEVEL ENHANCEMENTS IN THE HUMANITIES AND SOCIAL SCIENCES

HOSTING SUPPLEMENTARY MATERIAL

ARCHIVING SUPPLEMENTAL MATERIALS

DOIS FOR JOURNALS: LINKING AND BEYOND

NISO
How the information world CONNECTS

For restaurant recommendations* you have your choice of printed or web guides galore.
But, for scholarly articles, there is only one source: bX™, the Ex Libris Scholarly Recommender Service.

Benefit from Users' Collective Wisdom:
the Ex Libris bX™ Scholarly Recommender Service

Facing a rapidly expanding universe of online information, scholars are in need of new tools to help them find and assess research and study material in a manner that incorporates immediacy, relevance, and a standardized measure of quality.

The bX™ Scholarly Recommender is a new service that taps into the power of the networked scholarly community to generate recommendations based on article usage. Brought to you by members of the same team that created SFX® and the OpenURL standard, it represents the growing recognition of the importance of user-driven content and an important step in the convergence of Web 2.0 and the scholarly world. bX is offered as an on-demand service. bX recommendations are delivered directly to your library's system, so users start seeing and benefiting from recommendations immediately!

NOW OVER 700 SITES STRONG!

ExLibris bX

You can have bX up and running in a matter of hours.
For more information on bX, visit
http://www.exlibrisgroup.com/category/bXOverview.

Ex Libris Group Toll Free: 1-800-762-6300 · Email: infousa@exlibrisgroup.com · www.exlibrisgroup.com

*CHOW.com and Chowhound.com are registered trademarks of CBS Interactive, Inc.

ISQ

SUMMER 2010 | VOL 22 | ISSUE 3 | ISSN 1041-0031

NISO
How the information world CONNECTS

INFORMATION STANDARDS QUARTERLY (ISQ) is a publication by the National Information Standards Organization (NISO). ISQ is NISO's print and electronic magazine for communicating standards-based technology and best practices in library, publishing, and information technology, particularly where these three areas overlap. ISQ reports on the progress of active developments and also on implementations, case studies, and best practices that show potentially replicable efforts.

NISO MANAGING DIRECTOR and ISQ PUBLISHER | Todd Carpenter
ISQ MANAGING EDITOR | Cynthia Hodgson
NISO STANDARDS PROGRAM MANAGER | Karen A. Wetzel
NISO BUSINESS DEVELOPMENT & OPERATIONS MANAGER | Victoria Kinnear
DESIGN | B. Creative Group, Inc.

CONTENTS

Kristen Fisher Ratan

FROM THE **GUEST CONTENT EDITOR**

Scholarly publishing started in 1665 with the Royal Society's *Philosophical Transactions*. The earliest articles were reprinted letters between scientists. In those early days, articles were primarily text, only occasionally enhanced with hand-drawn figures. Over the following three centuries text, figures, and tables were typeset but little else really changed. Further enhancements to articles included author affiliations, author-supplied keywords, and reference sections. In the 20th century, color images and complex equations were considered cutting edge.

2: Observables upon a Monstrous Head *Phil. Trans. 1665 1:85-86; doi:10.1098/rstl.1665.0037*

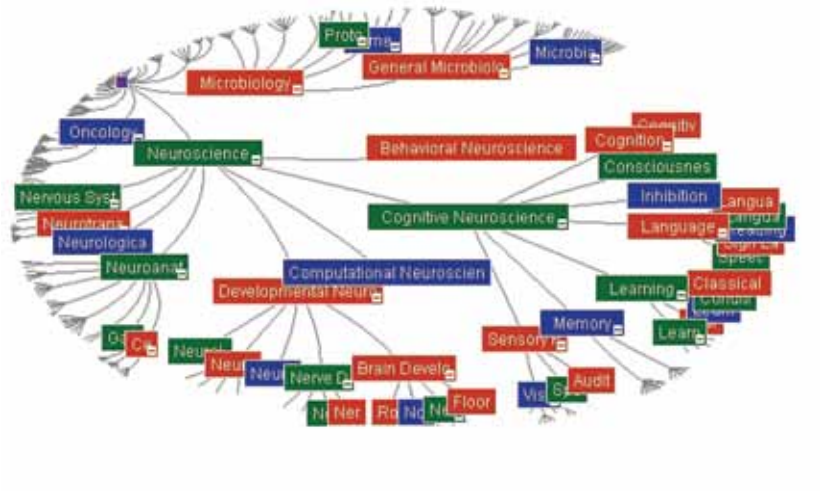
As content went online in the mid 1990s, more enhancements were added, crowding the space around the article. The web-enabled hyperlinking and this layer of added intelligence through HTML coding remains one of the most important enrichments to articles. Different versions of the article became available through digital content services. Articles were now searchable and indexed by search engines, with links to articles coming in from all directions. Data supplements began appearing, linking readers to the supporting data behind the published record. The first request HighWire received to post a data supplement was in 1998. Interestingly, it was a video, one of the fastest-growing forms of article enhancements now.

Soon, publishers were adding article-level services that enhanced the reading experience: related content, forward links to articles that have cited the current one, taxonomic tagging, links to external databases, alerts, and more.

Today we are seeing podcasts, pubcasts, and video versions of articles appearing on the publisher's site or on third party sites like SciVee, which hosts video versions of content and connects back to the article. Social bookmarking has taken the place of e-mailing services and there is a lot of buzz about the potential of the semantic web.

While many article enhancements enrich the content, what will prove even more important is what can be done with these enhancements. Having access to a data supplements is helpful, but large-scale mining of, utilizing, and visualizing data supplements may prove far more interesting. DOIs for components of articles help to provide links, but also provide metadata that is layered on at the component level, allowing for cross-publication collections of article components to be created on the fly.

The connections between articles and their enhancements could bring us closer to what's referred to as the semantic web, a more intelligently connected web of information. Semantic tagging is one way to do that. Extracting and connecting similar elements and metadata is another.



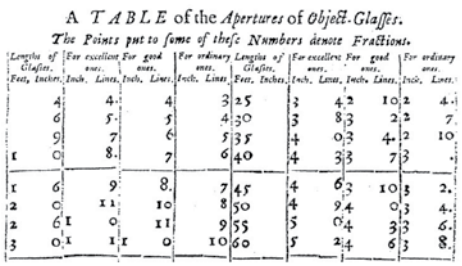
The connections between articles and their enhancements could bring us closer to what's referred to as the semantic web, a more intelligently connected web of information. Semantic tagging is one way to do that. Extracting and connecting similar elements and metadata is another.

This issue of ISQ focuses on the current challenges and opportunities surrounding different article enhancements, with particular focus on supplemental materials. Dean Smith and Wendy Queen review some of the changes that have occurred in enhancing journal articles, particularly in the humanities and social sciences. They describe Project MUSE and how it is supporting multimedia materials. Andrea Laue provides the perspective of a hosting service and the challenges involved in hosting supplemental material. She offers suggestions on how publishers and hosting services could standardize on the format and processing of supplementary content. The third feature perspective comes from David SH Rosenthal and Vicky Reich who address the issues in archiving and preserving supplemental materials and identify four areas that could reduce the costs involved in ensuring long-term access to these article enhancements.

NISO and NFAIS have jointly sponsored a new initiative to develop standards and best practices for supplemental materials. Linda Beebe, one of the co-chairs of the new group, describes the pre-work roundtable that led to the initiative and their future work plans. A survey conducted by Sasha Schwarzman was the impetus for the roundtable; an extract of his survey report illustrates the issues that publishers are encountering.

Our spotlight this issue is on the Digital Object Identifier (DOI) standard that is already heavily used to identify journal articles and provide cross-platform and cross-publisher linking. Patricia Feeney discusses new developments in how the DOI can be used beyond the common article linking, including using DOIs for datasets and other supplementary materials.

There are few standards or best practices for handling journal article enhancements today and the existing standards are not well-integrated when it comes to supplemental materials. But as this issue of ISQ shows, that situation is rapidly changing. doi: 10.3789/isqv22n3.2010.01


A TABLE of the Apertures of Object-Glasses.
The Points put to some of these Numbers denote Fractions.

Lengths of Glasses. Feet, Inches.	For excellent ones. Inch. Lines.	For good ones. Inch. Lines.	For ordinary ones. Inch. Lines.	Lengths of Glasses. Feet, Inches.	For excellent ones. Inch. Lines.	For good ones. Inch. Lines.	For ordinary ones. Inch. Lines.
4	4.	4	3	25	3	4,2	10	2	4.
6	5.	5	4	30	3	8,3	2	2	7
9	7	6	5	35	4	0,3	4.	2	10
1 0	8.	7	6	40	4	3,3	7	3	.
1 6	9	8.	7	45	4	6,3	10	3	2.
2 0	11	10	8	50	4	9,4	0	3	4.
2 6	1 0	11	9	55	5	0,4	3	3	6.
3 0	1 1	0	10	60	5	2,4	6	3	8.

1: Monsieur Auzout's Judgment Touching the Apertures of Object-Glasses, and Their Proportions, in Respect of the Several Lengths of Telescopes Phil. Trans. 1665 1:55-56; doi:10.1098/rstl.1665.0027

Kristen Fisher Ratan | *Assistant Director, Business Strategy, HighWire Press*

In the mid-to-late 1990s, electronic publishing visionaries held ambitious dreams. They predicted the end of the print journal and the emergence of virtual ones. Journal brands would eventually disappear and article bundles would take their place. Portals were the rage. Building community was a piece of cake. Scientific breakthroughs would happen faster with electronic journals. Ideas raced across the information superhighway.

DEAN SMITH AND WENDY QUEEN

ENGAGING THE READER

ARTICLE-LEVEL ENHANCEMENTS IN THE HUMANITIES AND SOCIAL SCIENCES

The scholarly journal and its collection of articles in a given area provided the ideal format for online delivery. Its time had come. Newly minted webmasters scoffed at PDF articles and became zealots for HTML and JavaScript. It was the beginning of a transitional phase in scholarly publishing that is still progressing today—and the pace is picking up. Early advances in electronic journal publishing came from the STM publishers whose substantial resources and sense of innovation provided a blueprint for publishers in the humanities and social sciences.

Experiments such as the TULIP Project, Red Sage, and the work of individual publishers led the digital transition from print to web publication and drove early enhancements at the article level. Speed to publication, especially in the scientific disciplines became an industry differentiator. Initiatives undertaken to publish articles in advance of print such as ASAP Articles (As Soon As Publishable), first developed by the American Chemical Society in 1998, positioned the web article as the version of record—elevating it to prominence as distinct from printed copies of a journal article.

Large STM publishers such as Elsevier, Springer, and Wiley reinvented their journal collections by creating integrated platforms with marketing-driven titles such as Science Direct, LINK, and InterScience respectively. Magazines such as Nature and Science leveraged their strong brands to build innovative and dynamic web offerings. Society publishers such as ACS, AIP, and IEEE built engaging platforms for electronic delivery. Project MUSE and HighWire played a leadership role in assisting smaller publishers in other disciplines with lesser resources.

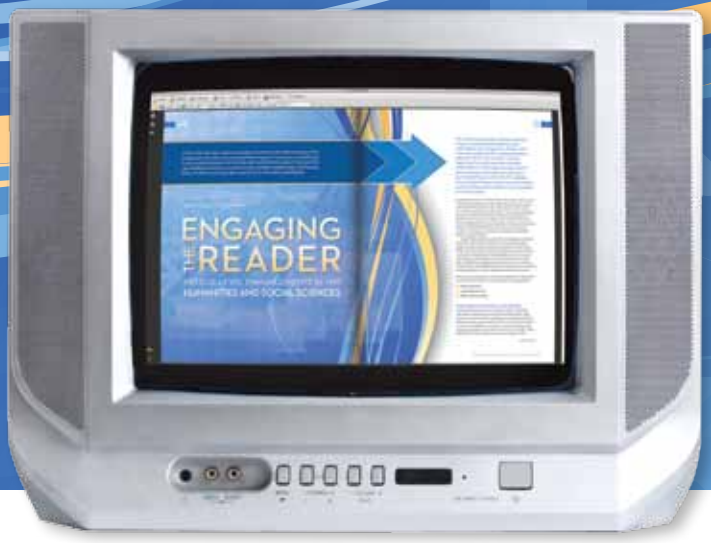
Three major developments over the last decade have shaped the environment for article level enhancement and engagement:

1. **linking initiatives,**
2. **usage statistics, and**
3. **Web 2.0 technologies.**

Usage Stats Focus Attention on the Article

A groundbreaking initiative launched in 2002, COUNTER provided an international standard for measuring the usage of electronic journals and it established a level playing field for libraries to analyze and establish value. It also focused attention on the article. Publishers were able to analyze referring traffic and measure pathways into their content such as Google, Yahoo, EBSCOHost, Scopus, Web of Science, and ProQuest. They

CONTINUED »



At the article level, end users voted with their keyboards and made PDF the predominant format. It was counter-intuitive to the medium of the web—liberating and free—to have an electronic version of the print facsimile, described by one publisher as "the pages of the novel on television."

CONTINUED »

could also measure the length of individual user sessions—emphasizing engagement. The possibility of digging deeper into the end-user mystery presented intriguing possibilities for publishers. We ended up measuring the "meat and potatoes"—article downloads, page views, and comparing HTML and PDF.

Standardized usage statistics presented a quantitative picture of article usage but we've only scratched the surface when it comes to understanding user behavior. COUNTER drove the evolution of metrics and fostered a yearning for more. Librarians now had the ammunition for cancellations in a way never before realized. Price per article download joined impact factors in the discussion and publishers devoted their energies to driving usage. New positions were created that focused on "search-engine optimization," "user behavior" and "business intelligence."

Article as End Game...and Beginning

Reference-linking initiatives such as CrossRef and secondary databases such as Web of Science and CAS created robust systems of connections into and out of the full text across publishers. Google joined the fray with Google Scholar, creating a focused search environment for researchers. Coupled with their expansive vision for content digitization, this new scholarly search interface helped position Google as a leading tool for search and discovery of scholarly information. Subscription agents EBSCO and SWETS entered the content aggregation game. The fully searchable abstract gained prominence as the "snippet" or "gist" of content designed to draw readers into the full-text. Once there, publishers began working on ways to expand the reader's options and envisioning the article as the launching point for discovery.

"In the future, we will look back at linking as the beginning of this idea of 'article as gateway'," said Judy Luther, President of Informed Strategies.

In the trenches at the article level, end users voted with their keyboards and made PDF the predominant format. It was counter-intuitive to the medium of the web—liberating and free—to have an electronic version of the print facsimile, described by one publisher as "the pages of the novel on television." We went deep into the labyrinths of our legacy content databases and digitized our back files. Our R&D functions no longer had time to ponder the future of the journal and innovation; they were too busy feeding the online production stream. As we looked at our platforms after the turn of the century, we were somewhat underwhelmed. We'd created PDF PEZ® dispensers. What happened to the community we were building?

"What we have innovated on is primarily "distribution" (i.e., using a website and a network and a printer to move what is basically the same information), not the content of what is being communicated," said John Sack, Director of HighWire.

Web 2.0 Technologies Impact Article-level Engagement

Web 2.0 technologies and XML provided the inspiration for enhancements at the article level and breathed new life into the notion of enhancing discoverability, driving usage, and building community. There was talk of open peer review. Publishers started building digital end-to-end workflows. They purchased content management systems. Functionality came back into vogue. Beta sites for such prestigious publications as the NEJM and JACS started experimenting with widgets and applications found in consumer environments such as Amazon and iTunes. Blogs, wikis, podcasts, citation management programs, and RSS feeds enabled publishers to push article content out. Image galleries and cover displays were added to journal sites to engage a new generation of researchers. The visual aspects of content and design merged. The GUI, pronounced "gooey," or "Graphical User Interface" took precedence for the first time.

"More Like This" functionality as seen on Amazon started appearing on scholarly journal sites. Publishers

became committed to user interface design and testing of their websites. The "drip-castle" approach to building platforms required a more robust commitment to information architecture and more cash. Collaboration, connecting, and sharing established themselves as core values for researchers. Websites addressed the constituent-based needs of authors, researchers, end users, librarians, and publishers with enhancements and resources delineated for each group.

The concept of "My Journal" and the idea of the semantic web provided a glimpse into what promises to be an exciting future for article-level enhancement. Adding new features to keep space with the speed of technology is what is needed to capture mindshare and remain relevant.

"Perhaps the most surprising single enhancement that just 'took off' was download to PowerPoint. When we were first developing it, we thought, people just Save Image As, or copy/paste into PowerPoint, so this won't be a big deal. We were wrong," said Sack.

Enhancing Articles in the Humanities and Social Sciences

Significant differences exist in the digital evolution from print to web between STM journals and those in the humanities and social sciences. Published in February 2009, Mary Waltham's study *The Future of Scholarly Journals Publishing Among Social Science and Humanities Associations* identified some of these characteristics. Rejection rates and longer articles contribute to higher costs for journals in the humanities and social sciences when compared to those in STM. University Press publishers lack resources to build platforms and print versions remained popular.

"A survey of Perspectives on History readers conducted in 2008 found that 63 percent read only the print copies of our serial publications," wrote Robert Townsend, Assistant Director of the American Historical Association, on his blog.

University of Chicago and the University of California built platforms for their journal content. JSTOR served the academic community as an archival resource for libraries. A unique collaboration between a publisher and a library to provide not-for-profit publishers with a platform for their journals and libraries with a cost-effective solution to acquiring scholarly research, the Johns Hopkins University Press and the Milton S. Eisenhower Library launched Project MUSE in 1995.

"The cost of journals was going up higher faster than anything else and those cost increases were squeezing our ability to buy books," said librarian Scott Bennett, who teamed with Press Director Jack Goellner to develop the original concept for Project MUSE.

Additional publishers joined MUSE in 2000, including Indiana University Press, MIT, University of Toronto, University of Nebraska, and the University of North Carolina. The MUSE publishing model delivers a sizable share —70 to 80%—of its subscription revenues back to publishers while providing a large collection of journals to libraries at a minimal cost per title. Delivering more than $70 million to publishers since 2000, Project MUSE is celebrating its 15th year and has demonstrated a track record for success, transparency, and excellence in providing high-quality scholarly content from not-for-profit publishers to the academic community.

Balancing the interests of 114 publishers, 450 journal communities, and millions of end users around the world, Project MUSE has achieved outstanding results by creating a robust digital environment for its readers. A new XML production workflow, Web 2.0 technologies, and social networking tools influenced the reinvention of the Project MUSE website in 2007. The focus of the redesign was to give the end user more tools at the article level with the goal of increasing the amount of time a user spends on the site. In addition to a more flexible layout, social bookmarking links, journal alerts, and linked subject headings were added for launch.

Immediately, the redesign received positive feedback from subscribers and end users. In addition to the enthusiasm, usage increased which inspired the MUSE staff to adopt a feature-per-month program. The program aims to incorporate recommendations from users as well as internal staff. The program, while ambitious, has provided MUSE the opportunity to have a fresh, constantly evolving site while collecting additional end-user behavior statistics. The statistics have provided the metrics to further learn what tools are widely used and how users are engaging with the content. The first year of the program produced a wide-variety of new discovery tools (see sidebar).

Project MUSE also adds value at the article level through an indexing process by providing a controlled vocabulary and name authority recognition. The indexing is contributed by an internal staff of six professional librarians. As a result, MUSE provides clickable subject headings on every article, as well as table of contents and search results, which bring together topics and subjects in a more precise fashion. These are also used in the "More Like This" algorithm and in search relevance ranking.

Overall, the program and article level enhancements have been a success. The creation of additional functionality at the article level has increased the amount of time a user spends on the site. The average session length has increased from 1.25 articles per session in 2008 to 2 articles in the 1st half of 2010.

Project MUSE plans to expand the program through 2011 by enhancing existing functionality, improving alerting technologies, and adding tools that increase awareness in specific subject areas. In addition, MUSE will continue to expand the platform and tools to create even greater connections between content and formats. The usage metrics have provided the basis for a forward-thinking blueprint to embrace innovation and remain relevant. As the program continues, the aim is to continue to increase the session time and overall usage.

CONTINUED »

Project MUSE

Feature-Per-Month Discovery Tools:

- Journal Search - ability to search within a specific journal
- More by Author - ability to search within MUSE, Google, and WorldCat
- Summary Pages - highlighting abstracts and Library of Congress Subject Headings
- Frequently Downloaded Articles browse at the journal level
- Frequently Downloaded Articles browse at the site level
- Tweet This option
- More Like This
- Subject Browse through the exploration of Library of Congress Subject Headings
- Reference-linking
- RSS feeds for Table of Contents
- Social bookmarking tools
- PDF covers - enhancing the platform and journal branding while providing links to additional content

CONTINUED »

"Increasingly the value of content is being influenced by its functionality," said Judy Luther. "Providing an array of tools and content types including video and datasets for additional exploration approaches this idea of 'article as gateway'."

Thinking forward, MUSE plans to further expose its metadata and create an opportunity to develop even richer partnerships within the research community. MUSE is also exploring user-contributed data—evoking participation beyond the basics of social media, i.e., Twitter and Facebook.

MUSE & Multimedia

Project MUSE has been supporting articles with audio and video since the late 90s. Journals like *Postmodern Culture* and *Advertising & Society Review* have enhanced the reader's experience by embedding multimedia. Some of the most widely used articles in MUSE have embraced the use of interactive content. As publishers conceptualize new approaches to constructing an article enriched by multimedia, MUSE strives to provide a state-of-the-art technical approach. *Advertising & Society Review* and its digital textbook companion, *ADText*, published by the Advertising Educational Foundation (AEF), include video examples and maximize the use of technology: video, audio, rich graphics, and other dynamic content.

"ADText: An Online Curriculum is distributed together with Advertising & Society Review on Project MUSE," said Paula Alex, CEO of the AEF. "ADText is a 20-unit living textbook about communications, marketing, and advertising. It is the first comprehensive digital textbook about advertising and its impact on society."

Through these publications, the AEF has built an interactive bridge between the advertising industry and the halls of academe where the cultural position of Madison Avenue is examined.

ADText is about to add another innovative feature at the article level, an audio interpretation by paragraph in Spanish (available in the Fall) and Mandarin Chinese (in 2011) to facilitate usage and understanding around the world.

"TV commercials and print ads are embedded in the text. It's completely innovative from a teaching and education point of view," said William M. O'Barr, Ph.D., Duke University, author of ADText.

MUSE publishers continue to express an interest in a more dynamic community-based web presence. In 2011, Indiana University Press will launch a new multimedia journal, *African Conflict and Peacebuilding Review*, which will once again alter the traditional journal paradigm.

"ACPR will set new standards for enhancing the journal at the article level by leveraging today's online capabilities to offer a media-rich, interactive electronic edition of the journal, including podcasts and venues where contributors and users can interact," said Kate Caras of Indiana University Press.

Going Mobile

The emergence of mobile technologies will be a game changer. An article entitled "Looking Ahead at Social Learning" in *Training and Development Magazine* reports that by 2015, more people will connect via the Internet through their mobile device than by PC.

"Mobile gives us the opportunity to rethink the article 'box,' because articles don't just 'shrink to fit' for mobile devices," said John Sack. "I am hopeful that we can apply that rethinking outside of planning for mobile. For example, as we think about use cases for mobile—which should guide what we 'mobilize' and how—we should also think about the use cases for information in other media and devices," said John Sack.

Smartphones, iPads, and e-readers are already in the process of transforming the face of publishing forever. The mobile climate provides MUSE the opportunity to create new and different tools resulting in a fresh end-user experience. While accessible on handheld devices now, MUSE is working on a mobile site to launch in 2011 and investigating discovery apps.

"After 15 years of putting the print online, we need to be sure we don't spend the next 5 years putting the online website on mobile wheels. That would be a lost opportunity," said Sack.

Towards a Content Community

The future looks bright for articles in the Humanities and Social Sciences. Much progress has been made on the electronic front, but the migration from print to web has been gradual. There is a lot of excitement around e-books and initiatives that are underway to digitize the scholarly monograph and make it discoverable. The scholarly monograph's time has come.

Project MUSE will launch an integrated content platform in July of 2011 that will include a collection of e-books from a selection of university presses. A similar initiative funded by Mellon will bring 50 university presses together on an e-book platform.

As we continue to enhance the user experience at the article level, traditional content formats are beginning to blur. A search for information on any academic library website will return results that include journal articles and book chapters in one place, side-by-side, and stripped of the branded publisher environment.

"Over the past two years, we're beginning to see books and journals appearing together on the same platform," said Judy Luther.

A move towards the integration of multiple content formats—journals, books, reference works, datasets, YouTube videos, and others—on a fully-discoverable platform has begun.

"I do think its importance will increase as more and more researchers depend on the online content to provide supplemental data to the more traditional scholarship," said Cason Lynley of Duke University Press.

In the book *Groundswell*, authors Charlene Li and Josh Bernoff wrote that the Net Generation will "use technologies to get the things they need from each other, not from traditional institutions." Freedom, fun, collaboration, and customization are important values to the Net Generation. The reader of the future will use technology to push the boundaries of creativity, collaboration, connectedness, and community-focused interaction.

"Research in humanities—at least as I did it years ago—is pretty different. A lot of work with primary sources," said John Sack. "We need to find ways to tie the resources that humanities scholars and social science researchers use into the online 'web' of information."

Project MUSE has created a dynamic content community for scholars in the humanities and social sciences, established a home for publishers and libraries with a common purpose, and will continue to innovate through a shared sense of collaboration at the article level.

"Project Muse just keeps getting better and better, and inclusion in the database is essential to the success of any journal—especially a new title—in today's increasingly online world," said Kate Caras.

| FE | doi: 10.3789/isqv22n3.2010.02

DEAN SMITH <dsmith@press.jhu.edu> is Director of Project MUSE and Wendy Queen <wendy@muse.jhu.edu> is Associate Director of Project MUSE.

RELEVANT LINKS

ADText
www.adtextonline.org

COUNTER
www.projectcounter.org

Google Scholar
scholar.google.com

JACS Beta
pubs.acs.org/JACSbeta/

Looking Ahead at Social Learning. Training and Development Magazine, July 2010.
www.astd.org/TD/Archives/2010/Jul/Free/1007_LookingAheadAt.htm

Project MUSE
muse.jhu.edu

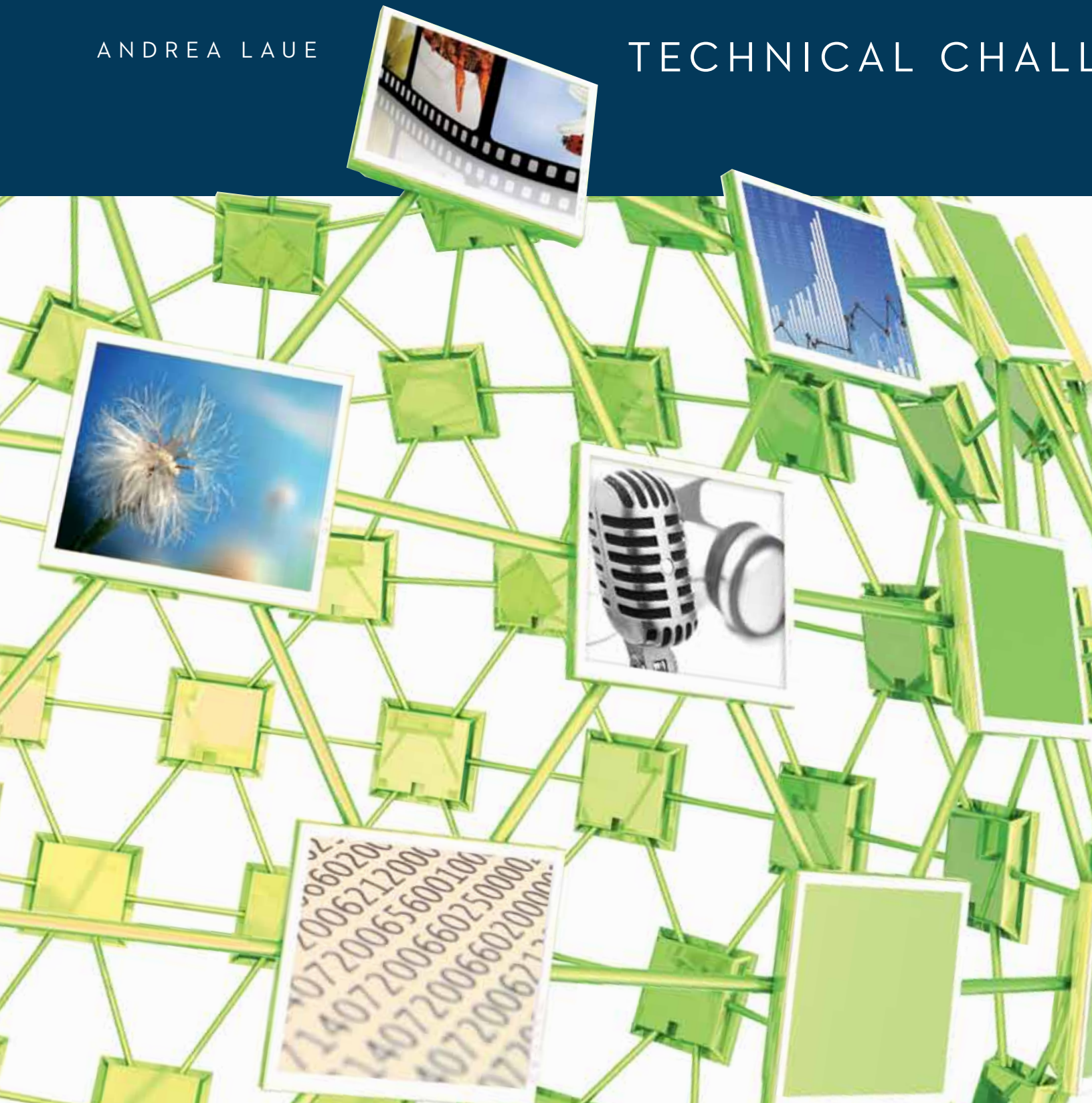
Townsend, Robert A. Is There a Future for Journals in the Humanities? AHA Today Blog, September 2, 2009.
blog.historians.org/articles/865/is-there-a-future-for-journals-in-the-humanities

Waltham, Mary. The Future of Scholarly Journals Publishing Among Social Science and Humanities Associations.
www.nhalliance.org/bm~doc/hssreport.pdf

HOSTING SUPPLEM

ANDREA LAUE

TECHNICAL CHALL



IENTARY MATERIAL:

LENGES & SUGGESTED BEST PRACTICES

Debates about the definition of supplementary material point to a larger transition in scholarly and technical publishing from print to electronic articles as versions of record. In the past, a piece of evidence was deemed "supplemental" because it couldn't be presented in print. The medium defined what constituted rhetorically valid evidence. As online versions overtake print, the impulse is to imagine that medium is no longer a constraint. Any and all supplemental information can be "included" online.

In her now famous editorial in *Cell*, Emilie Marcus writes of setting "limits" in defining supplementary materials. Marcus defines supplemental materials via their proximity to the main argument. The point is to admit only evidence that is integral to the paper and appropriate for the medium, whether that medium be electronic or print. What's appropriate for the electronic article may currently exceed and continue to expand faster than that of print, but it's not without bounds.

Commenting on online display conventions for supplemental materials, Ian Brown of HighWire Press issued the caveat that "'supplemental materials' are truly that—they should supplement the primary information presented in the article, but not themselves be central to it. Technical limitations of various media make integration of data more difficult (e.g., it may be hard to effectively incorporate audio datasets into print), but those limitations ideally would never affect whether core information was treated as 'supplemental' or not." Brown gets at the issue of medium from a slightly different tack here, arguing that materials considered "supplemental" in print might best be modeled as part of the core article in an electronic environment.

The challenge, then, is twofold. First, the rhetorical challenge of presenting the most appropriate evidence given the medium of publication. Second, the technical challenges associated with any medium of publication, whether it be print or online.

The task of describing, displaying, and searching supplemental materials online remains underexplored territory. Simply dumping those materials on the web and providing links is only marginally better than citing a video in a print article. Hosting of those materials poses technical and business challenges, and issues surrounding preservation are just now getting the attention they deserve.

Technical Challenges

Supplemental materials present several challenges to online publishers of scholarly and technical content. Five technical challenges will be discussed here—the first second and fourth in some depth and the remaining two in a more cursory manner:

1. Display
2. Search
3. Hosting
4. Article markup
5. Preservation

CONTINUED »

1. DISPLAY

Audience and device are two key variables to consider when designing display of supplemental materials. Different audiences are likely to have different levels of interest in supplemental materials, and different devices are likely to have different capacity for displaying them.

Ian Brown, HighWire, suggests that "a 'complete' presentation of an article should include descriptions of, references to, and (where possible) ready access to all material used to create the article. This so that researchers are able both to understand the authors' conclusions (by reading the article itself) and evaluate the methodologies used and quality of the conclusions (by looking at supporting data)." Descriptions should characterize the content of the supplementary materials as well as the technical requirements for viewing them. An article should contain explicit links to all supplementary items as well, connecting that extra material to a specific argument made within the article.

Some publishers now include supplementary materials sections in their (electronic) articles. "Publishers opting for such presentation might consider grouping materials into sets, one that supports the conclusions of the article and another that prompts additional questions and future research," suggests Brown. Evidence that complicates findings might also be addressed and presented in a separate section.

Different audiences are likely to have different levels of interest in supplementary materials. John Sack, Director of HighWire, has noted that clinicians are less likely to consult supplementary materials than researchers. Clinicians—readers generally—are also increasingly likely to read research articles on a mobile device, which poses access and data management issues when it comes to supplementary materials.

Ironically, mobile devices and printed PDFs pose similar challenges when it comes to supplementary materials. Both make simple descriptions of the supplementary materials desirable, as access to the materials themselves cannot be assumed. Sack advised that the PDFs have supplementary materials copied into the PDF so as to avoid taking readers by surprise when they discover that the article they printed isn't complete. This might also be addressed by offering two versions of the article PDF, one with supplementary material included and one without it.

Brown suggests a "package" download, whereby a reader could retrieve the article and all associated supplements. This "one stop shopping" presented alongside the article PDF would address the needs of clinicians and researchers both, while presenting each with the other option.

Concern over length and complexity of PDFs reminds me of earlier anxieties about connection speeds and file sizes of PDFs. Perhaps we'll soon see "article-only PDFs" and "full PDFs," the latter containing all supplementary materials suitable for print, much like we once had low-resolution "screen" PDFs for reading and high-resolution "print" PDFs for downloading.

2. SEARCHING

Searching locally-hosted, non-XML supplementary resources and remotely-hosted resources of any format presents particular challenges. Some of these challenges are technical (e.g., tagging of video so that transcripts can be searched and hits accurately located within the video's binary object) and others are conceptual (e.g., does a text search operate on the article and dataset in the same manner). Adding the complexity of federated search (local and remote resources) to that presents a daunting challenge.

Todd McGee, Assistant Director of Application Systems at HighWire Press, argues for integrated searching of article and supplemental material content. "By definition supplementary data does not stand on its own," observes McGee, and "that being the case the supplemental data should be indexed along with the original article and should, by default, be included in a simple fulltext search." When it comes to search, supplemental materials should be treated like other components of an article, McGee feels. Search options should "follow the pattern that a publisher follows for other sections of an article/chapter."

Search results lists that include "hits" within supplemental materials should clearly indicate the parent/child relationship of the article and supplemental materials. A results list from a general search should take pains to label supplemental materials and clearly associate the "hit" with the argument that invoked that material as evidence. If a given piece of supplemental material is hosted remotely, the search result should include a link directly to the supplemental object.

A targeted search of supplemental data should follow the same conventions as other section- or object-specific searches, suggests McGee. "Typically this would be to return a link to the parent item; optionally that link might be to a positional anchor to the section of the article where the supplementary data can be viewed."

McGee advises publishers to request transcripts for audio/video supplements to make this material searchable and accessible to all users, regardless of their abilities. Supplements supplied as Text, HTML, PDF, etc. should be indexable by most search engines. Compressed files or very large datasets present special cases and in most cases would not be directly indexable by many search engines.

Including descriptive metadata of supplements not susceptible to standard indexing would offer some means for improving searchability. This metadata might be indexed by default, or it might be called on only in those cases where transcripts or other surrogates were unavailable. The NLM Journal Publishing DTD (see #4 below) isn't optimized for this purpose, however, so pursuing this strategy will likely require some creative interpretations of or extensions to the standard content model.

3. HOSTING

Hosting supplementary materials presents a variety of challenges and requires constant attention to an always shifting set of formats and requirements. From a vendor's perspective, the key decision is whether to host in-house or contract with a third party for the technical expertise and transmission bandwidth to serve supplementary materials.

Digital video is a good example here. Competing formats (Flash, QuickTime, WindowsMedia) with different server and player requirements are constantly moving targets. Investing in and maintaining the infrastructure to host and securing the bandwidth to serve the videos involves potentially significant cost. Alternatively, there are third-party vendors willing to partner with publishers and take on this portion of the risk. Then the technical challenge becomes one of linking from locally-hosted content to remotely-hosted supplementary materials, a task that is relatively well understood and manageable. Ideally, the use of standard identifiers such as DOIs facilitates this approach, as DOI-based query services and searches on DOIs themselves relieve the need for literal URLs in source content and promote the general discoverability of supplemental materials wherever they are hosted.

4. ARTICLE MARKUP

The NLM Journal Publishing DTD version 2.3[1] offers the supplementary material element (<supplementary-material>) for modeling supplemental material. The element was designed to contain references to "additional data files that contain information directly supportive of the document, for example, an audio clip, movie, database, spreadsheet, applet, or other external file." Acknowledging the multiple meanings of the word "supplementary material," the Tag Library identifies as candidates for this element: extra tables, supportive materials "too voluminous" to appear within the narrative, and materials such as quizzes and forms that "enhance" the article content.

The content model of supplemental material allows one to describe any number of media objects, graphics, or tables, and the supplementary material element can be repeated in those locations where it's allowed. As such, one may model multiple pieces of supplemental evidence as individual items or grouped into any number of sets.

Four points of association have been suggested for supplemental materials:

- **Article-level association**
- **Structural unit or point within the article body**
- **Specific figure**
- **Separate section, often at the end of the article**

The DTD accommodates all of these use cases, although support for associating supplemental material with a specific figure is limited.

The Tag Library describes two use cases for its supplementary material element. First, listing of supplementary materials in the article metadata so that those materials can be accessed from the article. Second, positioned inline like a figure (<fig>), in which case a position attribute may be applied so as to indicate whether the supplementary material is intended to be located at its point of reference or instead relocated to a point convenient for the particular display.

The first use case, listing of supplementary materials in the article front matter (<front>), models accurately the association of those materials with the article as a whole. Any number of supplementary materials elements can be included.

CONTINUED »

1 No major changes to the modeling of supplementary material were introduced in version 3.0 of the NLM DTD.

The *Tag Library* describes two use cases for its supplementary material element. First, listing of supplementary materials in the article metadata so that those materials can be accessed from the article. Second, positioned inline like a figure (<fig>), in which case a position attribute may be applied so as to indicate whether the supplementary material is intended to be located.

CONTINUED »

The second use case, associating supplementary material at a specific point within an article, may be accomplished either by positioning it (<supplementary-material>) within the narrative flow or by encoding a cross-reference to supplementary material tagged in the article front matter. The first model facilitates inline display but implies exclusion based on some constraint of the medium rather than the nature or role of the referenced material. Encoding the supplementary materials in the front matter and encoding cross references at mention in the narrative flow seems in keeping with the notion that supplementary items extend rather than support the basic argument of the article.

The content model of section (<sec>) includes supplementary materials as a child. Thus, a list of supplementary materials may be modeled as a (final) titled or untitled section in an article. As article navigation often keys on section titles, this content model may offer built-in support for navigating the supplementary materials without any additional development work.

Modeling supplementary material as an extension of a particular figure poses a challenge to users of the NLM DTD. The supplementary material element is not included in the content model of figure (<fig>). In most contexts supplementary material may be a sibling of figure, and adjacency might be used to imply ancestry, although that approach seems fraught with peril.

The NLM DTD also offers an inline supplementary material element (<inline-supplementary-material>) to model links to such materials. Allowed as a child of any paragraph, this element might be used within a figure caption to point to supplementary materials associated with a particular figure. The dearth of attributes allowed on the element limit the semantic richness with which the referred-to materials might be described, however. Presumably the text of the parent paragraph (part of the caption) could describe for the reader what she will find at the other end of the link, but reproducing that in a format actionable by the machine poses different challenges.

The content model of inline supplementary material does offer external link (<ext-link>), which in turn offers a type attribute (@ext-link-type) that might describe the reference material. The current Tag Library offers a hodgepodge of values from "doi" to label a DOI to "genpept" to identify a specific database. While "doi" is too general to identify the supplemental materials, values such as "genpept" might facilitate some inferences as to what's at the other end of the link. The list of allowed values for this type attribute is not restricted, but the community might benefit from agreeing on some standard values.

A DOI may be associated with supplementary material by means of object identifier (<object-id>) child elements (with pub-id-type attribute values of "doi"). The list of allowable identifier types is not restricted, although the published list is under-descriptive. It's advisable to register DOIs for supplementary materials, and that's likely to remain a primary identifier. However, as (or if) the number of repositories of datasets and other supplementary materials grows, the need for additional values will likely grow. The catch-all **content-type** attribute may also be applied to object identifiers, allowing for additional labeling.

The supplementary material element offers a limited set of attributes for encoding metadata about the referred-to materials. MIME type and language may be defined using standard attributes. The catch-all attribute **content-type** is available to capture special semantic intent of the tagged content. Version 3.0 of the NLM Journal Publishing DTD adds the **specific-use** attribute, which "specifies distinctions in the applicability of a particular element." The Tag Library advises using this attribute to specify audience for or medium in which the contents of the element are relevant.

Child elements such as attribution (<attrib>) and permissions (<permissions>) may be included to indicate authorship or intellectual property restrictions that differ

It's advisable to register DOIs for supplementary materials, and that's likely to remain a primary identifier.

from that of the article as a whole. The content model of attribution does not allow for formal tagging of authors, however.

Supplemental materials may require additional metadata to support online display. For example, displaying an inline video may require a call to a player. That call may require a few basic display parameters, e.g., height, width, and duration. Depending on the format of the video and the preferred player, the available parameters and the syntax required when passing them may vary.

At HighWire Press, we've defined a separate manifest file that accompanies each video. This XML file accommodates a variety of metadata values designed to facilitate online display. The manifest knows the video by its DOI, as does the article XML. The article and all associated manifests are part of the processing context, so display variables are available during generation of the browser XHTML, for example.

This design acknowledges the relative stability of the NLM XML DTD and the relative instability of digital video encoding and players. The article XML refers to the video as an external object known via standard identifier. The manifest handles the specifics of that object. This model would allow a change in video format without a concomitant update to the article XML, for example. In short, this model externalizes the aspects of video content tagging that are most likely to change.

5. PRESERVATION

The questions of how to preserve supplementary materials and where they should be deposited are intimately linked. Online publishers have invested significantly in data models and storage architectures designed to preserve core article content. Accomplishing this requires, among other things, the adoption of some shared standards and basic formats. There is no such uniformity in the area of supplementary materials.

Consortiums and vendors have been slow to volunteer to take the lead in defining standards and offering centralized storage of data sets. Much has been invested in the adoption of XML and an industry standard XML DTD for journal publishing, but no similar effort has been undertaken for datasets. While authors can be relatively confident that the text of their articles will be processable (by a machine) for years to come, there is little reason to be as confident about datasets. The PARSE project (Permanent Access to the Records of Science in Europe) offers some hope in this area, although its scope is currently limited to European research.

Conclusion

Without any consensus on standards or best practices in the handling of supplemental materials, hosting services are left with setting their own requirements for publishers who submit content or with dealing with each publisher's own policies and practices—a laborious and costly approach. Several projects currently underway are intended to develop standards and best practices for supplemental materials. The NLM Journal Archiving and Interchange Tag Suite version 3.0, the three journal article schemas, and the documentation are currently being shepherded through the NISO standardization process to become a consensus ANSI standard and attract a wider audience of publishers. A joint NISO/NFAIS effort to define best practices for supplemental materials is just underway (see article on page 33). The technical working group for that effort will no doubt consider the issues described in this article. As these standards and best practices are completed, both hosting services and publishers can look forward to adopting some common approaches to the handling of supplementary materials.

| FE | doi: 10.3789/isqv22n3.2010.03

ANDREA LAUE <andrea@highwire.stanford.edu> is Manager, Content Services at HighWire Press <highwire.stanford.edu>. She is a member of the NISO/NFAIS Supplemental Journal Materials Technical Working Group.

RELEVANT LINKS

DOI® System
www.doi.org

Marcus editorial in Cell
download.cell.com/pdf/PIIS0092867409011817.pdf

NLM Journal Publishing Tag Set
dtd.nlm.nih.gov

PARSE
www.parse-insight.eu

DAVID S. H. ROSENTHAL AND VICTORIA A. REICH

ARCHIVING SUPPLEMENTAL MATERIALS

It has long been considered important that institutions other than the publisher preserve academic journals. Libraries fulfilled this role when the publishing medium was paper. Shortly after journals started their transition to the Web in the mid-90s, The Andrew W. Mellon Foundation started studying how they should be preserved.[1] These studies bore fruit; now institutions other than the publisher routinely preserve e-journals.

THERE ARE TWO ARCHITECTURES IN USE: centralized and distributed. The LOCKSS Program represents the distributed approach[2], but some years of experience in operating systems with both architectures shows that the differences are matters of detail. The two approaches share many major issues, and in particular those caused by the importance of preserving supplemental materials.[3]

We examine this problem from the perspective that the same institutions that already preserve the primary journal content should use the same technologies to preserve supplemental materials. We have relevant experience; the LOCKSS program currently preserves supplemental materials in this way. Setting up these institutions, providing them with viable business models, and developing the necessary technologies has over the last decade proven to be a major effort. It is unrealistic to believe that a similar but separate effort could be undertaken on behalf of supplemental materials, which by their nature are normally regarded as less valuable than the primary content.

Sustainability

It has become clear that the single most difficult issue in digital preservation in general, and in e-journal preservation in particular is "economic sustainability." To be blunt, on a cost-per-byte basis it costs too much. Even the most cost-effective approach known, that of the Internet Archive, is too expensive to meet its goals. The vastly more expensive techniques being used for e-journals are similarly inadequate.[4] The LOCKSS program has been cash-flow neutral for some years, but that may not be enough for long-term sustainability. As for the sustainability of Portico, the other major e-journal preservation system, early this year an audit reported:

> "the ongoing business viability of Portico as a service is not yet assured, judging from financial information disclosed to date."[5]

Examination of the latest tax returns available (2008[6]) for Ithaka, the parent organization, shows that Portico was at that time short of cash-flow neutrality by a substantial margin.

Adding a broad range of supplemental materials to the task of preserving primary e-journal content, however it is done, will inevitably add costs and thus make the problem of economic sustainability worse. How significant are these costs likely to be?

Ingest

The question thus becomes "how to minimize these additional costs?" Experience has shown that the dominant cost in e-journal preservation is the ingest process.[7] This is not unexpected. Long-term studies of the cost of storing data at the San Diego Supercomputer Center[8] have shown that it decreases over time, although not as fast as the famous "Innovator's Dilemma"[9] exponential drop in the cost per byte of disk storage would lead one to expect. The much-feared costs of format migration have not in practice been incurred, since formats have not been going obsolete at anything like the rate predicted by Jeff Rothenberg in the mid-90s.[10] Ingest costs cannot be delayed to take advantage of the time value of money.[11]

CONTINUED »

Adding a broad range of supplemental materials to the task of preserving primary e-journal content, however it is done, will inevitably add costs and thus make the problem of economic sustainability worse.

CONTINUED »

The reason why ingest is the dominant cost is that it consumes staff time, which is expensive in absolute terms and tends to increase over time. In order to ingest an e-journal's content, the archive must first obtain permission from the copyright owner to do so. This takes negotiation between the archive and the publisher, and often involves lawyers on both sides. Second, the preservation system must be adapted to the peculiarities of each journal. Third, the ingest process must be carefully monitored to ensure that routine problems such as intermittent network and publisher outages, or unannounced changes in the publishers' systems, do not interfere.

It is relatively cost-effective to ingest content from major publishers such as Elsevier. Although the negotiations are time-consuming, this cost is amortized across a large number of journals. Their systems are well engineered, consistent across many journals, stable, and well documented.[12] But the content of the major publishers is at low risk of being lost, so the value of archiving the large amount of content obtained in this way is low.

The content at high risk of loss comes from smaller publishers. Although negotiations with smaller publishers are typically easy, each results in only a small amount of content. Their systems are more diverse, less well documented, and less stable than those of major publishers. Thus the efforts involved in adapting the preservation system to the journal, in monitoring the ingest process, and in handling the more frequent exceptions detected, are all much greater both absolutely and on a per-byte basis.

Thus we see that the two major staff time sinks in the ingest process are diversity and low volume. Supplemental materials are by their nature more diverse and lower volume than the primary content, and can thus be expected to be more expensive both on a per-item and a per-byte basis.

This expectation is reinforced by preliminary results from a survey of data preservation costs under the auspices of the UK's JISC. This survey found that ingest costs including the process of obtaining permission to do it were over half the total:

> "The cost of archiving activities (archival storage and preservation planning and actions) is consistently a very small proportion of the overall costs and significantly lower than the costs of acquisition/ingest or access activities... As an example the respective activity staff costs for the Archeology Data Service are: Access (c.31%), Outreach/Acquisition/Ingest (c.55%), Archiving (c.15%)."[13]

Best Practices

The application of best practices, or even better standards, to the process of publishing supplemental materials can reduce diversity and aggregate the materials into larger, uniform collections. Doing so will not merely reduce the cost of preservation but also the cost of publishing, finding, and accessing them. In particular, these best practices should be aimed at reducing staff time, since this is the biggest cost component in each of these tasks.

In what areas are best practices likely to have the greatest impact in reducing staff time? We identify four:

1. Intellectual Property
2. Location and Structure
3. Technical Metadata
4. Bibliographic Metadata

1 Intellectual Property

The existing intellectual property constraints, both informal and formal, on sharing of data are diverse, unclear and in flux. This complicates the archive's task, which requires clarity as to the permission that the archive has to keep copies of the publisher's intellectual property, and about the terms under which the data is to be preserved and in future accessed.

As regards informal constraints, surveys of authors' attitudes to sharing data[14,15,16] uniformly report a great diversity among fields, funders, and practitioners. In some fields, such as astronomy, sharing is the norm, albeit modulated in some cases by delays allowing those researchers who captured the data "right of first publication." In others, particularly those with the potential of valuable patents, sharing is the rare exception. In these fields it is to be expected that any supplemental materials actually published will be of low value; they will have been sanitized to ensure they do not compromise the potential for commercial exploitation.

As regards formal constraints, the legal situation is complex. Some supplemental materials are copyrighted, but it isn't clear that the same copyright terms apply to them as to the text of the article to which they are attached. Some data are just facts, so are not subject to copyright. Some data represents a compilation of facts, so may be subject to copyright or may in the European Union (but not elsewhere) be subject to "database right."

Furthermore, despite the efforts of Science Commons,[17] there is no widely used equivalent of the Creative Commons (CC) license for copyrighted data.[18] The reason is not hard to understand; the CC license is grounded in well-established copyright law. Because the legal framework surrounding data is much less clear, it has been much harder to establish a strong means for allowing the right to use data while providing a guarantee of credit that is the most frequent desire of researchers. The recent release of the Open Data Commons "BY" license for databases[19] is a step in the right direction.

Similarly, there is no equivalent of the machine-readable means for labeling content with the appropriate CC license.[18] Lacking such means, ingest programs that collect data from supplemental materials are on shaky legal ground.

Best practice efforts are thus urgently needed in two related areas:

- *An analog of the CC "attribution" license*, allowing researchers to grant general permission to use their data provided credit is given. This would satisfy a substantial proportion of researchers. Additional license versions could be added later to satisfy other groups of researchers.
- *A machine-readable form of this license*, similar to the RDF form of the CC license, allowing automated harvesting of data from supplemental materials.

2 Location and Structure

There are two basic forms in which e-journal content can be ingested:

- *"Source" content*, in which the publisher packages up the content it wishes the archive to preserve in some form different from that in which the content was originally delivered to readers and then transmits it, often via FTP, to the archive for processing and preservation. Source is in practice something of a misnomer. Sometimes, the "source" content includes the actual source (e.g., SGML markup) but it almost always includes exactly the same rendered form of the content that was delivered to some readers (e.g., PDF).
- *"Presentation" content*, in which the archive behaves exactly as a reader would, accessing the e-journal's website and ingesting the same HTML, CSS, PDF, JavaScript, and other formats that the reader's browser would interpret.

Note that for primary e-journal content this distinction has some long-term relevance. If a "source" publisher supplies actual source, for example SGML markup, then the archive will contain information not normally available from a "presentation" publisher. Conversely, the archive of a "presentation" publisher will contain information, for

CONTINUED »

example the CSS and JavaScript implementing the e-journal's look-and-feel, not normally available from a "source" publisher.

But for supplemental materials, and especially data, this distinction is unlikely to have long-term relevance. This data is frequently neither source to be processed by the publisher's web infrastructure into a form usable by a web browser nor is it a presentation to be interpreted by a web browser. It is normally raw input to some other program, in particular one different from that used by the original authors.

Thus the only important difference is whether the archive collects the data in the same way that readers would (presentation) or in some form packaged by the publisher (source):

- In the presentation case, the archive's ingest web crawler must be able to identify those links in an article pointing to supplemental materials and any associated metadata that should be preserved along with the article. For example, a recent article in *Science* [20] illustrates AAAS' approach to supplemental materials. A link in the *Article Views* sidebar *Supporting Online Materials* points to a landing page[21] describing and linking to a single PDF file with a Materials & Methods section, figures, and tables. The ingest web crawler needs to know that it should follow this chain of links.
- In the source case, the archive's ingest process must be able to identify in the package form created by the publisher (often a tar or zip archive), the relationship between the article text, any associated components, the supplemental materials, and any associated metadata. For example, a recent article in the *Journal of Monetary Economics*[22] as it appears on the Web has a link near the end of the paper's text to a single PDF file with supplementary material. In the packaged source format Elsevier uses[12], this PDF appears in the same directory as the PDF, XML, and raw ASCII of the primary article. There is no XML or raw ASCII for the supplement.

Best practices for making these connections that were robust enough to enable similar automatic processing across a range of e-journal publishing technologies would be useful. Weaker best practices would have little effect, either on preservation or other tasks.

The LOCKSS software currently ingests supplemental materials in both presentation and source forms, but only from major publishing platforms such as HighWire Press[23] and Elsevier[12]. In both cases there are one-time and continuing per-publisher costs involved in doing so, but they are not large. We would expect these costs to increase as smaller publishers and smaller publishing platforms increase their use of supplemental materials; effective best practices would reduce the expected increase.

3 Technical Metadata

We have argued elsewhere[4] that the advent of the Web triggered a switch from documents as private to applications to documents published for many applications, and that this effectively turned document formats into network protocols, which are almost immune from the backwards-incompatible changes that cause format obsolescence. We have also argued[4,24] that the increasing importance of open source has similar effects for similar reasons.

A corollary of these arguments is that the technical metadata provided by the Web (Mime type, magic numbers, etc.) is adequate, since it clearly enables web browsers, including open source browsers, to render the content.

Some of these arguments are weaker when applied to supplemental materials in the form of data. Although it is not the private property of a particular application, it is also less "published" and more dependent on metadata other than the basic web metadata. These considerations raise the importance of publishing supplemental materials in forms that can be accessed by open source tools, such as XML with public DTDs. Best practices codifying this would be useful both for preservation and the kinds of data-mining activities championed by, for example, Peter Murray-Rust.[25]

In addition, standards for representing the technical and scientific metadata that supplemental materials need in addition to the basic web technical metadata would be very useful, although the benefits would not accrue primarily to preservation but rather to the eventual users of the preserved materials.

4 Bibliographic Metadata

There are standards for the attribute names[26] and to a lesser extent for the formats[27] and vocabularies for the bibliographic metadata describing the articles in journals. They are not as well observed in practice as one might hope, but they are useful. Extending them to cope with supplemental materials would be useful, as would best practices stressing the importance of conformance to metadata standards, and tools verifying such conformance.

It is noteworthy that while Elsevier's source format[12] includes the most comprehensive bibliographic (and technical) metadata about primary articles of any publisher we have worked with, it includes no metadata about supplementary materials except an MD5 digest of the file. It is not even possible to discover from the supplied metadata whether or not an article has supplementary material.

Conclusion

We have identified that one goal of codifying best practices, or even standardization, with respect to supplemental materials should be to reduce the cost of ingest by eliminating tasks needing human intervention. Suggested areas with the potential to do so are:

- An analog of the CC "attribution" license, allowing researchers to grant general permission to use their data provided credit is given.
- A machine-readable form of this license, similar to the XML form of the CC license, allowing automated harvesting of data from supplemental materials.
- Uniform means for connecting articles, their supplemental materials, and the metadata for the supplemental materials, both in e-journal websites and in the packaged formats used by "source" publishers.
- Standard representations of the metadata needed by supplemental materials in addition to the basic web metadata.
- Publishing data in supplemental materials in forms that can be processed using open source tools.
- Extensions to existing metadata standards and practices to allow for detailed description of supplemental materials.

| FE | doi: 10.3789/isqv22n3.2010.04

DAVID S.H. ROSENTHAL <dshr@stanford.edu> is Chief Scientist and **VICKY REICH** <vreich@stanford.edu> is the Director at LOCKSS (www.lockss.org). David's blog on digital preservation is available at: blog.dshr.org/.

REFERENCES

1. Cantara, Linda (ed.). Archiving Electronic Journals. Digital Library Federation, 2003. www.diglib.org/preserve/ejp.htm
2. LOCKSS Program www.lockss.org/
3. Roundtable on Best Practices for Supplemental Journal Article Materials January 22, 2010 Washington, DC. Co-sponsored by NFAIS and NISO. www.niso.org/topics/tl/supplementary/
4. Rosenthal, David S. H. How Are We Ensuring The Longevity Of Digital Documents. Coalition for Networked Information, April 2009. blog.dshr.org/2009/04/spring-cni-plenary-remix.html
5. Report on Portico Audit Findings. Center for Research Libraries, January, 2010. www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories/portico
6. Ithaka tax returns available from: www2.guidestar.org/Home.aspx
7. Eakin, Loraine, et al. A Selective Literature Review on Digital Preservation Sustainability. Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2009. brtf.sdsc.edu/biblio/Cost_Literature_Review.pdf
8. Moore, Richard L., et al. Disk and Tape Storage Cost Models. Archiving 2007, Arlington, VA, May 2007, pp. 29-32. www.imaging.org/IST/store/epub.cfm?abstrid=34413
9. Christensen, Clayton M. The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Harvard Business School Press, June 1997. ISBN: 978-0060521998
10. Rothenberg, Jeff. Ensuring the Longevity of Digital Documents. Scientific American, 272(1), 1995.
11. Rosenthal, David S. H. Format Obsolescence: the Prostate Cancer of Preservation. dshr's blog, May 2007. blog.dshr.org/2007/05/format-obsolescence-prostate-cancer-of.html
12. ScienceDirect OnSite Product Specifications. Elsevier, March 2006. info.sciencedirect.com/implementation/implementing/sdos/
13. Beagrie, Neal, et al. Keeping Research Data Safe. JISC, 2010. www.beagrie.com/KRDS2_selectioncriteria.pdf
14. Key Perspectives, Ltd. Data Dimensions: Disciplinary Differences in Research Data Sharing, Reuse and Long Term Viability. Digital Curation Centre, January 18, 2010. www.dcc.ac.uk/sites/default/files/documents/publications/SCARP-Synthesis.pdf
15. Harley, Diane, et al. Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines. Center for Studies in Higher Education, UC Berkeley, January 2010. cshe.berkeley.edu/publications/publications.php?id=351
16. Nelson, Bryn. Data Sharing: Empty Archives. Nature, 461, pp. 160-163, 2009. doi:dx.doi.org/10.1038/461160a
17. Scholar's Copyright Project. Science Commons. sciencecommons.org/projects/publishing/
18. About Licenses. Creative Commons. creativecommons.org/about/licenses
19. ODC Attribution License (ODC-By). Open Data Commons, June 24, 2010. www.opendatacommons.org/licenses/by/1.0/
20. Gibson Daniel G., et al. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science, 329 (5987), pp. 52–56, July 2, 2010. doi:dx.doi.org/10.1126/science.1190719
21. Science magazine supporting material example: www.sciencemag.org/cgi/content/full/sci;science.1190719/DC1
22. Balia, Turan G., and Robert F. Engle. The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations. Journal of Monetary Economics 57, (4), pp. 377-504, May 2010. doi:dx.doi.org/10.1016/j.jmoneco.2010.03.002
23. HighWire is Hosting. HighWire Press. highwire.stanford.edu/about/
24. Rosenthal, David S. H. Format Obsolescence: Scenarios. dshr's blog, April 28, 2007. blog.dshr.org/2007/04/format-obsolescence-scenarios.html
25. Murray-Rust, Peter. Data-driven Science – A Scientist's View. NSF/JISC Repositories Workshop, April 10, 2007. www.sis.pitt.edu/~repwkshop/papers/murray.html
26. Dublin Core Metadata Initiative. dublincore.org/
27. Term Name: date. *In*: DCMI Metadata Terms. DCMI Usage Board, January 14, 2008. dublincore.org/documents/dcmi-terms/#terms-date E.g., "Recommended best practice is to use an encoding scheme [for dates], such as the W3CDTF profile of ISO 8601."

NISO IS WORKING

to make electronic content more accessible.

Researchers and library patrons are increasingly expecting instant access to the information they need. While the availability of electronic content grows daily and standards such as OpenURL have drastically improved discovery, impediments still remain. **At NISO, a number of current projects are underway to improve discovery, access, and delivery of content:**

KBART: Phase II

PHASE I of the joint NISO/UKSG KBART (Knowledge Bases and Related Tools) project resulted in practical recommendations for exchanging metadata between content providers and knowledge base developers. These recommendations are intuitive, easy for content providers to implement, and easy for knowledge base developers to process.

PHASE II builds on that work to focus on the more advanced, complex issues that cause problems in this area. Learn how to implement the recommendations from Phase I and about the next stage of this work at www.niso.org/workrooms/kbart.

IOTA: Improving OpenURLs Through Analytics

IOTA is a two-year project to investigate the feasibility of creating industry-wide, transparent, and scalable metrics for evaluating and comparing the quality of OpenURL implementations across content providers. At this time, nearly 9 million OpenURLs have been analyzed from log files. The reports created from this analysis allow publishers to see where they can make improvements to their OpenURL strings so that the maximum number of OpenURL requests can be resolved—bringing more readers to their products. Visit openurlquality.niso.org to view the metrics and learn how to add your data to the project. Find out more at www.niso.org/workrooms/openurlquality.

E-Journal Presentation & Identification

Unless journal websites accurately and uniformly list all the titles under which content was published, user access to desired journal articles is considerably diminished. When journals change titles or publishers, their content must remain easily accessible. This new working group will be developing recommendations that will provide much-needed guidance on the presentation of e-journals to publishers and platform providers—particularly in the areas of title presentation, accurate use of the ISSN, and citation practices—that will solve some long-standing concerns of serials librarians. See www.niso.org/workrooms/ejournalpresentation/ for more information.

ESPReSSO: Establishing Suggested Practices Regarding Single Sign-On

This NISO Chair's Initiative was launched to develop recommendations that will improve the user experience when using diverse electronic services by providing transparent single sign-on authentication across distributed service providers. The end result of this work will be small, smart conventions for moving the user within a session from one licensed site to another, so that publisher content can be accessed easily and seamlessly. Find out more at www.niso.org/workrooms/sso.

DO YOU HAVE A SUGGESTION FOR NEW WORK? WE'D LIKE TO HEAR FROM YOU!

www.niso.org/standards/suggest OR visit www.niso.org/workrooms for more information.

OP [OPINION]

A judgment formed about something;
a personal view, attitude, or appraisal

Sasha Schwarzman

SASHA SCHWARZMAN

Supplemental Materials Survey

In October 2009, Alexander (Sasha) Schwarzman at the AGU (American Geophysical Union) conducted an informal survey of scientific journal publishers to learn how other publishers were dealing with the issue of "supplemental materials." Conducted mainly through the e-mail listservs of CrossRef TWG and eXtyles, Schwarzman's questions "touched a raw nerve" and generated more responses than he had been expecting. This article is an extract of the full survey report, issued in November 2009, which is available from the AGU website (see relevant links at the end of this article). Schwarzman's article was the impetus for a January 2010 Supplemental Materials Roundtable meeting on the subject co-sponsored by NISO and NFAIS and the subsequent Working Group on Supplemental Journal Materials that the two organizations launched (see article on page 33).

Problem Statement

As Emilie Marcus, Editor in Chief of *Cell*, put it in her editorial, *Taming Supplemental Material* (*Cell* 139(1):11 (2009), doi: 10.1016/j.cell.2009.09.021):

Unfortunately, over the years supplemental material has evolved into a seemingly limitless repository for additional "stuff". ...It has become a mechanism for expanding the overall content of a paper without any delineated change in editorial standards. ... Authors often feel compelled, by their own desire to be comprehensive and in response to questions raised in the review process, to include increasingly large amounts of data that exceed the traditional restrictions of the printed article. Reviewers may feel responsible, as the supplemental material is ultimately published as part of the peer-reviewed publication, to assess this information with the same attention and standards as the main body of the article, which often means that they are asked to evaluate the equivalent of two papers in the place of one. And readers may find it difficult to navigate through large supplements and may be unsure about how carefully the supplemental material was evaluated in the review process.

What is the definition of supporting material?

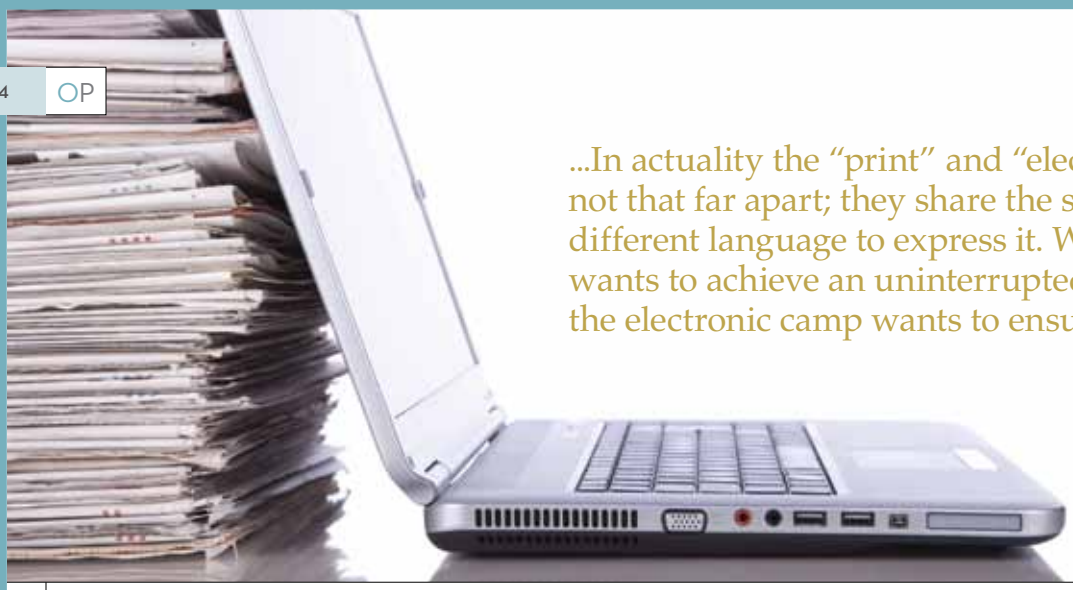
There is a clear split within the publishing community between those who declare the electronic article the copy of record and those who don't. The supporting material definition is easier for those publishers who consider the print journal to be the normative copy; for them, anything that cannot be printed automatically falls into the category of supporting material.

For those of us, however, who define the electronic article as the copy of record, the decision is not so obvious. [The *Cell* editors in] Elsevier's "Article of the Future" initiative distinguish between **three major conceptual categories:**

1. evidence that provides deeper support for the points made in the main paper,
2. large data sets and multimedia that can only be presented online, and
3. detailed information about the methods.

Other publishers think along similar lines, e.g., "material that is not critical to the overall message of the paper but which supports it," "information that will be of interest to some readers but is not essential to the central message of the paper," "data and other materials that directly support the main conclusions of a paper but are considered additional or secondary."

CONTINUED »



...In actuality the "print" and "electronic" groups are not that far apart; they share the same concern but use different language to express it. While the print camp wants to achieve an uninterrupted flow of narrative... the electronic camp wants to ensure smooth navigation.

CONTINUED »

Does the notion of supporting material make sense in electronic-only environment?

There is no clear consensus here. Interestingly, many respondents who are currently dealing primarily with print tend to think that the notion may not be applicable in the electronic world. Yet, this optimistic view is not shared by electronic publishers. It seems to me, however, that in actuality the "print" and "electronic" groups are not that far apart; they share the same concern but use different language to express it. While the print camp wants to achieve an uninterrupted flow of narrative (and to do so dumps the offending interrupters "on the Web"), the electronic camp wants to ensure smooth navigation (and to that end dumps the culprits on the lower levels where they are less visible—either through an ingenious user interface or by providing a link instead of displaying an item right away).

Who is to decide what supporting material is?

There is a virtual consensus here that while the initial division between "main" and "supporting" material comes from the author, the ultimate decision must rest with the editor who has to have guidance from the publisher.

Personally, I think that once a conceptual decision of what constitutes supporting material is made, a submission system interface can help a great deal in guiding the author in this respect.

How do you ensure uniform application of "supportiveness" criteria?

Everyone seems to be resigned to the fact that there can be no uniformity in applying the "supportiveness" criteria across different journals published by the same publisher, much less across the entire scientific discipline. However, I would think that a publisher should articulate what the criteria are for a given title and insist that editors apply them consistently. Otherwise, a publisher risks that the decision will be made selectively or arbitrarily, and the editors will be left in the "I know it when I see it" situation.

Should supporting material be peer-reviewed?

I am happy to report that everyone, without exception, thinks that supporting material must be peer-reviewed.

What are different kinds of supporting material? Does it exist on the level of article components only or that of an entire article?

It appears that we can distinguish between two main kinds of supporting material, each treated somewhat differently:

- Supporting components, e.g., supporting tables, figures, multimedia, computer programs, etc. (Data sets are a special type of this component.)
- Supporting structural section, e.g., text (narrative), possibly containing math and a separate reference list.

To state the obvious, while supporting components exist on the component level, structural sections exist on the level of an entire article.

Some publishers explicitly stipulate how many [supporting] components, and of what type an article may contain. Other publishers have no explicit restrictions on how many supporting components can be accepted. Importantly, there is often a difference between the "main" and "supporting" components in (a) their acceptable formats, and (b) whether and to what extent they are processed.

When it comes to data sets, we can distinguish between two rather different cases: (1) those data sets that have been deposited to one of the official data centers and those that have not. CrossRef accepts metadata deposits for data sets, so a data set can have a DOI. In the area of geophysics there is a World Data Center System Roster and I suspect that similar approaches exist in other disciplines, such as astronomy, biology, etc. The important point here is that when a data set is deposited with an official data center the whole "supporting vs. main" issue becomes irrelevant; the component is now an external resource that can be cited in the references by its metadata and [identifier]. It seems to me, it would be in the publisher's best interests to make every effort to encourage

authors to deposit their data sets to an official data center or even insist that they do so once the manuscript has been accepted.

Some journals, especially those where articles conform to a rigid format, define very clearly which sections fall into the "supporting" category. For others the picture is less clear; there is no consensus on what constitutes in-article Appendix versus online supporting material. The same kind of derivation of a formula or a proof of a lemma can in one case be part of an Appendix, while in the other appear only online.

What about readability, usability, preservation, and reuse?

Why does a scientist need a publisher? Well, of course we shepherd the manuscript through peer review, but we also add value to the content in a number of other ways:

- make it readable through copy editing;
- make it navigable and accessible through user interface;
- make multichannel publishing, e.g., Web/HTML, Web/PDF, Print/PDF, PDAs, iPhone/Blackberry, e-Readers, etc., possible by applying markup in accordance with de-facto semantic and syntactic best practices;
- facilitate the relationship of an article to its scientific context and promote its discoverability by linking references, building citation indices, assigning DOIs to the article and sometimes to its components, and depositing/disseminating article metadata through abstracting and indexing services;
- preserve the narrative by printing it on an acid-free paper or/and marking it up; and
- preserve the components by ensuring they are submitted in/converted to formats that have a good chance of survival or could at least be migrated with lossless conversion.

When we look at supporting material we discover:

- with rare exceptions, supporting material is not being copy edited;
- supporting material items are usually not presented the same way as their "main" brethren, e.g., instead of an individual HTML document/section or a carefully processed image one will see a link to a PDF or MS Word [file], or to the whole group of documents, sometimes of different type (tables, figures, text) stitched together;
- supporting structural sections are universally not being marked up;
- supporting references are not being deposited and are not being linked;
- supporting components are often presented in author-submitted formats that do not meet archival standards or won't be easily migrate-able;
- even when supporting material is provided in standard formats, e.g., PDF/A, such formats are less likely to be usable than more robust ones, such as XML.

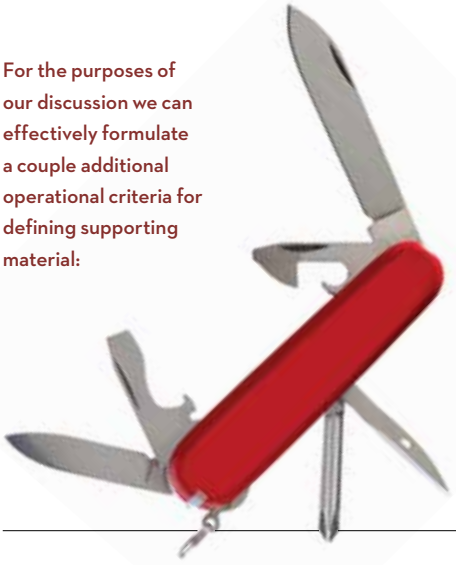
The implications:

While a publisher makes a reasonable effort to ensure that the main content of the article lends itself to multichannel publishing, the probability is lower for supporting material.

While a publisher can be reasonably confident that the scientific content of the article can be recreated in the future as technology changes, the same cannot always be said about supporting material with the same degree of confidence.

CONTINUED »

For the purposes of our discussion we can effectively formulate a couple additional operational criteria for defining supporting material:



Usability. Supporting material is not likely to be as versatile, robust, and usable as the main article when it comes to multichannel publishing.

Longevity. While the main article is going to enjoy eternal life with many reincarnations along the way, supporting material is likely to rot and die, with very little possibility of resuscitation.

CONTINUED »

It is a common concern that supporting material has become a "back-door to publication."Two approaches to stemming the abuse are: enacting strict editorial guidelines, and charging authors for supporting material.

What this means for the purposes of our discussion is that, effectively, we can formulate a couple additional operational criteria for defining supporting material:

- **Usability** – Supporting material is not likely to be as versatile, robust, and usable as the main article when it comes to multichannel publishing.
- **Longevity** – While the main article is going to enjoy eternal life with many reincarnations along the way, supporting material is likely to rot and die, with very little possibility of resuscitation.

Still there is no free lunch

Even though supporting material is not processed nearly to the same degree as the main article, it is still not without a cost to a publisher. Supporting material needs to be integrated with the main article; some degree of quality control needs to be exercised; the material's existence needs to be reflected in the metadata; minimal markup needs to be applied, etc. Yet, with only one exception, all respondents have indicated that they do not charge authors for supporting material.

It seems to me, there are two aspects to the question of supporting material cost: on the one hand, a publisher absorbs supporting material processing expenses; on the other hand, a publisher saves costs by not holding supporting material to the same standards of usability and longevity as the main article.

The majority of respondents stated that they take a "pragmatic" approach to dealing with supporting material. Leaving aside the non-printability issue, it appears that "pragmatic" here can refer to two different things: (a) arriving at a working definition of what is essential and what is not to the scientific conclusions of the article, and (b) achieving a trade-off between saving costs by sacrificing usability and longevity and providing access that should suffice at least in the short run.

Preventing abuse

It is a common concern that supporting material has become a "back-door to publication."

I could discern two approaches to stemming the abuse: (a) enacting strict editorial guidelines, like imposing a limit on the total number of supporting components, and (b) charging authors for supporting material.

Tagging practices

The NLM Journal Publishing Tag Set allows one to tag a supporting section the same way as any other structural section and give it a requisite title. When it comes to tagging a component, the element <supplementary-material> allows one to treat it in a variety of ways. The approach of the Tag Set is to consider <supplementary-material> to be an element on par with <graphic> or <media> elements, rather than to be able to indicate that a particular <graphic> or <media> plays a "supporting" role.

There is a consensus that eXtyles has no problem exporting supporting material to XML, which is not surprising, given the fact that supporting sections and components markup is minimal.

Summary

While all of the publishers surveyed were distributing [supplemental] materials, there was little consistency in how they were handled. There was consensus in the view that all supplemental materials should be peer-reviewed, but not necessarily about the rigor of that review. The size and scope of the supporting materials was an issue, as well as if and where those materials reside online. Publishers generally responded that supplemental materials did not go through the same production processes, such as editing, layout, consistent markup, etc. While ensuring that the supporting data remained intact and unchanged, this lack of production management could lead to problems when a publisher wants to archive the information or migrate it to a future system.

| OP | doi: 10.3789/isqv22n3.2010.05

ALEXANDER (SASHA) SCHWARZMAN <SSchwarzman@agu.org> is an Information Systems Analyst-Designer at the American Geophysical Union (AGU) <www.agu.org>. He is co-chair of the NISO/NFAIS Technical Working Group on Supplemental Journal Materials.

RELEVANT LINKS

Supplemental Materials Survey (full report)
www.agu.org/dtd/Presentations/sup-mat/

Cell's Supplemental Information Guidelines
www.cell.com/supplemental_information_guide

CrossRef Info for Publishers
www.crossref.org/02publishers/

Data Set Prototype (Dryad)
datadryad.org/repo/handle/10255/dryad.20)

eXtyles (Inera)
www.inera.com/extylesinfo.shtml

NLM Journal Publishing Tag Set
dtd.nlm.nih.gov

Taming Supplemental Material editorial
download.cell.com/pdf/PIIS0092867409011817.pdf

World Data Center System Roster
www.ngdc.noaa.gov/wdc/list.shtml

SP [SPOTLIGHT]

Patricia Feeney

PATRICIA FEENEY

DOIs for Journals: Linking and Beyond

As most readers of ISQ probably know, Digital Object Identifiers (DOIs) are alphanumeric strings assigned to digital objects. Each DOI is unique and, once assigned to an item, remains a constant locator, not changing even as object moves from URL to URL. DOI names are assigned to a range of content but have been most readily embraced by the world of scholarly publishing and by researchers looking for consistent links to mutable resources.

T he DOI system is managed by the International DOI Foundation (IDF), an organization that provides oversight to DOI registration agencies and maintains the DOI resolver. CrossRef, a non-profit membership organization dedicated to promoting collaboration between scholarly publishers, is the official registration agency for scholarly materials including journals, books, reports, and conference proceedings, and has registered over 41 million DOIs on behalf of our members. The vast majority (over 36.5 million) of CrossRef DOIs have been assigned to journal articles. This article will focus primarily on CrossRef's implementation of the DOI, with some coverage of how other organizations are delivering DOI-linked content in ways that enhance journal articles.

From an end user perspective, DOIs are used primarily in citations, both in print and online. The most recent edition of the *Publication Manual of the American Psychological Association* recommends that authors include DOIs in their references, allowing researchers to easily locate a cited item by clicking on (or in the case of print typing in) simple DOI links. DOIs for journal articles, books, and conference proceedings have become the standard persistent identifier for most scholarly publishing disciplines. CrossRef DOIs are primarily assigned to individual articles, but publishers opt to assign DOIs on a broader level, using DOIs to link to title and issue level pages, as well as tables of contents. An example of this is http://dx.doi.org/10.1002/(ISSN)1522-2454, a title-level DOI that links to the home page of *Vakuum in Forschung und Praxis* on Wiley InterScience.

Each CrossRef journal DOI must link to a response page containing bibliographic information and a means to access full text—a DOI does not grant access to content, instead it provides a publisher-approved route to accessing full text. Most DOIs link to text-based journal content, but a DOI can link to alternate formats as well. *The Journal of Visualized Experiments*, a peer reviewed video journal for biological research, assigns DOIs to what are essentially video articles (example: http://dx.doi.org/10.3791/1733). Emerging formats present challenges on many fronts, but from an identifier perspective all formats are the same, provided the metadata describes the object.

The rules for creating DOIs are defined in the standard, *Syntax for the Digital Object Identifier* (ANSI/NISO Z39.84). To create a DOI, publishers obtain a DOI prefix from CrossRef, assign individual DOIs to digital objects, and deliver XML-encoded metadata to the CrossRef database. The CrossRef system in turn registers the deposited DOI and URL with the IDF. The deposited metadata consists of basic citation information that can be used to identify and describe a digital

CONTINUED »

As most readers of ISQ probably know, Digital Object Identifiers (DOIs) are alphanumeric strings assigned to digital objects. Each DOI is unique and, once assigned to an item, remains a constant locator, not changing even as object moves from URL to URL.

```
<citation_list>
  <citation key="1">
    <issn>0386-2615</issn>
    <author>Takata Seiji</author>
    <volume>32</volume>
    <issue>4</issue>
    <first_page>162</first_page>
    <cYear>2005</cYear>
  </citation>
  <citation key="2">
    <doi>10.1007/BF01969578</doi>
  </citation>
```

Figure 1: Sample citation deposit

Figure 2: Sample Multiple Resolution page from Graft
Source: http://dx.doi.org/10.1177/1522162802239753

Reference linking benefits publisher members by driving traffic cross-publisher, and gives end users a reliable route to finding cited articles online.

CONTINUED »

object. No full text, abstracts, or other content is deposited—only data necessary to describe and locate the item is necessary. A journal article deposit, for example, contains bibliographic metadata such as the journal title, ISSN, volume, issue, page numbers, article title, and author names, as well as other identifying data such as internal publisher identifiers, codens, title abbreviations, language used, and contributor roles.

DOIs and Reference Linking

CrossRef is very much a collaborative effort. Publisher members commit to depositing and maintaining DOIs for all online journal content, but members also commit to querying the CrossRef system to harvest DOIs deposited by other members. The retrieved DOI links are then included in their reference lists published online. This practice, known to CrossRef members as reference linking, is an integral part of CrossRef as an organization. The reference linking process is powered by the metadata submitted with each DOI. Reference linking benefits publisher members by driving traffic cross-publisher, and gives end users a reliable route to finding cited articles online.

Cited-by Linking

CrossRef members may also participate in cited-by linking, an optional service that allows publishers to display citations from other publications that cite their content, providing an easily implemented way to display cross-publisher citations. Participating publishers must include citation metadata for reference lists within their article DOI deposits (see Figure 1), and in turn are able to query the reference lists of other publications. The citations are submitted as XML metadata or as already-deposited DOIs. This service is currently only available for journal content—almost 16 million

Components comprise an ever expanding assortment of data types, ranging from figures and tables to images, video, audio, and PowerPoint presentations. They allow publishers to create durable links to figures, tables, and supplemental content that can be easily updated.

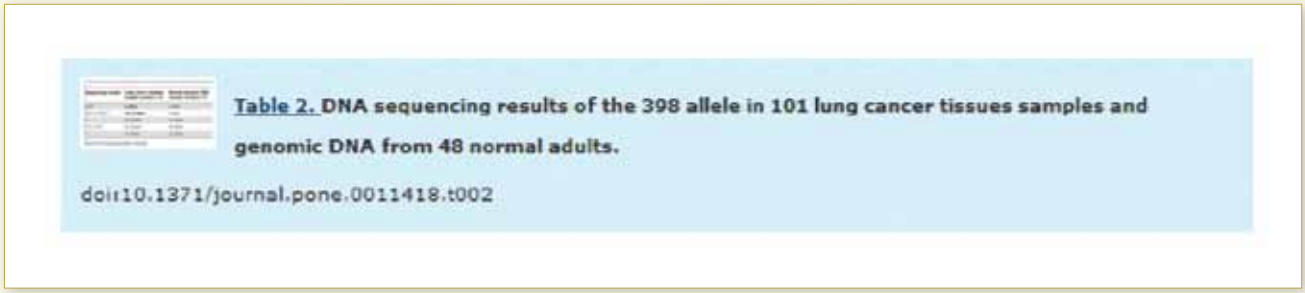


Figure 3: Component DOI
Source: http://dx.doi.org/10.1371/journal.pone.0011418

journal DOIs (or 34%) have at least one cited-by link. The cited-by linking network only accesses the data of participating members, but the number of participants grows constantly.

Multiple Resolution

CrossRef DOIs conventionally link the user to a single source of material but, in select circumstances, an item might exist in multiple locations or formats. The DOI specification supports a practice called multiple resolution in which multiple URLs may be attached to a single DOI. As implemented by CrossRef, instead of delivering the user directly to content, the DOI resolves to an interim page containing citation metadata and multiple links to an item. This feature has been enthusiastically adopted by members who co-publish journals, as it allows them to dually host an authoritative version of an article.

A recent focus on preserving online journal content has resulted in cooperative efforts between archiving institutions to preserve and provide continuing access to titles that have ceased publication and are no longer maintained by the original publisher. The multiple resolution process allows DOIs assigned to these journals to resolve to multiple hosts, allowing end users to choose between the archiving organizations that host the content. DOIs are currently assigned to *Auto/Biography* and *Graft*, originally published by SAGE, and *Brief Treatment* and *Crisis Intervention* from Oxford University Press (OUP), both of which have been archived by Portico and CLOCKSS (see Figure 2).

DOIs for Supplemental Content

Publishers are increasingly delivering supplemental journal content online, and DOIs can be assigned to supplemental materials as well. Publishers generally use two content types to link to supplemental materials: components and datasets. Supplemental materials are typically not cited on their own and as such aren't discoverable by querying the CrossRef system, but assigning DOIs allows publishers to easily create and update durable links to content that otherwise might not survive platform migrations and ownership changes. Other registration agencies facilitate assigning DOIs to supplemental content as well, particularly data not provided by the publisher such as datasets, videos, maps, and raw scientific data.

Components

Components comprise an ever expanding assortment of data types, ranging from figures and tables to images, video, audio, and PowerPoint presentations. They allow publishers to create durable links to figures, tables, and supplemental content that can be easily updated. Only a small number (~300,000) of CrossRef's 41 million+ DOIs are components, but the number grows daily. The CrossRef definition of component is fairly loose, viewing them as a container element and allowing the publisher to determine how their supplemental material is classified. Consequently required component metadata is simple, consisting of sparse metadata describing the content and file type. The metadata perhaps most relevant to the

CONTINUED »

Figure 4: Multiple component DOIs
Source: http://dx.doi.org/10.1107/S1600536810020751

CrossRef collects a number of dataset DOIs but they are also increasingly being registered by organizations devoted to delivering datasets and other types of raw scientific data.

CONTINUED »

component is that of the item the component is supplementing, also known as the parent DOI. Other component metadata consists of an item description, format (or file type), and of course the DOI and URL. Optional elements include item titles, contributor information, and publication dates.

A component must be associated with a parent DOI that has been created for a CrossRef content type (journal, book, conference proceeding, technical report, working paper, standard, dissertation, or dataset). The majority of deposited components are associated with journal articles. Although components have not been widely adopted across the membership, several CrossRef members have successfully integrated them into their content. The Public Library of Science uses components to link to tables and figures for their journal PLoS ONE. The tables and figures appear within the text in both the print and online versions of an article, with the DOI listed below an image thumbnail (see Figure 3). This component DOI links directly to the full-sized table or figure.

The International Union of Crystallography (IUCr) uses components to provide durable links to crystallographic information files (CIFs) and other supplemental materials in a variety of formats, as well as including a DOI directed at an HTML page containing all supplemental material for an article (see Figure 4). IUCr also includes the component DOI within the HTML version of the article.

Datasets

DOIs may also be assigned to datasets, a content type dedicated to database records. Datasets typically exist as stand-alone databases, but individual dataset records or a database as a whole may be used to supplement journal article data. CrossRef collects a number of dataset DOIs but they are also increasingly being registered by organizations devoted to delivering datasets and other types of raw scientific data. Dataset providers and journal publishers are able to provide

In the coming months CrossRef will be launching a new project, CrossMark, that will allow users to retrieve information about publisher-maintained versions of a document—including the status of a document, publisher metadata, and of course the CrossRef DOI assigned to the document.

Data from: A test of the "flexible stem" model of evolution: ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation

When using this data, please cite the original article:

Wund MA, Baker JA, Clancy B, Golub JL, Foster SA (2008) A test of the "flexible stem" model of evolution: ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation. American Naturalist 172: 449-462. doi:10.1086/590966

Additionally, please cite the Dryad data package:

Wund MA, Baker JA, Clancy B, Golub JL, Foster SA (2008) Data from: A test of the "flexible stem" model of evolution: ancestral plasticity, genetic accommodation, and morphological divergence in the threespine stickleback radiation. Dryad Digital Repository. doi:10.5061/dryad.1159

Dryad Package Identifier	doi:10.5061/dryad.1159
Dryad Data Files	http://hdl.handle.net/10255/dryad.1160 http://hdl.handle.net/10255/dryad.1161
Abstract	If an ancestral stem group repeatedly colonizes similar environments, developmental plasticity specific to that group should consistently give rise to similar phenotypes. Parallel selection on those similar phenotypes could lead to the repeated evolution of characteristic ecotypes, a property common to many adaptive radiations. A key prediction of this "flexible stem" model of adaptive radiation is that patterns of phenotypic divergence in derived groups should mirror patterns of developmental plasticity in their common ancestor. The threespine stickleback radiation provides an excellent opportunity to test this prediction because the marine form is representative of the ancestral stem group, which has repeatedly given rise to several characteristic ecotypes. We examined plasticity of several aspects of shape and trophic morphology in response to diets characteristic of either the derived benthic or limnetic ecotypes. When marine fish were reared on alternative diets, plasticity of head and mouth shape paralleled phenotypic divergence between the derived ecotypes, supporting the flexible stem model. Benthic and limnetic fish exhibited patterns of plasticity similar to the marine population; however, some differences in population means were present, as well as subtle differences in shape plasticity in the benthic population, indicating a role for genetic accommodation in this system.
Scientific Names	Gasterosteus aculeatus

Figure 5: Dataset DOI landing page with original journal article citation and DOI
From source: http://datadryad.org/handle/10255/dryad.1209

durable links to supplemental content by cross-linking between hosted datasets and journal content.

DOI linking between datasets and journal content is nascent but off to a promising start. One example of this is Dryad, a newish data repository focused on evolutionary biology and ecology that partners with a number of major evolutionary biology and ecology-centric publications. Dryad lists among its goals to "preserve all the underlying data reported in a paper at the time of publication, when there is the greatest incentive and the ability for authors to share their data." Accordingly, they register DOIs for the datasets they deliver. When compiled in conjunction with a published article the dataset landing page provides DOI links to the parent article, as shown in Figure 5, an example of data supporting a paper published in *Molecular Ecology*.

In the Dryad example, the parent article does not provide a link back to the dataset. Collaboration between dataset providers and journal publishers does exist, as evidenced by a dataset hosted by PANGAEA, an open access network for geo-scientific and environmental data. A journal article DOI link is provided in the citation, and the response page for the article DOI contains a link to PANGAEA supplementary data.

DOI registration for journal articles has become an accepted practice, as have CrossRef enhancements such as reference and cited-by linking. More and more publishers struggling to represent their supplemental data online are using DOIs for linking. Recent efforts to assign DOI links to raw data are encouraging and more reciprocal linking can be expected in the future, as can other DOI-related enhancements. In the coming months CrossRef will be launching a new project, CrossMark, that will allow users to retrieve information about publisher-maintained versions of a document—including the status of a document (withdrawn, corrected, enhanced, etc.), publisher metadata, and of course the CrossRef DOI assigned to the document. The CrossMark process will be enabled in part by the metadata deposited with CrossRef DOIs. Although this project is still in the planning stages, it's a promising sign of how DOI use can evolve. | SP | doi: 10.3789/isqv22n3.2010.06

PATRICIA FEENEY <pfeeney@crossref.org> is the Metadata Quality Coordinator at CrossRef.

RELEVANT LINKS

CrossRef website
www.crossref.org

CrossMark
www.crossref.org/crossmark.html

CLOCKSS
www.clockss.org

Dryad
www.datadryad.org/factSheet

PANGAEA
www.pangaea.de

Portico
www.portico.org

Publication Manual of the American Psychological Association
www.apastyle.org

Syntax for the Digital Object Identifier (ANSI/NISO Z39.84)
www.niso.org/standards/z39-84-2005/

SERU

SERU: A SHARED ELECTRONIC RESOURCE UNDERSTANDING (NISO-RP-7-2008)

IT'S TIME

SERU IS FOR YOU

An alternate to e-resource licenses

Libraries and Publishers save time and money.
SERU offers libraries and publishers the option to reference a set of common understandings as an alternative to negotiating a signed license agreement.

Developed by a NISO working group comprised of librarians, publishers, subscription agents, and lawyers, SERU is a recommended practice that is designed to streamline the acquisitions/sales process.

The SERU recommended practice is available for free download from: www.niso.org/standards/resources/RP-7-2008.pdf.

Libraries and publishers rapidly adopting SERU

More than 70 Libraries • Eight Consortia • Over 30 Publishers

Publishers and librarians agree on the products for which they wish to reference SERU and forgo a license agreement. The SERU Registry helps to identify publishers and libraries interested in using SERU for electronic resources. Publishers who wish to use SERU with any of their products and librarians who would like to request that SERU apply to some of their products are quickly joining, using, and appreciating the benefits of SERU. Follow their lead and sign up to the SERU Registry today! **www.niso.org/workrooms/seru/registry/**

Benefits of SERU include:

- ✓ Easier e-resource subscription transactions
- ✓ Rapid acquisition and minimal delay for access
- ✓ Time and cost savings for both libraries and publishers

How SERU can work for you

- ✓ Sign the registry to show your interest in using SERU
- ✓ Select products or services to which SERU may apply
- ✓ Reference SERU in the purchase documents
- ✓ Link to SERU on the NISO website

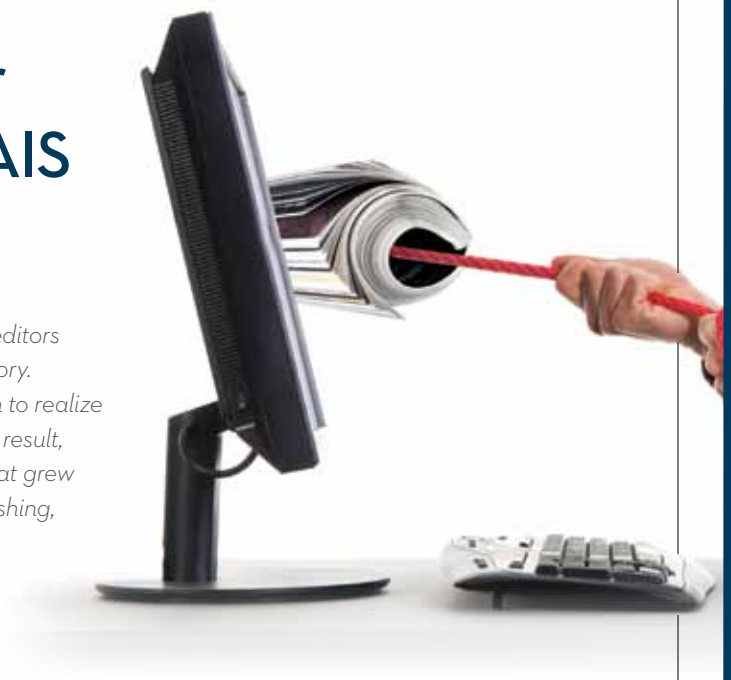
NISO

www.niso.org/workrooms/seru/registry/

Linda Beebe

LINDA BEEBE

Supplemental Materials for Journal Articles: NISO/NFAIS Joint Working Group



The tug of war between authors who wish to show all their work and editors concerned both about reader acceptance and page limits is an old story. Once electronic publishing was firmly established, both parties began to realize that supplemental materials could perhaps satisfy their concerns. The result, depending on the discipline and journal policy, was an initial trickle that grew rapidly into a flood for some. As with many aspects of electronic publishing, there were no standards or recommended practices for dealing with supplemental materials.

*What emerged was a messy stew of **different** approaches.*

In fact, there exists no clear consensus on what constitutes supplemental materials. Some journal editors practice peer review, others do not. Some journals post supplemental materials along side the article within the journal; others post the materials to the open web. Often the latter are missing journal article connections. That is, the reader of the article will find a link to the supplemental materials, but the individual finding the materials serendipitously may not learn what they might have been attached to or what the context is for the data found in a table or figure. Journals rarely offer a recommended citation for the materials; indeed some journals warned initially that supplemental materials were never to be cited separately from the journal article. Most frequently, supplemental materials suffer from a lack of descriptive metadata.

Solving this messy problem is the purpose of one of NISO's newest working groups. The goal of the Joint NISO/NFAIS Working Group is to "create a Recommended Practice for publisher inclusion, handling, display, and preservation of supplemental journal article materials." The impetus for the formation of this working group started two years ago.

Prelude I—NFAIS Working Group

In 2008, an NFAIS Working Group looked at the practices of publishing electronic journal articles overall. Initially directed toward best practices for article-by-article publishing, the group, representing primary and secondary publishers as well as librarians, in the end considered practices for all journal articles published in electronic form. Among the problems they focused on were workflow issues, confusion among versions, problems with citation structures, linking problems, and discoverability issues when articles did not reach Abstracting and Indexing (A&I) services promptly.

Best Practices for Publishing Journal Articles received final approval from the NFAIS Board of Directors in February 2009. That document included recommendations on 1) affirmation

CONTINUED »

Once published, the supplemental materials should be considered part of the journal's archival record and should not be changed without a clear statement of correction.

CONTINUED »

of the journal; 2) article retrieval; 3) version management; 4) supplemental materials; 5) content creator; 6) indication of length; 7) article identifiers; 8) citation elements required and publisher display of recommended citation; 9) tables of contents and indicators of completeness; 10) journal editor identification; and 11) copyright statement.

One key recommendation on Supplemental Materials was that the journal make a clear connection between an article and the supplemental materials that accompany it. Once published, the supplemental materials should be considered part of the journal's archival record and should not be changed without a clear statement of correction. Publishers, the document noted, should always supply a recommended citation as well as good, descriptive metadata for those materials. A&I services covering the journal article should include the presence of supplemental data in the article record, indicating file types and DOI.

Prelude II—Schwarzman White Paper

In Fall 2009, Sasha Schwarzman, Information Analyst-Designer at the American Geophysical Union, surveyed a number of his technical colleagues in other organizations about their experience with supplemental or supporting materials. Based on the thoughtful responses he received from seven publishers, he wrote a white paper describing the lack of consensus around handling these materials.

Some variances stemmed from differences between print and electronic. If the print is considered the version of record, supplemental materials are likely to be items that cannot be printed because of their medium or format or due to page limits. If the electronic is the version of record, the distinction seems to be less clear and varies from publisher to publisher. Some journals restrict the number of components in supplemental materials. For example, *Cell* limits them to no more than twice the number of figures and tables in the article. Some designate as supplemental only those items that cannot be delivered in print.

Schwarzman found that there was no consensus on whether the materials should be included as an appendix to the article or placed online separately. Then, he noted, there are issues around readability, usability, preservation, and reuse. With rare exceptions, he found that publishers are not copy editing this content, treating it with the same deference they do article content, or marking it up. These deficits suggest that the odds of the content being as versatile and useful as the article itself or even having a very long life are slim. However, he noted that correcting these problems would result in substantial financial costs that would need to be borne by the publisher or the author. [See article on page 23 for more findings from this survey.]

Expanding the Discussion

Because of the growing interest in the topic and the amount of discussion that Schwarzman's paper generated, Todd Carpenter, NISO's Managing Director, suggested that NISO

and NFAIS jointly sponsor a roundtable to discuss the issues that had been raised. When the group convened in Washington, D.C. in January 2010, nearly 70 people were engaged, either on site or on the phone for a meeting that lasted several hours.

Three speakers were invited to set the stage for discussion. First, Sasha Schwarzman expanded on his white paper to question where supplemental materials belong structurally within scientific articles. He noted it is essential to maintain these articles as a literary genre. Supplemental materials, he said, pose a threat of abuse: abuse of reviewers who are asked to review a catchall array of content and abuse of readers who may wade through an unreadable jumble of content. He discussed the costs and benefits and concluded with concerns about business models and uniform application of policies across the industry or even within a single journal.

Eefke Smit, Director of Standards and Technology, International Association of Scientific, Medical, and Technical Publishers (STM), presented a report on PARSE (Permanent Access to Records of Science in Europe). PARSE is a European Union project that aims to highlight the vulnerability of digital content. For the past two years, the project has been creating an inventory with surveys and case studies. Their findings are that datasets and supplemental materials are the least organized for preservation. The characteristics of an ideal system, they note, include a good linking system, reliable metadata, certification of repositories, and registration of datasets.

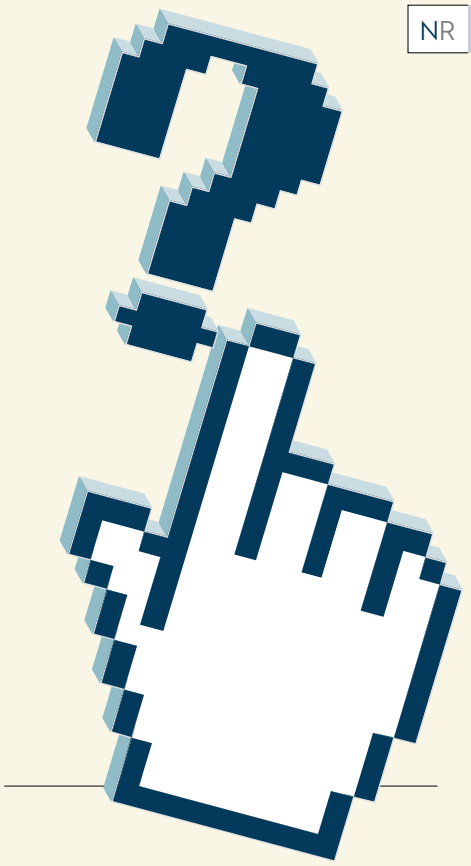
Scott Dineen, Deputy Senior Director of Publications, Optical Society of America (OSA), provided an update on the Interactive Science Publishing project OSA undertook in publishing large databases with its journals. The experiment, a partnership with the National Library of Medicine, included a DSpace architecture (MIDAS) that would allow them to accept datasets. They then created viewing software to allow readers to rotate, crop, zoom, and analyze. Thus, a reader could, for example, view 3D lung cancer datasets in context. One of the key issues was a lack of metadata. OSA is now looking at funding and business models to continue the work.

Defining the Problem

The first question for discussion was what exactly are supplemental materials. Although the group rather quickly developed a potential list of content types (see sidebar), the deeper question of what is designated as supplemental went unanswered. Different organizations approach the materials differently. For example, AGU's position is that these materials provide the next step for the reader who wishes more information, but should not be essential to the reader's understanding of the article. On the other hand, AAAS, the publisher of *Science*, treats supplemental materials as a tool for authors to make their case, thereby making supplemental materials "essential to the scientific integrity of the article." Articles published in *Science* are frequently very short, whereas the supplemental materials may be quite extensive.

Decisions about what is supplemental have thus far been largely subjective and may be made independently by any of several players. The group discussed potential roles for authors, peer reviewers, editors, publishers, and libraries/data centers. All agree that supplemental materials should receive the same level of peer review that an article does;

CONTINUED »



What *exactly* are supplemental materials?

- Figures (including high-resolution)
- Tables
- Movies
- Software/scripts (or network files)
- Videos
- Appendixes
- Audio Files
- Images
- Text
- Datasets

Although the group rather quickly developed a potential list of content types, the deeper question of what is designated as supplemental went unanswered.

CONTINUED »

however, there are real questions about how that is done today and how publishers might cover the costs of the review.

Findability, participants agreed, is a big issue in part because A&I services receive inconsistent notification. Some publishers supply good metadata and others do not. Some assign DOIs and others do not. **Among the significant questions the Roundtable posed are these:**

- What exactly are supplemental materials? Should they be considered part of the main article or might they be linked, but separate items?
- What is the impact for preservation, citations, and copyright agreements?
- Some materials from extensive studies may be shared across articles and perhaps across authors. How will this work?
- How can costs be managed?
- How might issues around sharing data be handled? Among these issues are sensitive information, such as patient data; permission or use restrictions; embargoes; and the growing requirement to share data.
- How will publishers weigh competing user needs around supplemental materials?

Moving Forward

In discussing the formation of a follow-up working group, participants in the Roundtable identified some general issues for potential Recommended Practices. **Among them are the following:**

- Clear, consistent indicators of content
- Metadata needs
- Universal agreement on citation practices
- Consideration of use of the DOI
- Potential cost recovery
- Common vocabulary
- Peer review
- Preservation and interaction with repositories
- Archiving
- Clearly defined specific responsibilities for the parties involved in scholarly publication.

Further discussion can be found in the report of the Roundtable Meeting.

Supplemental Journal Article Materials Working Group Rosters

Business Working Group

Linda Beebe,
American Psychological Association (Co-chair)

Marie McVeigh
Thomson Reuters (Co-chair)

Annette Flanagin
JAMA and Archives Journals

David Gillikin
National Library of Medicine

Bruce Kiesel
Thomson Reuters

Amy Kirchhoff
ITHAKA

Bonnie Lawlor
NFAIS

Alison Loudon
American Institute of Physics

Skip Maier
APA Journals

Jill O'Neill
NFAIS

Eefke Smit
International Association of STM Publishers

Technical Working Group

Dave Martinsen
American Chemical Society (Co-chair)

Sasha Schwarzman
American Geophysical Union (Co-chair)

IJsbrand Jan Aalbersberg
Elsevier

Ken Beauchamp
American Society for Clinical Investigation

Jeffrey Beck
NCBI, National Library of Medicine

Tshawna Byerly
Byerly Editorial Services

Rachael Hu
California Digital Library

Chuck Koscher
CrossRef

John Kunze
California Digital Library

Kathy Kwan
NCBI, National Library of Medicine

Deborah Lapeyre
Mulberry Technologies, Inc.

Andrea Laue
HighWire Press

John Meyer
ITHAKA

Dharitri Misra
National Library of Medicine

Nancy Murray
ITHAKA

Ira Polans
IEEE

Craig Rodkin
Association of Computing Machinery

Kathleen Sheedy
American Psychological Association

Amy Stout
Massachusetts Institute of Technology

Keith Wollman
Reed Elsevier

Note: Final rosters are pending approval of the Content and Collection Management Topic Committee.

Participants agreed that it would be important to move forward with a defined proposal to create Recommended Practices for Supplemental Materials under the NISO Recommended Practice publication series. **Given the scope of the problem, they agreed that the Working Group should be composed in three parts:**

1. **Stakeholders Interest Group** – a larger group to be kept apprised of development, to serve as a source of feedback on drafts, and to provide community vetting of a final document.
2. **Business Working Group** – a small group to draft recommendations related to the semantic aspects of the Recommended Practices. These include what constitutes supplemental materials, definitions, recommended roles, business practices, and policy questions.
3. **Technical Working Group** – a small group to look at the syntactic, structural issues, such as syntax, linking, interoperability, markup, and metadata.

The groups are now developing their mission statements and working plans. They will provide regular updates to the Stakeholders Group and interested parties through an e-mail discussion group on the NISO website.

Current Status

The NISO Content & Collection Management Topic Committee approved the proposal in late Spring 2010. The two small working groups have been formed with Linda Beebe from American Psychological Association and Marie McVeigh from Thomson-Reuters co-chairing the Business Working Group and David Martinsen from American Chemical Society and Sasha Schwarzman from American Geophysical Union co-chairing the Technical Working Group. The groups both include representatives from a broad spectrum across the scholarly information community.

The groups are now developing their mission statements and working plans. They will provide regular updates to the Stakeholders Group and interested parties through an e-mail discussion group on the NISO website. Interested parties can join the list by sending an e-mail to: suppinfo-subscribe@list.niso.org. Because the issues are thorny, it is likely to be sometime later in 2011 before Recommended Practices are finalized for acceptance. | NR | doi: 10.3789/isqv22n3.2010.07

LINDA BEEBE <lbeebe@apa.org> is Senior Director of PsycINFO, American Psychological Association.

RELEVANT LINKS

NFAIS. Best Practices for Publishing Journal Articles.
www.nfais.org/files/file/Best_Practices_Final_Public.pdf

NISO/NFAIS Supplemental Journal Materials Roundtable Report
www.niso.org/topics/tl/suppmatls/

OSA Interactive Science Publishing
www.opticsinfobase.org/isp.cfm

PARSE Project
www.stm-assoc.org/standards_and_technology_parse.php

Schwarzman, Alexander (Sasha). Supporting Material Survey Results.
www.agu.org/dtd/Presentations/sup-mat/

Schwarzman, Alexander (Sasha). Towards Formulating Criteria for Supplemental Material.
www.agu.org/dtd/Presentations/sup-mat/sup-mat-points.shtml

Michael Clarke

MICHAEL CLARKE

Society of Scholarly Publishing's 2010 Annual Meeting: Sustainability and Transition

The main take-away from the Society for Scholarly Publishing's (SSP's) 2010 Annual Meeting was that this is the year "sustainability" became a bad word. Usually, you don't see a whole conference turn against a single word. A few people in a session here or there might voice a concern, but to have what can only be described as a relentless and sustained assault across sessions, in and out of hallway and exhibit hall conversations, and across multiple days is truly remarkable. **I pity the word that gets on the bad side of an SSP conference.**

Powerpoint-Karaoke is a spin-off from the traditional Karaoke, however instead of singing songs, the participants must present an impromptu presentation based on a random presentation, projected on a screen, to an audience.

But what was all the fuss about really? Sustainability is such an unassuming word. It is quiet and keeps to itself. Some might say it is a bit milquetoast, but really is that cause for such hostility?

The problem is precisely its milquetoast quality. It implies that all one needs is to get by. All one needs is to break even, to come out without loss, to be no worse for wear. In short, it implies that milquetoast is OK.

And sometimes milquetoast just is not acceptable. Sometimes one needs a bit of what **Joe Espisito** would call reckless enthusiasm. Or at least something a bit ahead of breakeven.

The case against sustainability was first made by **Geoffrey Bilder** in his PowerPoint Karaoke session with **Kent Anderson**. What, you might ask is PowerPoint Karaoke. Here I will defer to Wikipedia:

> *Powerpoint-Karaoke is a spin-off from the traditional Karaoke, however instead of singing songs, the participants must present an impromptu presentation based on a random presentation, projected on a screen, to an audience.*

In the case of SSP, meeting attendees were invited to submit slides to the session moderators in advance. The moderators then showed the slides to Bilder and Anderson, neither of whom had seen them before. They were then prompted to extemporaneously talk, providing their perspective on the slide.

One of the slides was about sustainability. Bilder took issue with the notion that aiming to simply be sustainable was sufficient. Even not-for-profit organizations need to do more than simply sustain their business, to "grimly hold on," as he put it. Successful organizations need to generate surpluses so that they can experiment, invest, and improve. New technologies, new publication models, new products and services all require surpluses. Moreover, attracting and retaining talented staff requires more than scraping by. Bilder suggested that all organizations, commercial and not-for-profit, need to aim for "thriving" not merely sustainability.

We are just now beginning to rethink the rest of the box—the containers we are so familiar with, whether in electronic or print form: journals, books, articles, chapters, and even websites.

Implicit in Bilder's critique, and explicit in Joe Esposito's comments on sustainability during the closing session (the Scholarly Kitchen's "Food Fight" was covered extensively with a link provided below), is the notion that we are in a period of transition, where the need for experimentation and new development is paramount.

This theme of transition permeated the meeting and was probably best described by **John Sack** during his plenary talk, *Publishing in the Post-Web World*. Sack made the compelling case that the transition period we are entering is very different from the one we have just gone through. Over the last 15 years we have rethought the "distribution box" for scientific and scholarly content, with most books and journals now available electronically. However, the content itself—and the formats it is produced in—have scarcely changed for scholarly publishers. We are just now beginning to rethink the rest of the box—the containers we are so familiar with, whether in electronic or print form: journals, books, articles, chapters, and even websites.

The role of publishers is to help develop, curate, and distribute knowledge in whatever formats and media it is needed, not the production of specific containers. Sack urged publishers to consider this shift as an opportunity. As the center of gravity shifts from the Web to a more diverse array of communication tools and technologies, opportunities are emerging for publishers to create new products and services that continue to add value to the scholarly communication ecosystem.

Some of these emerging areas of opportunity were explored in more detail in other sessions. This author moderated a session on scientific applications, for example. The session included a presentation from **Steve Welch** of the American College of Chest Physicians and SiNae Pitts from Amphetamobile on SEEK, the ACCP's sleep medicine application for the iPhone and iPad. The interesting thing about SEEK was that much of the information in the application came from a print book. The information was updated and, in some cases, redeveloped for the new format, but the ACCP essentially jumped from a print book to a mobile application without developing a website for the content.

Keir Mierle from Google provided a demonstration in this same session on Astrometry.net. Astrometry.net is not a Google product but rather a collaborative, not-for-profit initiative that has provided both amateur and professional astronomers with a powerful new tool for identifying the position of images taken of the sky. Anyone can submit a photograph of the night sky to Astrometry.net and the site will identify the coordinates and objects in the image. It is able to do this by matching relational star positions in the image against positions of stars in its extensive database. Astonometry.net is therefore a database,

CONTINUED »

CONTINUED »

search engine, reference work, and workflow tool in one—a new scientific resource outside of the traditional containers.

The discussion of mobile applications continued in another session moderated by **Darrell W. Gunter** of Collexis. This session included overviews by **John Barker** of Wolters Kluwer, **Daniela Barbosa** of Dow Jones, and **Kim Murphy** of Elsevier on the range of mobile applications offered by their respective companies. Barker described the two development paths at Wolters Kluwer. The first is a top-down approach where applications are developed via a centralized technology group. The second is a bottom-up approach whereby local offices develop narrow applications for niche (often geographically defined) audiences (e.g., architects in Germany). Barker also described how Wolters Kluwer was careful not to repeat common "shovelware" mistakes made by publishers in transitioning to the Web a decade and a half ago. By this he meant that one cannot simply shovel content developed for other purposes and formats into an application and expect it to work. Content must be carefully selected, and in some cases updated or redeveloped, for the purpose at hand.

The session titles below are, to my mind, representative of the conversations, experiments, and investigations going on throughout the industry. I was unfortunately not able to attend all these sessions but look forward to catching up as they are posted on the SSP website.

One presentation I did see but will definitely revisit is **Brewster Kahle**'s keynote. If his presentation was perhaps a bit off-topic from the perspective of content transition, he can be excused. Kahle, founder of the Internet Archive and the Open Content Alliance, described his vision for a distributed

The re-defining of content containers stuck out as a theme, clearly visible in the titles of many of the meeting presentations:

- » Connecting Content: An Integrated Approach to Electronic Materials
- » Rich Media: Projects and Prospects
- » The Future Scientist: Will We Provide the Right Tools for this New Generation of Researchers? The Future of Reference Publishing
- » Re-inventing Reference: This is NOT Your Grandmother's Encyclopedia!
- » From a Production Industry to a Technology Industry
- » Semantic Technology: New Tools & New Rules for Search+Visualization

electronic book dissemination system (Sack's "distribution box") that does not rely on a few proprietary mega-stores. Ebooks could be purchased, rented, "checked-out" (in the case of library copies), or freely accessed (in the case of public domain works) using a distributed content-finding system called BookServer. The idea is that a search for a book via BookServer will return all the options and the user can then select from a menu of accessing and format options (e.g., PDF, ePub, Mobi, XML, etc.).

To quote from the BookServer website:

> As the audience for digital books grows, we can evolve from an environment of single devices connected to single sources into a distributed system where readers can find books from sources across the Web to read on whatever device they have. Publishers are creating digital versions of their popular books, and the library community is creating digital archives of their printed collections. BookServer is an open system to find, buy, or borrow these books, just like we use an open system to find Web sites.

The BookServer has already been built and many public domain works are now available. The question is whether BookServer will gain traction with libraries, publishers, booksellers, and book readers.

The landscape for the last 15 years has been familiar. Yes, there have been revolutions in the way scholarly content is distributed but the content itself has remained fairly stable. Now we are headed into uncharted waters—blank spots on the map. What is beyond? Who is to say? The interesting thing about the future is that it is unknown. I think it is safe to say, however, that the future will bring challenges and opportunities—opportunities not merely to sustain the status quo but to thrive in an evolving information landscape.

| CR | doi: 10.3789/isqv22n3.2010.09

MICHAEL CLARKE <michael@clarkepublishing.com> is the founder of Clarke Publishing Group <clarkepublishing.com>, a consultancy that helps scholarly, professional, and society publishers navigate today's rapidly changing information landscape. He is a frequent blogger on SSP's Scholarly Kitchen site <scholarlykitchen.sspnet.org>.

Now we are headed into uncharted waters—blank spots on the map. What is beyond? Who is to say? The interesting thing about the future is that it is unknown.

RELEVANT LINKS

Anderson, Kent. Publishing Through the Lens of Sustainability, Quality, and Chaos. The Scholarly Kitchen
scholarlykitchen.sspnet.org/2010/06/07/tackling-tomorrow-publishing-through-the-lenses-of-sustainability-quality-and-chaos/

Astrometry.net
www.astrometry.net

BookServer
www.archive.org/bookserver

Esposito, Joe. Let's Here it for Reckless Enthusiasm! The Scholarly Kitchen
scholarlykitchen.sspnet.org/2010/01/01/lets-hear-it-for-reckless-enthusiasm/

PowerPoint Karaoke
en.wikipedia.org/wiki/Powerpoint-Karaoke

Wills, Stewart. Live Blogging the Food Fight. The Scholarly Kitchen
scholarlykitchen.sspnet.org/2010/06/04/live-blogging-the-food-fight/

KBART Recommended Practice Endorsed by Content Providers

Following the January 2010 release of the joint NISO and UKSG recommended practice on Knowledge Bases And Related Tools (KBART), a number of major organizations in the scholarly information supply chain have publicly endorsed the recommendations by confirming that their systems can process KBART formatted files.

The KBART Recommended Practice (NISO RP-9-2010) contains practical recommendations for the timely exchange of accurate metadata between content providers and knowledge base developers. All content providers, from major databases to small publishers, are encouraged to publicly endorse the KBART Recommended Practice by submitting a sample file to the KBART working group. Once the file's format and content has been reviewed and approved, and the provider has made it publicly available (in line with the recommendations), the provider will be added to a public list of endorsing providers.

The organizations currently endorsing KBART are Alexander Street Press, American Institute of Physics, Annual Reviews, EBSCO, Ex Libris, Innovative Interfaces, Inc., Serials Solutions, and OCLC. Further information on endorsement is available from the KBART Information Hub.

The KBART working group is now undertaking Phase II, with a largely new set of volunteers. Sarah Pearson (University of Birmingham) the UKSG co-chair has been joined by Andreas Biedenbach (Springer Science+Business Media) as the NISO co-chair. Whereas the Phase I report provided minimum recommendations to improve knowledgebases, the Phase II recommendations will focus on the more advanced, complex issues that cause problems in this area, as defined in the "Next Steps" portion of the Phase I report. ■

RELEVANT LINKS

KBART Recommended Practice
www.uksg.org/kbart/s1/summary
www.niso.org/publications/rp/RP-2010-09.pdf

KBART Working Group
www.niso.org/workrooms/kbart/

KBART Information Hub
www.uksg.org/kbart/hub

International DOI Foundation Launches shortDOI™ Service

A shortDOI™ service has been made publicly available by the International DOI Foundation to create shortcuts for long DOI® names. A valid DOI string can be input to the online form at http://shortdoi.org/ and, upon submittal, a shortDOI will be assigned (or displayed if one had already been assigned). The shortDOI can then be used in e-mail, blogs, tweets, etc. to provide links that will resolve exactly the same as the full DOI string. ■

Example:

Full DOI string: 10.3789/isqv21n1
shortDOI: 10/afy
Shortcut URL: http://doi.org/afy

ISO International Standard Name Identifier Approved for Publication

Members of the International Organization for Standardization (ISO) Technical Committee 46 on Information and documentation, Subcommittee 9 on Identification and description have approved the publication of a new standard for an International Standard Name Identifier (ISNI).

ISO 27729:2010 specifies an identifier for the public identities of parties to disambiguate their identities in a digital environment and to support the management and exchange of media content. The ISNI, which consists of 15 digits and a check character displayed in blocks of four numbers, has no meaning built into the number.

TC46/SC9 has asked the ISO Technical Management Board to negotiate a contract with the newly created International ISNI Agency to be the Registration Authority (RA) for ISO 27729. The ISNI RA will be responsible for allocating unique ISNIs for a public identity, as requested, and maintaining a database of the registrant data, metadata, and administrative data associated with each ISNI.

Although the final required metadata will be determined by the ISNI RA, the standard cites as a minimum: name of public identity, type of party, and at least one of external data link, creation class, or role.

An informative annex to the standard describes the relationship of the ISNI to other identifiers and describes ISNI as a "bridge" identifier that will allow industry partners "to exchange information related to a party without the need to disclose confidential information."

The ISNI standard is expected to be published by the end of 2010. ■

W3C Launches Library Linked Data Incubator Group

The World Wide Web Consortium has announced the creation of a Library Linked Data Incubator Group, whose mission is to help increase global interoperability of library data on the Web, by bringing together people in the library community and beyond who are involved in Semantic Web activities focused on Linked Data. The group will explore how existing building blocks of librarianship, such as metadata models, metadata schemas, standards and protocols for building interoperability and library systems and networked environments, encourage libraries to bring their content, and generally re-orient their approaches to data interoperability towards the Web.

The groups charter was sponsored by Helsinki University of Technology, DERI Galway, Competence Centre for Interoperable Metadata (KIM), Library of Congress, Los Alamos National Laboratory, MIMOS, OCLC, Talis, University of Applied Sciences Potsdam, and Vrije Universiteit Amsterdam in recognition of the need for a shared standardization effort within the library community around Semantic Web standards and with the intent of involving stakeholders beyond libraries including cultural heritage institutions, partners from the publishing industry, and other relevant domains.

The group's chairs are Tom Baker (Dublin Core Metadata Initiative), Emmanuelle Bermes (Bibliothèque Nationale de France), and Antoine Isaac (Vrije Universiteit Amsterdam). A report presenting the landscape of Linked Data development in the library domain and related sectors and proposing a way forward is targeted for May 2011. ■

For more information, visit: www.w3.org/2005/Incubator/lld/workrooms/i2

Mellon Grant Awarded to IU and NISO for MESUR Sustainability

A $349,000 grant from The Andrew W. Mellon Foundation to Indiana University Bloomington will fund research to develop a sustainable initiative to create metrics for assessing scholarly impact from large-scale usage data.

Indiana University Bloomington School of Informatics and Computing associate professor Johan Bollen and NISO will share the Mellon Foundation grant designed to build upon the MEtrics from Scholarly Usage of Resources (MESUR) project that Bollen began in 2006 with earlier support from the foundation.

The new funding for *Developing a Generalized and Sustainable Framework for a Public, Open, Scholarly Assessment Service Based on Aggregated Large-scale Usage Data* will support the evolution of the MESUR project to a community-supported, sustainable scholarly assessment framework.

MESUR has already created a database of more than 1 billion usage events with related bibliographic, citation, and usage data for scholarly content.

The project will focus on four areas in developing the sustainability model—financial sustainability, legal frameworks for protecting data privacy, technical infrastructure and data exchange, and scholarly impact—and then integrate the four areas to provide the MESUR project with a framework upon which to build a sustainable structure for deriving valid metrics for assessing scholarly impact based on usage data. Simultaneously, MESUR's ongoing operations will be continued with the grant funding and expanded to ingest additional data and update its present set of scholarly impact indicators.

Data from more than 110,000 journals, newspapers, and magazines, along with publisher-provided usage reports covering more than 2,000 institutions, is being ingested and normalized in MESUR's databases, resulting in large-scale, longitudinal maps of the scholarly community and a survey of more than 40 different metrics of scholarly impact. ■

For information on MESUR, visit: www.mesur.org/

The new funding will support the evolution of the MESUR project to a community-supported, sustainable scholarly assessment framework.

NFAIS and ASIDIC Announce Merger

The National Federation of Advanced Information Services (NFAIS) and ASIDIC (formerly the Association of Information and Dissemination Centers) announced on June 30, 2010 the assimilation of ASIDIC members into the NFAIS Community as a result of the dissolution of ASIDIC.

"Since its founding in 1968 ASIDIC has been true to its mission and has served its members well," said ASIDIC President, Tim Ingoldsby. "But as the Board of Directors looked forward in these changing times, we determined that it was in the best interest of our members to recommend the dissolution of ASIDIC and to identify an organization with a similar mission that could not only serve our members, but also continue the spirit that has sustained ASIDIC throughout its history. After a call for proposals and several months of discussions with interested parties, I am pleased to announce that the Board, supported by a majority vote of its member organizations, has partnered with NFAIS to reach our objective."

According to NFAIS Executive Director, Bonnie Lawlor, the agreement will provide ASIDIC members in good standing with all NFAIS member benefits through June 30, 2011, after which they will have the option to continue their NFAIS membership through a three-year transition period. Any ASIDIC assets remaining after the merger will be used as requested by the ASIDIC Board and that is to attract conference speakers to NFAIS events that are in keeping with the mission and spirit of ASIDIC. In addition, a member of an ASIDIC Member organization will be invited to serve as a non-voting NFAIS Board member from the close of the merger through June 30, 2011 in order to assist the NFAIS Board in a successful membership transition and to ensure the preservation of the ASIDIC spirit and mission within NFAIS. ■

For more information, visit the NFAIS website: www.nfais.org

Editor's Note: NFAIS and ASIDIC are both NISO voting members.

Institutional Identifiers Working Group Releases Midterm Report

NISO's Institutional Identifier (I²) Working Group released a midterm report summarizing the findings and assumptions from the first phase of their work.

Charged with developing a robust, scalable, and interoperable standard for identifying the institution, a core entity in any information management or sharing transaction, the I² working group focused the first phase of their work in creating workflow scenarios. Over 300 stakeholders in the commercial supply chain, libraries, and institutional repositories were interviewed, confirming the need for the standard and providing feedback on draft metadata.

The initial investigative work done by the I² Working Group validated that a global institutional identifier is critical to enable information supply chain business scenarios. The environment envisioned by this working group is similar to what is being planned for the International Standard Name Identifier (ISNI) standard (ISO 27729) that will be published shortly by ISO (the International Organization for Standardization). As with ISNI, the I² identifier is proposed to be managed through a central registry that would assign identifiers to institutions, store core metadata about those institutions, and provide look-up services. A decentralized series of business-specific registries relying on the central core registry would be expected to provide expanded information about an institution as required by their business; for example, a registry that is supporting the access services for an institution may choose to add IP addresses as well as information about link resolvers.

The report identifies and describes the following six key attributes of the I²:

1. Identify organizations
2. Be opaque
3. Support the creation of a core metadata set that describes an institution
4. Support registration of institutions in a decentralized manner
5. Address community-specific registry needs
6. Allow URI(s) from 3rd-party registries to be submitted and stored

A number of existing identifiers were examined as part of this committees work. These included: the International Standard Name Identifier (ISO 27729), the MARC Code List for Organizations, the Standard Address Number (ANSI/NISO Z39.43), and the Dun & Bradstreet DUNS number. Each of these was analyzed with respect to the key attributes required for the I², with ISNI being the only standard that met all six attributes.

A set of metadata elements and subelements was proposed with [institution] name, location (country and city), contact information, and type [of institution] proposed as mandatory metadata.

Six workflows were depicted in the report to illustrate the purpose of the institutional identifier and provide a compelling rationale for its development and use:

1. Library wants to be a member of a consortium
2. Library subscribes to an electronic journal
3. Library A places ILL requests with Library B
4. Library places an ILL request via consortium
5. Library places an ILL request that has special circumstances
6. A regional electronic theses and dissertations repository harvests metadata from a participant repository

Public comments were solicited on the midterm report and the feedback is currently being compiled. The next release by the working group will be a draft standard for comment that will build on the feedback received to the mid-term report. ■

Visit the I² working group webpage at: www.niso.org/workrooms/i2

E-Science and Data Support Services: A Study of ARL Member Institutions

The Association of Research Libraries (ARL) has published a report from a 2009 survey of their member institutions on how libraries can contribute to e-science activities in their organizations. ARL's E-Science Task Force, launched in 2006, defines the domain of e-science as "those new methods that are large-scale, data driven, computationally intense, and often engaging research teams across institutional boundaries." A working group of the Task Force, chartered to develop a better understanding of the changing requirements to support e-science-based research, conducted the survey.

Catherine Soehner, Catherine Steeves, and Jennifer Ward, the authors of *E-Science and Data Support Services*, selected six respondents to highlight as case studies: Purdue University; University of California, San Diego; University of Illinois at Chicago; and the Massachusetts Institute of Technology.

Catherine Soehner, Catherine Steeves, and Jennifer Ward, the authors of *E-Science and Data Support Services*, selected six respondents to highlight as case studies: Purdue University; University of California, San Diego; University of Illinois at Chicago; and the Massachusetts Institute of Technology.

AMONG THE FINDINGS ARE:

45% of respondents have some type of designated unit for data curation and support for scientific research data.

A third of respondents have conducted an assessment about what types of services their researchers need.

73% of respondents' libraries are involved in e-science support with 48% reporting a team or committee leadership approach.

Library services provided include consultation services, reference, managing technology such as servers for data storage, and finding or developing tools to assist researchers.

The top three pressure points in providing e-science support are lack of resources, difficulty in acquiring appropriate staff and expertise, and a lack of unifying direction in the organization.

When reassigning staff, the MLIS degree was held by a majority of the position-holders, with science degrees or expertise also highly valued.

Collaboration was a key to success and took place on many levels: "between libraries of different institutions, between libraries and the departments they serve, between various departments to address interdisciplinary subject areas, and between institutions."

The authors conclude that "engagement by research libraries in e-science has been developing rapidly in the past few years," that "institutions are quickly rising to meet the challenge of managing data," and they are exhibiting a "great diversity in the strategies employed."

The report is available for free download from: www.arl.org/bm~doc/escience_report2010.pdf

Library of Congress Launches National Digital Stewardship Alliance

The Library of Congress announced on August 3, 2010 the formation of the National Digital Stewardship Alliance (NDSA), a partnership of institutions and organizations dedicated to preserving and providing access to selected databases, web pages, video, audio, and other digital content with enduring value.

The alliance is an outgrowth of the National Digital Information Infrastructure and Preservation Program (NDIIPP), which the Library has administered since 2000. In establishing the program, Congress directed the Library to work with other federal agencies and a variety of additional communities to develop a national approach to digital preservation. NDIIPP has achieved substantial success though partnering with more than 170 institutions to provide access to a diverse national collection of digital content.

The NDSA will build on this accomplishment by focusing on several goals. It will develop improved preservation standards and practices; work with experts to identify categories of digital information that are most worthy of preservation; and take steps to incorporate content into a national collection. It will provide national leadership for digital-preservation education and training. The new organization will also provide communication and outreach for all aspects of digital preservation.

The NDSA will launch with a core set of founding members drawn from current NDIIPP project partners. Those members will develop a roadmap for immediate action, including a process for expanding membership. |NW|

For more information, visit www.digitalpreservation.gov/ndsa/

The NDSA will build on this accomplishment by focusing on several goals. It will develop improved preservation standards and practices; work with experts to identify categories of digital information that are most worthy of preservation; and take steps to incorporate content into a national collection.

SD [STANDARDS IN DEVELOPMENT: *August 1, 2010*]

In Development or Revision

Listed below are the NISO Working Groups that are currently developing new or revised standards, recommended practices, or reports. Refer to the NISO website (www.niso.org/workrooms/) and *Newsline* (www.niso.org/publications/newsline/) for updates on the Working Group activities.

WORKING GROUP	STATUS
Cost of Resource Exchange (CORE) Co-chairs: Ed Riding, Ted Koppel	**NISO RP-10-2010, Cost of Resource Exchange (CORE) Protocol** Following DSFTU period, approved for publication as a NISO Recommended Practice instead of a standard.
DAISY/NISO Standard Advisory Committee Chair: George Kerscher	**Z39.86-201x, Specifications for the Digital Talking Book** Standard revision in development. *Part A, Authoring and Interchange Framework*, issued for public comment.
E-Journal Presentation & Identification Co-chairs: Cindy Hepfer, Steve Shadle	Recommended Practice in development.
ERM Data Standards & Best Practices Review Co-chairs: Ivy Anderson, Tim Jewell	Technical Report in development.
ESPReSSO: Establishing Suggested Practices Regarding Single Sign-On Co-chairs: Steve Carmody, Harry Kaplanian	Recommended Practice in development.
Institutional Identifiers (I^2) Co-chairs: Grace Agnew, Oliver Pesch	**Z39.94-201x, Institutional Identifiers** Standard in development. Midterm Work to Date document released for public comment.
IOTA: Improving OpenURLs Through Analytics (formerly OpenURL Quality Metrics) Chair: Adam Chandler	Technical Report in development.
Knowledge Base and Related Tools (KBART) Phase II *Joint project with UKSG* Co-chairs: Andreas Biedenbach, Sarah Pearson	**NISO RP-9-2010, KBART: Knowledge Bases and Related Tools** Phase I Recommended Practice issued January 2010. Phase II Recommended Practice in development.
Physical Delivery of Library Materials Co-chairs: Valerie Horton, Diana Sachs-Silveira	Recommended Practice in development.
RFID for Library Applications Revision Co-chairs: Vinod Chachra, Paul Sevcik	**NISO-RP-6-201x, RFID in U.S. Libraries** Revision in development.
Standardized Markup for Journal Articles Co-chairs: Jeff Beck, B. Tommie Usdin	**Z39.96-201x, Standardized Markup for Journal Articles** Standard in development.
Supplemental Journal Article Materials Co-chairs Business Working Group: Linda Beebe, Marie McVeigh Co-chairs, Technical Working Group: Dave Martinsen, Alexander (Sasha) Schwarzman	Recommended Practice in Development.

2010 NISO EDUCATIONAL EVENTS

SEPTEMBER

Two-Part Webinar: Measuring Use, Assessing Success

8 Measure, Assess, Improve, Repeat: Using Library Performance Metrics

15 Count Me In: Measuring Individual Item Usage

OCTOBER

7 **In-Person**
E-Resource Management: From Start to Finish (and Back Again)
Chicago, IL

13 **Webinar**
It's Only as Good as the Metadata: Improving OpenURL and Knowledgebase Quality

NOVEMBER

10 **Webinar**
The Case of the Disappearing Journal: Solving the Title Transfer and Online Display Mystery

DECEMBER

8 **Webinar**
Unprecedented Interaction: Providing Accessibility for the Disabled

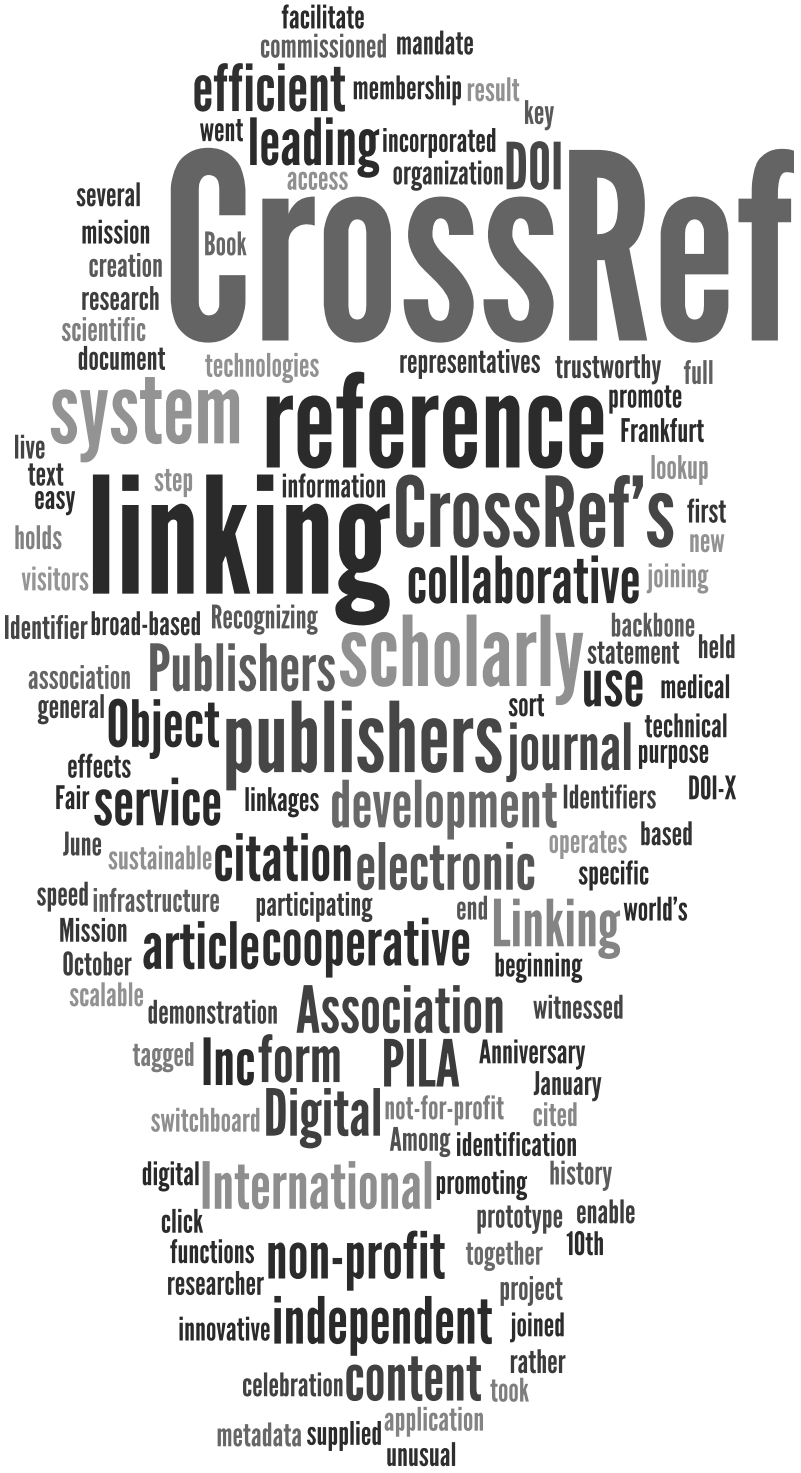
NISO Open Teleconferences

Join NISO on our free monthly conference calls to discuss projects underway in NISO and to provide the organization with feedback and input on areas where NISO is or ought to be engaged. NISO teleconferences are held from 3:00-4:00 p.m. (Eastern time) on the second Monday of each month (except July). To join, simply dial 877-375-2160 and enter the code: 17800743.

Webinar Subscription Package Discounts

Buy 4 get 2 free.
Buy 6 get 7 free.

www.niso.org/news/events



crossref.org