

INFORMATION STANDARDS QUARTERLY

SUMMER 2014 | VOL 26 | ISSUE 2 | ISSN 1041-0031

# ISQ

TOPIC

## OPEN ACCESS INFRASTRUCTURE

---

OA INFRASTRUCTURE: WHERE WE ARE  
AND WHERE WE NEED TO GO

---

THE ROLE OF STANDARDS IN  
THE MANAGEMENT OF OA RESEARCH  
PUBLICATIONS: A Research Library Perspective

A PUBLISHER'S PERSPECTIVE ON  
THE CHALLENGES OF OPEN ACCESS

---

THE NEED FOR RESEARCH  
DATA INVENTORIES AND THE  
VISION FOR SHARE

CHORUS HELPS DRIVE PUBLIC ACCESS

---

STANDARDIZED METADATA  
ELEMENTS TO IDENTIFY ACCESS  
AND LICENSE INFORMATION

**NISO**  
How the information world  
CONNECTS



# USING THE WEB AS AN E-CONTENT DISTRIBUTION PLATFORM: CHALLENGES AND OPPORTUNITIES

## About the Forum

Web technologies have changed, and continue to change, the way that content is delivered to libraries and to users. Currently e-journals are delivered through platforms but the final object is still often a PDF file. E-books are generally delivered as a downloadable file to a stand-alone e-reader. However, with the advent of the Open Web Platform using standards such as HTML5, we are looking at a new era of separating the content from the container.

The web as a distribution platform offers many new opportunities for more utilization of multimedia and streaming media, embedding of apps, increased linkages and interoperability between related content, greater interactivity with content, social sharing of user-generated content related to a "publication," text mining, and much more that hasn't even been imagined yet.

Join NISO for this two-day forum where speakers will discuss the current state, future trends, and implementation issues from the perspectives of both libraries and publishers.

**OCTOBER 21 - 22, 2014**

Gleacher Center - Booth School of Business,  
The University of Chicago, Chicago, IL

*Available in-person  
and by livestream*

With the generous support of SAGE, NISO is able to make the livestream of the Opening and Closing Keynote presentations **FREE**. Virtual space is limited for this offer, so interested participants must RSVP. Access to the full livestream of the two-day event requires paid registration.



FOR MORE INFORMATION AND TO REGISTER VISIT:  
[www.niso.org/news/events/2014/forum\\_2014/](http://www.niso.org/news/events/2014/forum_2014/)

Livestream Sponsor:



Supporting Sponsors:



**OCLC**<sup>®</sup>

The world's libraries.  
Connected.™



# ISQ

SUMMER 2014 | VOL 26 | ISSUE 2 | 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 EXECUTIVE DIRECTOR and ISQ PUBLISHER | Todd Carpenter

ISQ MANAGING EDITOR | Cynthia Hodgson

NISO ASSOCIATE DIRECTOR FOR PROGRAMS | Nettie Lagace

NISO MEMBER SERVICES AND ENGAGEMENT MANAGER | DeVonne Parks

NISO EDUCATIONAL PROGRAMS MANAGER | Juliana Wood

DESIGN | B. Creative Group Inc.

### NISO Publications Committee Members

Marshall Breeding, *Vanderbilt University*

Priscilla Caplan, *Florida Center for Library Automation*

Liam Earney, *Joint Information Systems Committee (JISC)*

Corey Harper, *New York University, Division of Libraries*

Sheila Morrissey, *ITHAKA*

Peter Murray, *LYRASIS*

Andrew Pace, *OCLC*

Kristen Ratan, *Public Library of Science (PLoS)*

Susan Stearns, *Boston Library Consortium*

Elizabeth Winter, *Georgia Institute of Technology Libraries*

### 2014 AD RATES

	1 ISSUE	2-3 ISSUES	4 ISSUES
Full page (8.5" x 11")	\$375	\$350	\$325
Half page (4.25" x 5.5")	\$250	\$225	\$200
Back cover (8.5" x 11")	\$700	\$600	\$550
Inside Cover Front (8.5" x 11")	\$500	\$450	\$400
Inside Cover Back (8.5" x 11")	\$500	\$450	\$400

For more information on advertising, visit [www.niso.org/publications/isq](http://www.niso.org/publications/isq)

©2014 National Information Standards Organization.

REUSE: For permission to photocopy or use material electronically from Information Standards Quarterly, ISSN 1041-0031, please access [www.copyright.com](http://www.copyright.com) or contact Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users.

## CONTENTS

3 From the Guest Content Editor

### FE FEATURE 4

4 Open Access Infrastructure:  
Where We Are and Where We Need to Go

### IP IN PRACTICE 15

15 The Role of Standards in the Management of  
Open Access Research Publications: A Research  
Library Perspective

22 A Publisher's Perspective on the Challenges  
of Open Access

### PI PROJECT INITIATIVES 29

29 The Need for Research Data Inventories and the  
Vision for SHARE

32 CHORUS Helps Drive Public Access

### NR NISO REPORTS 35

35 Standardized Metadata Elements to Identify  
Access and License Information

### NW NOTEWORTHY 38

38 NISO Publishes Three Recommended Practices  
on Knowledge Bases, Demand Driven Acquisition  
of Monographs, and Library Discovery Services

39 NISO and OAI Publish American National Standard  
on ResourceSync Framework Specification

39 Epub 3.0.1 Issued by International Digital  
Publishing Forum

40 Library of Congress Identifies Recommended Formats  
for Long-Term Preservation

40 UKSG Transfer Working Group Announces  
Improvements to the Code of Practice with Release  
of Version 3.0

### SD STANDARDS IN DEVELOPMENT 41



# NISO UPCOMING 2014 EDUCATIONAL EVENTS

## SEPTEMBER

NISO Two-Part Webinar:  
E-books for Education

- 10** **Part 1:** Electronic Textbook:  
Plug in and Learn
- 17** **Part 2:** Open Textbook Initiatives
- 24** Library Data in the Cloud  
(*Virtual Conference*)

## OCTOBER

- 15** 21st Century Resource  
Sharing: Which Inter-Library  
Loan Standard Should I Use?  
(*Webinar*)
- 20-21** Using the Web as an  
E-Content Distribution  
Platform: (*In-Person  
Forum with Live Streaming*)

## NOVEMBER

- 12** Keyword Search = "Improve  
Discovery Systems" (*Webinar*)
- 19** Can't We All Work Together?  
Interoperability & Systems  
Integration (*Virtual Conference*)

## DECEMBER

- 3** Connecting the Library  
to the Wider World:  
Successful Implementations  
of Linked Data (*NISO/NFAIS  
Virtual Conference*)

NISO Two-Part Webinar:  
Sustainable Information

- 10** **Part 1:** Digital Preservation  
for Text
- 17** **Part 2:** Digital Preservation  
of Audio-Visual Content

### Webinar Subscription Package Discounts

- » Buy 5 NISO webinars  
& get 4 free
- » Buy 9 NISO webinars  
& get 5 free - attend  
the whole series
- » Buy 4 NISO virtual  
conferences & get 2 free

### NISO Open Teleconferences

Join us each month  
for NISO's Open  
Teleconferences—an  
ongoing series of calls  
held on the second  
Monday of each month  
as a way to keep the  
community informed of  
NISO's activities.  
The call is free.





Liam  
Earney

## LETTER FROM THE GUEST CONTENT EDITOR

### 2013: THE YEAR OPEN ACCESS BROKE<sup>1</sup>

It's a pleasure to be introduce this issue of *Information Standards Quarterly* on Open Access Infrastructure. When we were considering this issue, we were very aware that we didn't want to revisit previous arguments about open access (OA), but rather take as our starting point the fact that 2013 seemed to have been a watershed for open access. Driven by a number of policy announcements from funding bodies and governments worldwide, the question is no longer whether open access will or should happen, but rather how will it be implemented in a sustainable way.

The articles that follow contain a wealth of insights from a wide variety of viewpoints—publishers, funders, universities, intermediaries, standards bodies, and open access experts. They were selected with the intention of providing the uninitiated (which is the majority) with an overview of where we are today, what the challenges are, available routes to overcoming those challenges, and some of the initiatives that have been put in place to overcome these challenges.

What comes across strongly from these articles is the complexity and interdependency of the issues that we face. When approaching the authors, I was struck by how wary many of them were about authoring an article, largely on the basis that they felt they lacked the necessary expertise or breadth of knowledge to do the topics justice. My personal opinion is that this is such a dynamic and novel area of work that such an attitude is a strength rather than a weakness.

While Sage's David Ross is speaking about the publishers' perspective, his comment, "that the uncertain and ever-shifting global framework presents publishers with unique challenges with respect to long-term strategic planning, shorter-term policy decisions, and the development of infrastructure and workflow solutions to support these," could just as easily apply to all of the other participants in this area of activity.

What also comes across strongly is the importance that all the authors place on the development and adoption by everyone involved of standards-based approaches to overcoming the challenges for a sustainable open access infrastructure. Kaufman and Goodrich are right to note that "many publishers and institutions are still struggling to set up working systems and processes to support OA workflow for APCs and licensing," though Moyle, Sharp, and Bracey

summarize the current state rather more bluntly: "Stakeholders within the sphere of OA research publication do not benefit from the effective standardization of even the relatively few key pieces of information that support commonplace OA interactions." There is clearly much to be done.

The final key theme that comes across is that of collective responsibility by all stakeholders to adopt/implement standards and engage in the various initiatives that are in place globally. Without such collective action, we are unlikely to move on from the hand-to-mouth, spreadsheet-based approach that characterizes far too many interactions between the various stakeholders.

Does this matter? Aren't there myriad examples already of standards that have failed to lift off and doesn't scholarly communication persevere regardless? Well maybe, but I would suggest there has rarely been such high-level institutional and political interest globally in the publication and dissemination of scholarly research. If we are unable to demonstrate our capability to meet the requirements of those who fund so much of this research, then we run the risk of marginalization.

I would like to thank all of the authors for taking the time to contribute some of their hard pressed time towards this issue, while they are all exceptionally busy dealing with the issues and challenges highlighted so eloquently in their articles.

doi: 10.3789/isqv26no2.2014.01

Liam Earney | Head of Library Support Services, Jisc

<sup>1</sup> With inspirational thanks to David Markey, Director of 1991: *The Year Punk Broke*.

# OPEN ACCESS

## INFRASTRUCTURE:

WHERE WE ARE AND WHERE WE NEED TO GO

CYNTHIA HODGSON

---

Although the open access movement can be traced back to the late 1980s and early 1990s, many consider the Budapest Open Access Initiative in February 2002, the Bethesda Statement on Open Access Publishing in June 2003, and the Berlin Declaration on Open Access in October 2003 as the tipping points for the movement.

---





**T**he number of institutions and funders issuing policies regarding the availability of their research in some form of open access (OA) grew from one in 2003 to over 350 by the end of 2013, according to ROARMAP (Registry of Open Access Repositories Mandatory Archiving Policies).

There's no doubt that open access is here to stay, but the underlying infrastructure needed to support and sustain OA publishing is still very much in its development stages. Systems and services are in early stages of adoption with little interoperability between them. Some needed standards like ISSN and DOI are widely, though not universally, used, while others such as ISNI and ORCID are just beginning to be adopted. Additional needed standards in the areas of metadata, APIs, and protocols are either in discussion stages or not yet even envisioned.

This article, through a series of interviews with experts in the OA arena, highlights some of the major areas of infrastructure that are needed including institutional policies, compliance tracking and reporting, publishing tools, new economic models and licensing, and sustainability.




---

## INSTITUTIONAL POLICES FOR OPEN ACCESS

---

**Peter Suber** | *Director of Office for Scholarly Communication (Harvard Library) and Director, Harvard Open Access Project (Berkman Center), Harvard University*

The ten-year anniversary statement of the Budapest Open Access Initiative reaffirmed the two strategies of OA through repositories (also called “green OA”) and OA through journals (also called “gold OA”). Additionally, Recommendation 4.2 stated, “We should develop guidelines to universities and funding agencies considering OA policies, including recommended policy terms, best practices, and answers to frequently asked questions.” A month later, the first public edition of *Good Practices for University Open-Access Policies*, which had already been in development for several years, was released by the Harvard Open Access Project.

The Good Practices guide was based on the type of policy first adopted at Harvard, which asked faculty to deposit scholarly articles in the university’s institutional repository DASH (Digital Access to Scholarship at Harvard). Additionally, researchers grant the university a nonexclusive, irrevocable right to distribute their scholarly articles for any non-commercial purpose. This ensures that the repository can distribute the articles and does not have to track down rights or have different rights for different articles—a common problem with many institutional repositories. While there is a provision for obtaining a waiver regarding these rights, fewer than five publishers systematically require such a waiver as a prerequisite to publication.

Harvard researchers are free to publish articles in any journal of their choice. The policy is strictly about green OA; researchers are not required to choose gold OA journals for their publication. Commercial, subscription-based publications are equally acceptable. However, the university does want to encourage OA publishing and hosts a fund to pay the APCs for publication in fee-based OA journals—as long as they aren’t hybrid. Hybrid journals rarely reduce their subscription fees even when receiving APC fees for selected OA articles, which would mean the university is paying twice for the same content.

CONTINUED »



# GLOSSARY

---

**GREEN OA** / The archiving of a scholarly publication for public access in a repository other than that of the publisher, e.g., an institutional repository (IR) or discipline-related repository service. The deposited version is usually the final manuscript accepted for publication, but may not be the version that includes the publisher's final design and format. Also referred to as *open access archiving*.

---

**HYBRID JOURNAL** / A journal where some articles are available in open access while others are available only by payment (individually or by subscription).

---

**MANDATE (OPEN ACCESS)** / A requirement by an institution, funding agency, or government body that published research outcomes be available in some type of open access (green or gold). Mandates may dictate additional requirements regarding acceptable reuse licensing.

---

**EMBARGO** / A requirement by the publisher of record wherein a green repository deposit must be delayed for some period following the official publication.

---

## CC BY LICENSE

A Creative Commons Attribution license that allows the content to be shared and adapted for any purpose, including commercial, providing appropriate credit to the creator(s) is given.

## OPEN ACCESS

Unrestricted, online access to a scholarly publication that is free to read (*gratis*), and may have additional free reuse rights (*libre*).

## GOLD OA

The publication of a scholarly article in open access in a journal, usually peer-reviewed, and financed through article publication charges.

## INSTITUTIONAL REPOSITORY (IR) /

A database of content that contains, among other things, copies of the research output of authors. Repositories can be institution-based (representing the broad output of an institution), subject-based (representing the output of specific or related subjects), funder-based (representing the output of a funding agency, such as the NIH) or national (representing the output of a country or geographical region). Repositories can hold published or unpublished articles, presentations, datasets, and/or metadata about them.

---

## ARTICLE PUBLICATION CHARGE (APC) /

A fee paid to the publisher—usually by the author, author's institution, or funding agency—to make an article available in open access. Essentially shifts the cost of production from the subscriber to the author. Also referred to as *article processing charge*.

Implementation of the green OA repository by the Office for Scholarly Communication (OSC) is dependent on a number of standards. The Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH) ensures the content is discoverable and searchable. SWORD (Simple Web-service Offering Repository Deposit) provides interoperability between repositories. OSC references the publisher's DOI for the published article, if one exists, and also uses the DOI to look up and ingest relevant metadata about the article. Researchers are encouraged to obtain an ORCID identifier and associate it with all their publications. The primary article format is PDF, which is what most publishers still use, although OSC would prefer well marked-up XML or even HTML. Tools to convert from PDF into XML are not yet reliable enough and require substantial manual intervention. However, when the tools are better, the Harvard repository will add buttons to let users convert deposited PDFs to other formats on the fly. The recently issued *PIRUS Code of Practice* for recording and reporting usage at the individual article level will be adopted soon. Currently some repositories are reluctant to share deposits with other repositories because it takes away from their usage data. With PIRUS, they will be able to collect usage from wherever the article is accessed, which should help to encourage sharing.

There is much that publishers could do to aid institutions in managing their repositories. Adoption of community- or discipline-specific metadata vocabularies that are more robust than Dublin Core would eliminate or reduce the manual classification of article deposits. Using and sharing standardized article metadata through accessible APIs would serve numerous purposes and be useful beyond just repositories. Publishers could require or incentivize researchers to get an ORCID and provide it with all submitted manuscripts. They could also do direct deposits themselves of the final accepted manuscript into institutional repositories, like many publishers currently do with PubMed Central. Using formats other than PDF or providing multiple formats, e.g. PDF + HTML or + XML would aid in machine-readability and reusability of content.

The Harvard Open Access Project (HOAP) is distinct from the Office for Scholarly Communication (OSC) and looks beyond OA at Harvard to OA everywhere. It provides a current awareness service called the Open Access Tracking Project, creates an ontology for classifying OA developments, and catalogs OA journals published by scholarly societies. It particularly tries to spread awareness of *Good Practices for University Open-Access Policies*, consulting pro bono with other universities to assist them in developing their own OA policies. →



---

## TRACKING AND REPORTING COMPLIANCE WITH OA POLICIES

---

**Robert Kiley** | *Head of Digital Services,  
Wellcome Trust Library*

The Wellcome Trust (WT) has been a vanguard of the open access (OA) movement over the last 10 years and expects recipients of its funding to provide free, online access to their published research results. Electronic copies of any research papers that have been accepted for publication in a journal have to be made available through PubMed Central (PMC) and Europe PubMed Central (Europe PMC) as soon as possible, but not later than six months of any publication. In April 2013, an additional requirement was introduced that if WT pays an article publication charge (APC), the article must be licensed using CC BY (Creative Commons Attribution).

Funding organizations, both governmental and private, as well as researchers' institutions need to be able to track and report compliance with OA policies, which can be difficult, time consuming to compile, and not 100% accurate. There is currently no standardized metadata that can be used consistently with search and discovery services for identifying that an article is published in some type of open access. The information about the funding agency and the grant number is often included in an acknowledgements section of the text and either not repeated in the metadata or not used with standard formats or syntax. WT encourages researchers to make full use of available identifiers and metrics, in particular ORCID (Open Researcher and Contributor ID) and persistent digital identifiers, such as DOI, for both articles and datasets. Standard identifiers for funding sources would also be helpful.

Since WT requires article deposits in PubMed Central, they can run an automated search every month to find the number of articles attributed to the Trust. These searches are showing about a 70% compliance level with WT's OA policies. The searches do occasionally pick up some false hits where Wellcome Trust is mentioned but is not a funder, and also miss some papers where WT funding is not properly attributed. They have been working with PubMed Central to more consistently index WT-funded



## IS IT FULL GOLD OA, OR HYBRID, WITH OR WITHOUT SUPPORT FOR GREEN ARCHIVING?

Deciphering this can be very difficult for researchers, especially where publishing is being done by one organization on behalf of another, such as a professional society.

While progress is being made, much more attention is needed to get infrastructure in place for compliance tracking and reporting.

research and put that information in the grant funding attribution field. WT manages the Europe PMC repository, on behalf of 26 other publishers, so they are able to ensure the infrastructure is in place there to search and report on the content. They are also an early adopter of the CrossRef service FundRef and would like to see more publishers use this system to report funding sources for published scholarly research.

Far more difficult to track is the compliance with the CC BY licensing requirement. License metadata isn't always included at the article level or done in a consistent way. One publisher, for example, included the license type as a footnote to the article. Other publishers are only identifying licenses at a journal level or the license information is only available within the publisher's internal system. A standardized method and taxonomy is needed to express licensing at the article level in a machine-readable way.

It's often not clear what the publisher's policy for open access is, even at the journal level. Is it full gold OA, or hybrid, with or without support for green archiving? Deciphering this can be very difficult for researchers, especially where publishing is being done by one organization on behalf of another, such as a professional society. Thus researchers are uncertain if they will be complying with WT policy if they choose a particular journal. WT has been providing some funding support for SHERPA-FACT to help get this information better collected and searchable in the SHERPA system. Much of this information still has to manually interpreted by SHERPA.

Machine-readable licensing terms and/or an API to this information in the publishers' systems could go a long way in enabling the collection and maintenance of policy and licensing information.

Not captured at all yet, outside of the publishers' systems, are the fees to publish an article. Wellcome Trust currently has to go back to each institution to see what was paid per article, per publication, and per publisher. WT will give institutions a block of money to use for APCs and the institutions have to send a yearly spreadsheet showing how they spent the monies. The data that is returned can be variable in content and format. Last year, WT put this data online (see: <https://docs.google.com/spreadsheets/d/1RXMhqzOZDqygWzyE4HXi9DnJnxjdp0NOhIHcB5SrSZo/edit#gid=0>) and used community crowdsourcing to enhance it with DOIs, OA status, and licensing. While a fairly successful effort, widely implemented standards for reporting could eliminate the need for such enhancement work.

Progress is being made, but much more attention is needed to get the needed infrastructure in place for compliance tracking and reporting. There is still inconsistent use of metadata and too much manual communications with spreadsheets being done. A great deal of the data needed is held by individual publishers and better tools and mechanisms are required to enable publishers to share the data they hold with funders and researcher institutions. →

CONTINUED »




---

## INTEGRATING NEW ECONOMIC MODELS FOR OA PUBLISHING

---

**Roy Kaufman** | *Managing Director for New Ventures and Executive-level lead on Open Access*

**Jennifer Goodrich** | *Director of Product Management Both with Copyright Clearance Center*

In the subscription model of STM journal publishing, the number of relationships between the publisher and the paying customers is fairly concentrated. Libraries are the majority of the subscribers and most libraries work through subscription agents like Swets or EBSCO. Individuals may also subscribe as members of a learned society, with the payments aggregated by the societies. So publishers have a small number of payers to deal with, many payments are made once a year, and the payments, which are often aggregated, are on a larger scale.

Gold open access (OA) publishing, where the economic model switches to the author (or author's institution) paying article publication charges (APCs) changes things considerably. The number of payments and paying individuals or organizations has increased exponentially, the payments are made throughout the year, and many of the individual payments are small in comparison to subscriptions. Additionally, the APC fees can vary depending not just on such technical issues such as page count and number of color illustrations, but also based on the location/currency of the author and by whether different types of discounts might apply, such as society memberships, institutional volume discounts, pre-paid deposit account discounts, or whether the institution subscribes to the journal.

Publishers' systems are often not set up to handle the volume and variations of these new OA payments and the workflows are not always established to tie payments to specific articles and track that payments have been received prior to publication.

Previously, many institutions had few or no systems and processes in place to track their researchers' publication activities. When any tracking was done, it was usually post-publication and often at the departmental level. Now institutions have to develop new policies regarding APC payments, as well as set up systems and processes to budget, fund, and manage such payments. The processes have to be initiated very early in the publication cycle, often

prior to article acceptance, rather than post-publication. Researchers, who were often used to dealing directly and alone with publishers about their articles, have many more institutional hoops to jump through before they can get published. And institutions have to set up reporting mechanisms to funding agencies to prove compliance with OA policies. If the funder's monies are used for APCs, these also have to be tracked and reported at the grant level or even by the specific article. For those institutions that are also doing Green OA repository publishing, even more processes and systems have to be established.

Article publication workflows are further complicated by, and increasingly tied to, licensing issues. Licenses used to be imposed by the publisher with little or no negotiation room. Often the licenses were standardized across a publisher's entire portfolio; if more granular, they may have been at a discipline level or at most a journal title level. With OA publishing, some funding agencies and author institutions are dictating the type of license that is required, often CC BY (Creative Commons Attribution), but other license variations may apply. (On the other hand, a March 2014 survey conducted for Taylor and Francis showed that the majority of authors preferred more restrictions on the reuse of their published research.) Thus licenses can vary at the article level, especially in hybrid publications. An article could also have more than one grant and funder associated with it, each with different or possibly conflicting publishing and licensing requirements. These license nuances have to be identified from the time of article acceptance through to publication and distribution to the end users. Licenses can also affect the APC rates, since publishers may lose their rights to sell reprints with certain licenses—a major revenue stream for some—and may increase the APCs in those cases to make up the difference.



To date, many publishers and institutions are still struggling to set up working systems and processes to support OA workflow for APCs and licensing. New software and services are being introduced, both commercial and open source, but are not yet widely used or well integrated.

To date, many publishers and institutions are still struggling to set up working systems and processes to support OA workflow for APCs and licensing. New software and services are being introduced, both commercial and open source, but are not yet widely used or well integrated. Standards will be critical to making these new services integrate with each other and with existing systems, both within and between organizations. Metadata attached to the individual article that travels with it throughout the workflow is especially important. Among the standards that need to be utilized in this metadata are researcher identifiers such as ORCID, author and institution identifiers such as ISNI, and article identifiers such as DOI. The use of the DOI is furthest along, but even after close to 15 years of standardization, it is still not universally used by all publishers. And DOIs are usually not assigned until the time of publication (or even afterwards). ORCID and ISNI are more recent standards and are in the early adoption stages. Missing standards are those addressing funding information, such as funder and grant identifiers, licensing terms that are machine-readable, identification of the type of open access article and ties to embargo periods that may apply, article versioning (especially where green and gold versions both exist), and APIs or protocols for moving commonly used data between disparate systems.

Two services that are gaining some traction in aiding publishers and institutions in implementing the new economic models are FundRef from CrossRef and RightsLink® for Open Access from the Copyright Clearance Center. With FundRef, publishers deposit funding information from articles using a standard taxonomy of funder names. This funding data is then made publicly available through CrossRef's search interfaces and APIs for funders and other interested parties to use and analyze.

RightsLink provides a service that integrates directly with a publisher's workflow to allow authors and institutions to pay, track, and manage APCs. Users can view estimated mandatory and optional charges before acceptance, as well as the final charges at time of acceptance. Payments can be made by credit card directly through the system, by crediting to a deposit account, or an invoice can be requested for one of seven different currencies. Monies are collected for and remitted to the publishers, eliminating their burden of handling these numerous payments. Various publisher reports are available at any time, including order history, manuscript status, and payment status. Forthcoming reports will show aggregated information by publication, institution, or funder. The service makes heavy use of metadata supplied by the publishers, utilizing APIs with their systems that allow the metadata in RightsLink to get updated as a manuscript moves through the publisher's workflow from submission through publication. Thus it is a perfect example of how widespread use of standardized metadata by the publishers can improve the information and services available to everyone who uses this service. →

Widespread use of standardized metadata by publishers can improve the information and services available to everyone.

CONTINUED »




---

## OPEN ACCESS PUBLISHING TOOLS

---

**Martin Eve** | *Lecturer in Literature at the University of Lincoln, UK, Academic Project Director of the Open Library of Humanities, and founding member of the Open Access Toolset Alliance*

Tools for open access publishing of scholarly journals run the gamut from proprietary systems and large software packages that cover the whole workflow to niche open source tools for a single function. Interoperability between different systems is nearly nonexistent as are standard APIs and protocols to move data between them. Systems available outside of the commercial arena are still developing and the learning curve for using them can be quite steep.

Open Journal Systems from the Public Knowledge project is one of the more widely deployed open source journal management and publishing systems, but is still missing some needed functionality on the production end, such as content editing and XML generation. PLOS uses the Ambra platform, but it has not been adopted by many others, even though it is open source, possibly due to lack of modularity in its design. Wordpress plug-in solutions, such as Annotum can take a blog and turn it into an OA journal, but do not address other needed parts of the workflow. Still missing is a single, modular system that would allow a journal to be designed with drag and drop functionality, have plug-ins for all the different modules of the workflow, and support standards for creation, discovery, and preservation. Even more problematic is the inability to migrate content from one platform solution to another, as export formats and protocols do not currently exist.

A key standard for making scholarly information more reusable and accessible is JATS (*Journal Article Tag Suite*, ANSI/NISO Z39.96) for XML markup. Most researchers, however, are still writing and submitting their manuscripts in word processing software and there are few tools to easily convert such text into the JATS XML language. Those that exist are proprietary, rather than open source, and can be expensive. Even where JATS is used, different viewers can produce different results for the end user. The JATS for Reuse (JATS4R) project is working to define best practice tagging guidelines, along with tools that can help publishers identify whether their content is compliant with those best practices.

Also needed are standardized preservation solutions. Some libraries and repositories are participating in semi-private networks like LOCKSS or CLOCKSS. Some commercial publishers are using services such as Portico. But many journal publishers, both open access and commercial, are not using such preservation solutions for their e-journals. With libraries no longer owning their e-journals, this dependency on the content creator for long-term preservation is a serious concern.

Currently, a high degree of expertise is needed to use the existing tools for open access publishing. New tools, both commercial and open source, are in development, but a substantial lowering of the barrier to entry for using these toolsets is needed. More awareness and education about all the elements that must work together and where standards like JATS fit into the workflow are also critical to expanding open access publishing.

The Open Access Toolset Alliance was formed in August 2013 to create open source tools for open access scholarly publishing, facilitate discussion and collaboration, and showcase relevant projects. Individuals or institutions who are engaged in open source initiatives related to open access publishing are welcome to join. →

A key standard for making scholarly information more reusable and accessible is JATS (*Journal Article Tag Suite*, ANSI/NISO Z39.96) for XML markup.

Creating and managing a sustainable OA infrastructure is a challenging task and much more joint, collaborative effort is needed to move successful projects and experiments into the mainstream.



## SUSTAINABILITY OF AN OA INFRASTRUCTURE

**Dr. Alma Swan** | *Director of European Advocacy Programmes for SPARC Europe, and Director, Key Perspectives Ltd.*

**Dr. Caroline Sutton** | *Publisher and Co-Founder, Co-Action Publishing*

Almost all of the infrastructure services for open access were created on project money and many significant services still depend on such "soft" funding sources. This is a major concern for the future sustainability of these systems and services. In an effort to secure their long term future, some of these services have developed business models that involve individually approaching libraries and institutional repositories every year to obtain ongoing funds. While this may suit these individual services and the libraries involved at the moment, it is clearly not a workable solution for the long term if every service adopts this model. So far, there have been few efforts made to group the services together or to approach library consortia or associations for a more sustainable funding method. Services are often ephemeral "proofs of concept" with no plan or intent for ongoing management.

The Knowledge Exchange—a joint project of CSC-IT centre for Science in Finland, Denmark's Electronic Research Library (DEFF), the German Research Foundation (DFG), Jisc in the United Kingdom, and SURF in the Netherlands—has undertaken work to look at the sustainability of the OA infrastructure. Their *Sustainability of Open Access Services* Phase 1 and 2 report identifies three strategic areas that are needed: "embedding business development expertise into service development; consideration of how to move money around the system to enable Open Access to be achieved optimally; and governance and coordination of the infrastructural foundation of Open Access." The Phase 3 report discusses "two critical elements to designing an effective sustainability model for a free-to-the user infrastructure service: 1) inducing potential participants to reveal their demand for the service, and 2) getting organizations to contribute voluntarily to its provision." It also states that "in some cases, a sustainable fee-based model—that enables an initiative to deliver key infrastructure services to those organizations in the value chain that most require them—may be preferable to the free dissemination of a less-robust service to a broader audience."

Infrastructure Services for Open Access (IS4OA) was formed as an umbrella entity that aims to shelter a set of complementary OA services and to obtain ongoing funding for them from the research community using a few-to-few approach, rather than the many-to-many methods currently done for each individual project. In support of their mission to facilitate easy access to Open Access resources, IS4OA assumed responsibility in December 2012 for the ongoing support and maintenance of the Directory of Open Access Journals (DOAJ). Since then, they have implemented new governance and workflow; created an Advisory Board consisting of publishers, institutions, and libraries; introduced a more extensive application form to describe each journal; and are piloting the use of associate editor positions to review the applications and validating the information before it is added to DOAJ. In May 2014, IS4OA added the Open Citations Corpus, an open access repository of scholarly citation data, as a supported service. As more services are sheltered under its umbrella, IS4OA anticipates being able to further reduce administrative overhead for duplicated activities. It also foresees being able to implement data feeds between the services, thus improving individual services and exploiting potential mutual benefits to the full.

Creating and managing a sustainable OA infrastructure is a challenging task and much more joint, collaborative effort is needed to move successful projects and experiments into the mainstream. Publishers, in particular, are needed to join and support such efforts and bring their comprehensive knowledge and expertise to the table. One opportunity for such collaboration is the Jisc Open Access Good Practice project, which is planning a series of workshops in 2014-2015 to explore various open access issues and solutions. ☉

| FE | doi: 10.3789/isq26no2.2014.02

**CYNTHIA HODGSON** (chodgson@niso.org) is the Managing Editor of NISO's *Information Standards Quarterly* magazine.

### Acknowledgments

The author wishes to express sincere thanks to Kristen Ratan, Product Director, Public Library of Science (PLOS) for her assistance in scoping this article, identifying issues, and providing contacts for interviews.



**Annotum**

<http://annotum.org/>

**Berlin Declaration on Open Access**

<http://www.berlin9.org/about/declaration/>

**Bethesda Statement on Open Access Publishing**

<http://dash.harvard.edu/handle/1/4725199>

**Budapest Open Access Initiative**

<http://www.budapestopenaccessinitiative.org/read>

**Budapest Open Access Initiative, Ten years on**

<http://www.budapestopenaccessinitiative.org/boai-10-recommendations>

**CLOCKSS**

<http://www.clockss.org/>

**Copyright Clearance Center white papers on open access**

[http://www.copyright.com/content/rightscentral/en/toolbar\\_main\\_content/resources/white-papers.html](http://www.copyright.com/content/rightscentral/en/toolbar_main_content/resources/white-papers.html)

**Creative Commons Licenses**

<http://creativecommons.org/licenses/>

**Directory of Open Access Journals (DOAJ)**

<http://doaj.org/>

**DOI (Digital Object Identifier)**

<http://www.doi.org/>

**Europe PubMed Central (Europe PMC)**

<http://europepmc.org/>

**FundRef (CrossRef)**

<http://www.crossref.org/fundref/>

**Good practices for university open-access policies**

<http://bit.ly/goodoa>

**Harvard Open Access Project**

<http://cyber.law.harvard.edu/research/hoap>

**Harvard University Library Office for Scholarly Communication**

<https://osc.hul.harvard.edu/>

**Infrastructure Services for Open Access (IS4OA)**

<http://is4oa.org/>

**ISNI (International Standard Name Identifier)**

<http://www.isni.org/>

**JATS (Journal Article Tag Suite)**

<http://www.niso.org/workrooms/journalmarkup>

**JATS for Reuse (JATS4R)**

<http://jats4r.github.io/>

**Jisc Open Access**

<http://www.jisc.ac.uk/open-access>

**Knowledge Exchange**

<http://www.knowledge-exchange.info>

**LOCKSS**

<http://www.lockss.org/>

**Open Access Good Practice project (Jisc)**

<http://openaccess.jiscinvolve.org/wp/>

**Open Access Toolset Alliance**

<http://www.oatools.org/>

**The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)**

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

**Open Citations Corpus**

<http://opencitations.net/>

**Open Journal Systems**

<https://pkp.sfu.ca/ojs/>

**ORCID**

<http://orcid.org/>

**PIRUS Code of Practice**

<http://www.projectcounter.org/pirus.html>

**Portico**

<http://www.portico.org/>

**Public Knowledge project**

<https://pkp.sfu.ca/>

**PubMed Central (PMC)**

<http://www.ncbi.nlm.nih.gov/pmc/>

**RightsLink (Copyright Clearance Center)**

<http://www.copyright.com/openaccess>

**ROARMAP (Registry of Open Access Repositories Mandatory Archiving Policies)**

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

**SHERPA/FACT**

<http://www.sherpa.ac.uk/fact/>

***Sustainability of Open Access Services: Report on Phase 1: Scoping the Challenge and Phase 2: Consulting the Stakeholders. Knowledge Exchange, September 2012.***

[http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOpen+Access%2fSustainabilityServices%2fSustainability\\_OA\\_services\\_12.pdf](http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOpen+Access%2fSustainabilityServices%2fSustainability_OA_services_12.pdf)

***Sustainability of Open Access Services: Phase 3: The Collective Provision of Open Access Resources. SPARC and Knowledge Exchange, March 2013.***

<http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOpen+Access%2fSustainabilityServices%2fSustainability+of+OA+Services+phase+3.pdf>

**SWORD**

[http://www.ukoln.ac.uk/repositories/digirep/index/SWORD\\_Project](http://www.ukoln.ac.uk/repositories/digirep/index/SWORD_Project)

**Taylor & Francis Open Access Survey, June 2014**

<http://www.tandf.co.uk/journals/explore/open-access-survey-june2014.pdf>

**Wellcome Trust**

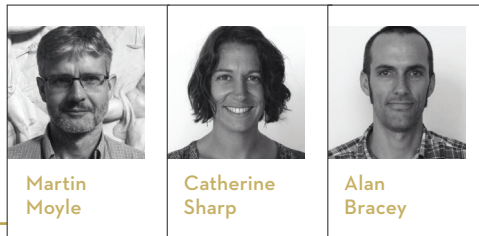
<http://www.wellcome.ac.uk/>

**Wellcome Trust Open Access Policy**

<http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm>



RELEVANT  
LINKS



Martin Moyle

Catherine Sharp

Alan Bracey

# The Role of Standards in the Management of Open Access Research Publications: A Research Library Perspective

MARTIN MOYLE, CATHERINE SHARP, AND ALAN BRACEY

UCL (University College London) is a multidisciplinary, research-led institution with approximately 5,000 staff and postdoctoral researchers and 4,500 research students. As London's Global University, UCL engages with the spectrum of research subjects, from arts and humanities to basic and applied sciences and medicine. UCL embraces open access (OA), supporting both the Gold and Green routes to OA. Academic freedom is a cornerstone of UCL's OA Publications Policy: UCL researchers are free to determine where to publish, how much to publish, and how often to publish.

## OA support at UCL

Open access at UCL is championed by the Vice-Provost (Research) and strategy is led by a Publications Board of senior academics and administrators. Operational responsibility for OA resides with UCL Library Services. The Library OA Team is responsible for UCL Discovery, UCL's Institutional Repository, ensuring that Green OA is supported efficiently and legally. The Team also manages Gold OA, with responsibility for overseeing UCL's Gold OA publications budget and for helping researchers to align their publication practices with the OA policies of funders, in particular those of the Wellcome Trust, Research Councils UK (RCUK), and, recently, the UK Higher Education Funding Council for England (HEFCE).

CONTINUED »

## Key OA activities

UCL Library Services supports OA at UCL through three key activities:

- » Arranging the payment of Article Processing Charges (APCs), including the negotiation and maintenance of agreements with publishers that secure value for money and/or administrative efficiencies for UCL.
- » Storing and providing access to Green OA. UCL encourages self-upload of research by authors; on receipt, permissions are checked and embargoes applied before access is enabled.
- » Collecting, analyzing, recording, and reporting management data.

## OA research publication: stakeholders and standards

Standards are particularly important in multi-stakeholder environments in which different parties need to share information with consistency. Open Access research publication is one such environment. The primary stakeholders in this context include authors, publishers, funders, and a range of administrative units within higher education institutions (HEIs) supporting OA management and research administration. The main stakeholders are supported by several service providers and agencies: for instance, vendors of research information systems and manuscript submission systems, registration agencies, open source developers and relevant organizations such as SHERPA and Creative Commons.

To illustrate some of the opportunities for interplay between stakeholders and standards in the area of OA research publication management, four brief case studies are outlined. The first considers the position of the author at publication; the following studies look more closely at the three “key OA activities” identified above.

### 1 Author obligations

At publication, authors face the task of identifying and complying with a range of requirements imposed by their publishers, funders, and institutions.

At the earliest stages of publication, authors need to work with their publisher’s submission system, often proprietary and custom-made. They must provide required author identifiers, which may include ORCIDiDs or proprietary identifiers, if known. They must give their affiliation—not without difficulty at UCL, whose many postgraduate institutes and interdisciplinary research centers offer a variety of possibilities for diversity of attribution. Publishers may ask authors for a code or

institutional affiliation to establish that an institutional prepayment scheme for Gold OA applies; authors need to be aware that these publisher-institution relationships exist, and to know how to find out whether their institution has an agreement with their publisher. A title-specific Copyright Transfer Agreement (CTA) must be endorsed or amended and, ideally, its implications fully understood by the signatory.

To ensure that funder requirements are met, authors need to be aware of all the different funding sources that contributed to the paper, and to understand where to find guidance on those funders’ open access policies. This may direct or limit their publication choices: the Wellcome Trust, for example, strongly recommends Gold open access; and many UK institutions have block grants to cover APCs for RCUK-funded papers. Funder requirements also have a bearing on licensing, obliging the authors to assign a particular license (usually CC BY in such cases), sometimes necessitating further negotiation with the publisher. Authors need to know about and act on any stipulations concerning the acknowledgement of funders and grant numbers, either of which may be elicited as free text contributions rather than selected from a controlled list. (Authors can be asked to provide grant details several times over, in fact—in the manuscript, as metadata at final submission, in correspondence with administrators about APC payment, and, later, as part of institutional processes around publication recording and reporting.) Other funder idiosyncrasies, such as RCUK’s mandate for a statement about the availability of underlying data, must also be observed at submission.

Researchers must understand and comply with any requirements placed upon them by their institution. Local open access policies may direct their final publication choices and affect their engagement with CTAs and licenses. Typically, the institution requires a record of every publication produced by its researchers; the author must understand how such data is compiled and contribute to its upkeep. Publication harvesting systems are becoming more commonplace, but can be rather “hit and miss” affairs, not least because of author and institutional identification issues mentioned above, as well as disciplinary differences in the coverage of available harvesting sources. An author may have to make a manual intervention to ensure that the local publication record is up to date. Finally, and to close the loop within institutional systems, the correct associations between individual research outputs and local records of the funders and grants under which those outputs were incubated must be recognized and recorded.



## 2 Arranging Gold open access

For HEIs, arranging Gold payment can be a complex process. Funder requirements for Gold OA must be matched to publishers' Gold offerings; authors often need advising on HEI, publisher, and funder policies; and payment arrangements developed for the subscriptions environment do not translate to the payment of article processing charges, necessitating new financial processes.

UCL's Open Access Team arranges APC payments on behalf of authors, who may be addressing the requirements of a funder that strongly encourages Gold open access (the Wellcome Trust, for example). Authors may choose Gold if immediate open dissemination is required. As the rapid growth of Gold offerings from publishers has been an organic process, the level of "openness" offered varies widely—as a result, funder requirements for Gold are not always met by publishers.

When the Team receives a request, they verify an author's affiliation and eligibility to use UCL funding. The verification process is obstructed when up-to-date data is unavailable in local HR or research systems, when an author is affiliated with a number of different institutions, or when a non-UCL author requests Gold open access on behalf of a UCL co-author. Widespread adoption of ORCID identifiers would undoubtedly contribute to a more efficient process for confirming eligibility. Bibliographic and funding details are obtained and recorded in the OA Funding Team's database, and the Team explicates funder policies to authors and clarifies if, and how, a journal's Gold option will comply. Many publishers offer Wellcome Trust- and RCUK-funded authors a choice of license, even though only CC BY is acceptable to the funder. The OA Funding Team directs authors to the type of license required by their funder, and troubleshoots cases where a non-compliant license has been chosen. Differing funder and publisher policies cause considerable confusion to authors and institutions. Funding details are an essential prerequisite for accurate assistance, but details provided to the Team by authors do not always match those given to the publisher, which may differ again from those acknowledged on the paper. The widespread adoption by publishers of a system for collecting and standardizing author's funding data (FundRef, for example) would greatly improve Gold open access workflows for all stakeholders.



Despite the existence of some relevant standards, supporting OA—certainly at UCL—depends uncomfortably on manual intervention and pragmatics.

At UCL, Gold payments are made either through publisher prepayment schemes or by invoice. Publisher schemes are preferred for the efficiency savings gained from not paying hundreds of individual invoices. The OA Funding Team records all payment, bibliographic, compliance, and deposit data centrally, updating records as transactions progress. Manual checks are performed to confirm whether funds have been released by University Finance, whether the publication has been made open access, whether the correct license type has been applied, whether funding is acknowledged (RCUK), and if the paper includes a statement on access to underlying research materials (RCUK). The Team also deposits the final PDF in UCL's institutional repository.

The potential for off-the-shelf APC management systems to deliver efficiencies is noted; however, the emerging systems have not yet transcended the complexities of servicing Gold OA on the scale at which the UCL Team operates.

## 3 Supporting Green open access

As with Gold open access, institutions need to be able to give authors accurate advice on how to engage with Green OA. Publisher rights cannot legally be disregarded, and institutions risk reputational damage (and perhaps financial penalties) if such copyrights are consistently breached. Accurate information about applicable Green rights is required at title level at the point of repository deposit. In the absence of standard input from publishers, it is difficult for an aggregator, like SHERPA (whether as SHERPA/RoMEO or SHERPA/FACT), to render accurately all permutations of a publisher's open access policies. The SHERPA APIs are potentially a great adjunct to repository and publication management systems, but, for full confidence in the legality of Green collections, guidance from the SHERPA suite can currently only be regarded as indicative.

CONTINUED »

The UCL Team frequently liaises directly with publishers and authors over the small print of CTAs. Every deposit is checked for legality and embargo periods identified and applied, before it is made openly accessible. Deposits to institutional repositories in the UK will soon increase dramatically in response to HEFCE's Open Access Policy, which requires the Green deposit, at acceptance, of a final manuscript of every article and proceedings paper that is to be eligible for the next Research Evaluation Framework (REF) assessment. To sustain support for Green OA at this level, institutions need reliable systems for the exchange of rights information, and to free them from the need to check every copyright agreement manually, while remaining fully confident that they are correctly observing publishers' rights and embargo rules.

The emphasis of the HEFCE OA Policy on deposit at acceptance also brings some new challenges in metadata management. Records will increasingly enter institutional systems at acceptance, when metadata is skeletal, and will need to be improved after publication so that they are fit to expose later to HEFCE for REF purposes. The publication harvesting systems already in use could help to automate local record enhancement, but their reliability is compromised by the fact that the unambiguous systematic identification of outputs only becomes possible at publication, when DOIs are disclosed (assuming they have been assigned at all). While HEFCE's support for Green OA is welcome, the risks of the duplication of records and an overall degradation of metadata quality in university systems are naturally increased by the terms of its mandate. There is a need to improve the ability of business systems accurately to disambiguate and merge metadata records to avoid such maintenance becoming an extremely time-consuming manual process for authors and/or administrators.

Institutions naturally wish to capitalize on their investment in repositories and one way of doing so is to celebrate the impact of open research. Monitoring the quantity and provenance of full text downloads is an obvious way of tracking impact. Benchmarking such impact across institutions, however, is made more difficult by non-standardized data collection. How much time should elapse between repeat downloads for each to be counted afresh? Are we sure that all crawlers and harvesters are being excluded from the figures? The IRUS-UK service, which aggregates COUNTER-compliant article-level statistics for cross-repository comparison, is a welcome development, and adoption of the same protocols in local repositories would help to boost the accuracy and credibility of repository impact assessment by institutions.

#### 4 *Compliance monitoring and reporting*

Accurate compliance monitoring is a challenge. To begin with, a means of identifying all the outputs associated with the institution and funded by each relevant research funder must be in place. Such a conspectus is difficult to achieve because of the vagaries of author and funder identification in the workflows outlined above, and so efforts to determine compliance are compromised at the outset.

In preparation for compliance reporting, institutions need to assemble bibliographic data about Gold and Green articles, including DOIs and institutional or subject repository identifiers; funder and grant information; information about licenses, acknowledgements, and other required statements; details of deposit in any specified subject repositories; financial information about APCs and their breakdown between funders; and details of acceptance dates, publication dates, and embargo periods. These administrative and bibliographic metadata need to be stored in readiness for report to different funders—who, of course, have different reporting requirements and emphases, not least because their mandates specify different criteria for compliance. Often, the data elements required for reporting are stored across several local systems, particularly where the institution does not have a current research information system (CRIS). Reports for consumption within the institution, meanwhile, also need attention; monitoring the level of researcher engagement with the HEFCE mandate, for example, is set to be of intense importance to heads of research in UK universities in the near future.

Publisher pre-payment systems have some merit in easing the burden of reporting, in that they typically require the publisher to make a periodic disclosure of how the money deposited by the institution has been used. Such reports, however, do not yet conform to any standardized format or content, and the number of schemes on the market is in any case few, so they are only a small, if helpful, part of the overall picture. There is very little escape from manual data collection and assessment in the sphere of compliance. The UCL OA Team spends much time interacting with authors and publishers to collect, verify, and store a wide range of information in support of future reporting needs.

### Current standards and services to facilitate OA research publication

The foregoing illustrations highlight the fact that, despite the existence of some relevant standards, supporting OA—certainly at UCL—depends uncomfortably on manual intervention and pragmatics. The illustrations also show that authors shoulder a great deal of responsibility for policy awareness and administrative information provision, areas into which ideally they would not have to be so deeply drawn. It is noteworthy that many of the concerns of

the stakeholders in OA research publication, such as the accurate identification of authors, funders, grants, publisher rights, and individual published outputs, are shared. Such repetition of need strengthens the argument for the development and implementation of standards-based interoperability.

Table 1 summarizes the authors’ perception of the current availability and utility of standards, and services with the potential for standards-based development and integration, that would facilitate the OA support work outlined above.

CONTINUED »

TABLE 1: RELEVANT OA STANDARDS AND PROCESSES AND CURRENT STATUS

AREA	REQUIREMENT	WHERE ARE WE NOW?
Researcher identification	Unambiguous identification of authors.	ORCID has much promise; it transcends proprietary researcher ID services and tracks individuals independently of institution. To be fully effective, requires adoption by all publishers and all researchers.
Funder identification	Unambiguous identification of the funder(s) of a research output—useful to funders, HEIs, and publishers.	FundRef offers publishers a normalized list for use in submission systems. Would require full take-up by publishers to be effective.
HEI identification	Correct association of authors and papers with parent HEIs, enabling accurate attribution of citations and supporting compliance monitoring.	ISNI standard (ISO 27729) has promise; and it is noted that ISNI and ORCID have recently announced joint agreement. However, ISNI is not currently widely adopted in submission systems, therefore variant nomenclature is in use.
Digital object identification	Unambiguous identification of a published output.	CrossRef is proven and has high levels of buy-in. DOIs only released at publication; would help HEIs immensely if DOIs could be assigned and routinely shared at acceptance.
Bibliographic metadata exchange	Seamless, automatic population of institutional publication systems.	Several metadata standards, not always employed. Each consumer system needs custom solutions per data source.
Administrative metadata exchange	Funder information, acknowledgements, license details, etc.	No standard. Metadata usually provided via ad hoc add-ons to bibliographic metadata supply, or derived manually by administrators.
Journal submission	Handover of final accepted manuscript to publisher.	Systems and workflows vary between publishers.

TABLE 1 CONTINUED »



## CONTINUED...

AREA	REQUIREMENT	WHERE ARE WE NOW?
APC processing	Regularized workflows to manage APC transactions, especially financial aspects.	Solutions being developed, but hitherto unsuitable for large HEIs. Current dependency on per-publisher arrangements.
Publisher rights	At-a-glance, ideally machine-readable, and accurate summary of self-archiving and embargo policies, per journal.	SHERPA-RoMEO is a useful dataset; has API; but currently cannot be relied upon to automate self-archiving without risk. Needs full publisher buy-in to a standard set of rights metadata.
Journal compliance with funder OA policies	Help authors and administrators to identify appropriate journal in which to publish.	SHERPA-FACT tool. RCUK and Wellcome Trust only. Not standards-based. Has potential if input/output can be standardized and coverage widened.
Licensing	Funders increasingly specify the assignment of particular licenses. Authors need clarity; publishers need to guide and support author choices.	Creative Commons. Mature concept, increasingly well understood, but not yet fully embedded in publisher workflows.
Repository downloads	HEIs would like to show impact of Green OA.	IRUS UK—aggregation based on COUNTER standard—good model. Similar standards not necessarily applied in local IRs.
Compliance monitoring	Ensure HEIs meet all obligations to research funders.	Need to combine and analyze HEI/author/funder/license data about publications. Some relevant standards in these areas (see above), but adoption too patchy to be dependable.
Compliance reporting	Accountability of HEIs to funders.	Requirements are funder-specific, although with much overlap between funders. Designation of a common framework for core elements of reporting would enable efficiency gains at report-making institutions.

## Conclusion

For standards to be useful to any given community, they must be adopted by all its members, and members' business systems must support the exchange of the information that has been standardized. The short case studies given above indicate that the stakeholders within the sphere of OA research publication do not benefit from the effective standardization of even the relatively few key pieces of information that support commonplace

OA interactions, let alone the exchange mechanisms that would enable such harmonized information to flow in a timely way between systems. Funders are channeling money into Open Access—for example, RCUK has committed £20 million to OA from the UK research budget in the current financial year—while HEIs are making increasing investment in repository services and OA administrators, who spend their days chasing down information from authors, publishers,

fundings, and finance systems, and maintaining and reconciling multiple spreadsheets for myriad different accounting and reporting purposes. Meanwhile, investment is clearly being made into relevant standards and services, but it is uncoordinated. Improvements are incremental and piecemeal, and meaningful, efficiency-delivering integration is frustratingly elusive.

All the stakeholders in OA research publication would benefit from further investment in the definition, refinement, promotion, and integration of relevant standards. There are common needs and shared problems, and there is new money in the ecosystem. It is interesting to reflect on the potential for publisher systems to act as a “hub” for the exchange of a significant quantity of information between the stakeholders in OA research publication. During the publication process, it is conceivable that author IDs, unambiguously-identified HEIs, DOIs, registry-sourced funder and grant details, standardized rights and license information, at-acceptance and at-publication metadata, financial transactional information, and, indeed,

the final accepted manuscript itself might seamlessly change hands between publishers, universities, and funders, drawing on registries and related services (and, additionally, leaving authors with more freedom to concentrate on writing). The traditional role of the publisher is challenged by OA, particularly as universities are taking more responsibility for the dissemination and curation of their published assets. New publisher-led initiatives, protocols, and systems to facilitate standards-based communication between the stakeholders in OA research publication might be warranted. Any such developments would certainly be welcomed by staff who are involved in the administration of open access at universities.

IIP | doi: 10.3789/isqv26no2.2014.03

---

**MARTIN MOYLE** (m.moyle@ucl.ac.uk; ORCID 0000-0002-2065-0508) is Group Manager, IT and Bibliographic Services, **CATHERINE SHARP** (catherine.sharp@ucl.ac.uk; ORCID 0000-0002-4472-943X) is Open Access Funding Manager, and **ALAN BRACEY** (a.bracey@ucl.ac.uk) is Open Access Compliance Officer—all with UCL Library Services, UCL (University College London), UK.

#### COUNTER Code of Practice for Articles

<http://www.projectcounter.org/counterarticles.html>

#### Creative Commons

<http://creativecommons.org/>

#### CrossRef

<http://www.crossref.org/>

#### FundRef

<http://www.crossref.org/fundref/>

#### HEFCE Policy for open access in the post-2014 Research Excellence Framework

<http://www.hefce.ac.uk/pubs/year/2014/201407/>

#### IRUS-UK

<http://www.irus.mimas.ac.uk/>

#### ISNI (International Standard Name Identifier)

<http://www.isni.org/>

#### Open Access at UCL

<http://www.ucl.ac.uk/library/open-access/>

#### ORCID

<http://orcid.org/>

#### RCUK (Research Council UK) Policy on Open Access

<http://www.rcuk.ac.uk/research/openaccess/policy/>

#### SHERPA/FACT (Funders & Authors Compliance Tool)

<http://www.sherpa.ac.uk/fact/>

#### SHERPA-RoMEO

<http://www.sherpa.ac.uk/romeo/>

#### UCL Discovery: Unlocking UCL research

<http://discovery.ucl.ac.uk/>

#### UCL Publications Policy 2012

<http://www.ucl.ac.uk/library/about/strategies-policies/publications-policy>

#### UK Higher Education Funding Council for England (HEFCE) Policy Guide Open Access Research

<http://www.hefce.ac.uk/whatwedo/rsrch/rinfrstruct/oa/>

#### Wellcome Trust Open Access Policy

<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/>



RELEVANT  
LINKS

David  
Ross

# A Publisher's Perspective on the Challenges of Open Access

DAVID ROSS

Open access (OA) publishing and archiving of academic research is becoming an important part of the scholarly communication process. This paper provides a publisher's perspective on the challenges faced developing effective infrastructure in response to this evolving, competitive landscape. More specifically, the paper offers the perspective of what has commonly become known as a "traditional subscription publisher." The views offered are based on first hand experiences of one of these publishers, SAGE, and while many of the issues identified will be common to other traditional subscription publishers it does not purport to be representative of the entire industry. It is an industry that has many players with very different levels of engagement with open access.

Publishers are not a homogenous mass. Scopus indexes over 26,000 academic, peer-reviewed journals from more than 5,000 international publishers. Although the market is dominated by large publishers with portfolios running into the hundreds, if not thousands, of titles, there is a very long tail of smaller operations. These publishers vary significantly in philosophy and corporate structure: from commercial to not-for-profit; university presses to multi-nationals; independents to corporate behemoths; august institutions with hundreds of years of history to relatively new entrants. For those opting to engage with open access, the configuration of their systems and their ability to manage open access publication will vary widely. This paper attempts to present an overview of some of the key challenges in developing open access infrastructures that are common to many but are certainly not universal.

## Uncertain Legislative Framework

Academic publishers work in a global environment and their author base is international. Whilst the well-established western markets of Europe and North America still dominate, the emerging economies—led by China and India—are contributing an ever increasing proportion of the research output. As a result, the plethora of national funder mandates provides a very challenging environment for publishers to work within. In addition to these government orders, the numerous private funding agencies have their own OA requirements.

As of June 2014, ROARMAP (Registry of Open Access Mandatory Archiving Policies), as shown in Table 1, lists 466 mandates with a further 27 proposed.



	EXISTING	PROPOSED
INSTITUTIONAL MANDATES	218	7
SUB-INSTITUTIONAL MANDATES	44	4
MULTI-INSTITUTIONAL MANDATES	9	5
FUNDER MANDATES	90	10
THESIS MANDATES	114	

Table 1: ROARMAP OA Mandates (Source: <http://roarmap.eprints.org/>)

Historically these have generally required green OA archiving of research but more recently mandates that make provision for gold OA solutions have also begun to be rolled out. The highest profile has been the Research Councils UK (RCUK) mandate that came into effect on April 1, 2013, but the German Research Foundation (DFG) policy provides funds for pure Gold OA publishing and allows direct grants to be used for Hybrid Gold, and the Dutch funding council has also suggested that they are considering some form of gold mandate.

In the US, there are no less than three initiatives on the table: the outcome of the Office of Science and Technology Policy (OSTP) directive, the Fair Access to Science and Technology Research (FASTR) Act, and the Frontiers, Innovation, Research Science, and Technology (FIRST) Act.

Just recently, on May 15, the National Natural Science Foundation of China, one of the country's major basic-science funding agencies, and the Chinese Academy of Sciences, which funds research at more than 100 institutions, entered the fray with the first major green deposit mandate for China.

Open access has shifted from being a bottom-up, scholar-led movement to top-down, funder-led. But it is not the intention here to examine the relative strengths and weakness of any approach. It is to make the point that the uncertain and ever-shifting global framework presents publishers with unique challenges with respect to long-term strategic planning, shorter-term policy decisions, and the development of infrastructure and workflow solutions to support these.

SAGE, like most publishers, strives to enable author compliance and welcomes well thought-out national mandates with reasonable embargo periods on the availability of the version of record. However, while there is a considerable amount of overlap between these mandates, there are also significant variations in conditions: most often their deposit criteria, embargo periods, and preferred license. Keeping abreast of the evolving framework poses its own challenge and publisher policy changes are often required to reflect these. As an example, in 2013 SAGE adopted one of the most liberal policies with regard to the authors accepted manuscript (AAM), allowing authors to post this in an institutional repository or their personal website immediately, with no embargo. This makes articles published by SAGE compliant with all mandates that have requirements for the AAM but to enable this we were required to consult with all our publishing partners before doing so. (SAGE Publishes on behalf of almost 300 learned societies, associations, and institutes.) It then necessitated alterations to author publishing agreements for all of our 700 plus journals.

Collectively, the industry has shown itself to be willing to engage and seek solutions to these challenges. It instigated the CHORUS project in the US as a possible solution to address the specific request by the OSTP for federal funding agencies to put forward open access solutions to make research derived from their funding public. The project was set up based on CrossRef's FundRef service and CrossRef itself was an industry-funded organization formed specifically to address the need to develop industry-wide standards and provide some infrastructure, originally in relation to DOI technology linking scholarly references. As the global mandate picture develops, it is likely that more initiatives such as this will be required.

The challenge is second guessing future requirements when developing systems and policies and helping authors navigate their way through what can be a complicated and confusing landscape.

### Education and Compliance

At a recent workshop, a librarian outlined the problem of an author who had to satisfy five different mandates: Higher Education Funding Council for England (HEFCE), Research Council UK (RCUK), a private foundation, the publisher, and his institution. For an author, interpreting all these is difficult enough; actually ensuring compliance is even more so as there is no silver bullet that will

CONTINUED »

satisfy all five mandates, even if the most liberal embargo period was in place. Authors need assistance and publishers, as well as librarians, have a role to play in both explaining the situation and enabling them to comply. It is likely that the authors or the institutions themselves (it is after all the institution that signs the grant agreement in the majority of cases) will ultimately be responsible for compliance and suffer any consequence of non-compliance, but publishers are being asked to intervene. Initiatives such as SHERPA/FACT, an author/funder compliance tool for RCUK and Wellcome Trust, which relies on publishers' data, may go part of the way in providing a solution at the national level—but there is no global system under development to mirror it worldwide. Calls have been made to make mandates machine readable, to enable automated compliance verification, and in particular to provide accurate solutions for multi-funder cases, but these have yet to be answered.

Another example is that, to ease compliance, the recent mandate announced by the four UK higher education funding bodies will quite possibly result in publishers having to develop entirely new workflows to enable the automatic deposit of AAMs, much in the same way they developed systems to automatically deposit National Institute of Health (NIH) funded papers in PubMed Central (PMC). The mandate requires that for outputs to be eligible for submission to the next UK Research Excellence Framework (REF), the AAMs be deposited in a university institutional repository (IR) at the point of acceptance. UK higher education institutions generally have no comprehensive tool to identify and collect all their published research and metadata, let alone the ability to flag articles at the point of acceptance. While it will be incumbent on authors to work directly with their institutions, it is likely publishers will have a role to play in assisting them to satisfy this REF OA requirement. At present, different stakeholders are developing individual solutions. Coordination and cooperation are required and standardized solutions need to be developed.

### Systems and Process

For hundreds of years publishers have operated journal-level workflows. Although the advent of online publication began a shift to article-level workflows, open access publication has accelerated this change. Truly continuous, open access publications operate more or less solely at an article level (with some exceptions in title-level indexing requirements).



**There is a need for publishers to interact with authors as paying customers in a way they have not done before.**

To facilitate this, publishers have developed new processes to accommodate the needs of open access publishing, often using manual or semi-automated work-arounds in the early stages. Enabling article deposit of NIH-funded papers in PMC, altering production and hosting processes, building systems for authors to pick their preferred Creative Commons license, and developing article processing charge (APC) collection interfaces are just a handful of examples.

One of the challenges for established publishers has been that, generally, they operate using legacy systems designed to service journal publications under the subscription model and these systems are generally not fit for OA purposes. New systems are required in addition to existing ones—the subscription business has not gone away and is not going to in the foreseeable future. Like all organizations, publishers have a multitude of strategic objectives and have to prioritize where they invest their resources. Couple that with the fact that, outside of biomedicine, revenues generated by open access are quite modest and you find the development of systems to facilitate a more streamlined approach to OA are often not deemed business critical.

### Article Processing Charge (APC) Collection

A key pain point has been the administration of APCs. Journal publishers are configured to transact large payments with libraries in annual cycles for subscriptions. While those that operate parallel book programs may have some direct interactions with individuals as customers, third-party booksellers handle the majority of financial transactions even in

that sector. There is now a need with OA to quickly process thousands of payments from individual authors and their institutions, and little or no experience by publishers in handling such transactions.

Automated systems need to be developed to take payment by credit card, issue invoices where needed, and apply VAT to European customers only (an issue exacerbated by frequently changing European Union VAT rules). Multiple prices, discounts, split payments, waivers, and currencies have to be handled; transactions and information logged; and reports produced. Institutions which have OA deals in place—some of which are prepaid, some of which are negotiated discount deals—operate under different rules. The institutional approval processes for the use of OA funds are often idiosyncratic but still need to be adhered to, and detailed receipts must be issued and regular reports provided. Internally, APCs must be allocated correctly and credit control rules adapted to reflect the much smaller invoice amounts being dealt with.

None of these issues are insurmountable and numerous third-party providers have stepped into the space—from new entrants such as Open Access Key (OAK) to existing intermediaries such as EBSCO, SWETS, and the Copyright Clearance Center (CCC)—but APC handling requirements have required publishers to invest considerable funds and time in developing modified processes, in addition to existing operations, even when partnering with a relevant intermediary.

Finally, there are often unforeseen effects. For example, with direct payment by authors come greater customer expectations. This direct B2C (business to customer) transaction is seen by some as one of the key drivers of a functional APC market, but it also has implications for customer service functions, expected speed of publication, and additional author services. There is a need for publishers to interact with authors as paying customers in a way they have not done before.

Once again, these new functions are additional to existing operations. Although our interactions with authors may be evolving, SAGE believes the library will continue to be the main transactional partner in the future. The way that libraries are taking on the administration of open access demonstrates this is likely to be the case even in an OA environment.

## Licensing

Under the open access publishing model authors are often allowed a choice of a Creative Commons (CC) license, or something broadly similar. Until now publishers have generally operated a single common license for all the research they publish, whether it be through an assignment of copyright or the granting of an exclusive license to publish. There have always been some occasional exceptions, such as those covering US government employees, but the open access publishing model, with a choice of license type, fundamentally changes the legal relationship between the publisher and author.

Again, the intention here is not to debate the merits of CC licenses but to point out that infrastructure implications are not insignificant. Until now, the publisher has been the administrator—and defender—of the copyright in academic research works. Under a CC BY license, that responsibility remains with the authors themselves whereas some derivatives, such as CC BY-NC, require the publishers to retain some limited capacity in that role. Systems have had to be developed to automatically recognize license types, add them to article metadata, display them correctly with the associated article, apply the correct permissions criteria, and record the terms in a contracts database for future reference.

Although this task has been made easier by the existence of standard contract templates as established by Creative Commons, the true long-term implications of a large-scale shift of copyright administration to the author has yet to be felt.

## Standards and Identifiers

Possibly the biggest hurdle to developing scalable and interoperable systems in any industry is the development and implementation of common standards. It is no different in open access publishing. A great deal of progress has been made in some quarters but little in others.

As an illustration, one of the key problems faced by many in the industry is the inability to automatically identify an author's institution. As mandates proliferate and more and more institutional OA funds are set up with different business rules, the requirement to identify an author's institution to enable automation is becoming key. Returning to the UK REF OA mandate as a case in point, this requires that the AAM of almost

FROM THIS...



TO THIS ?

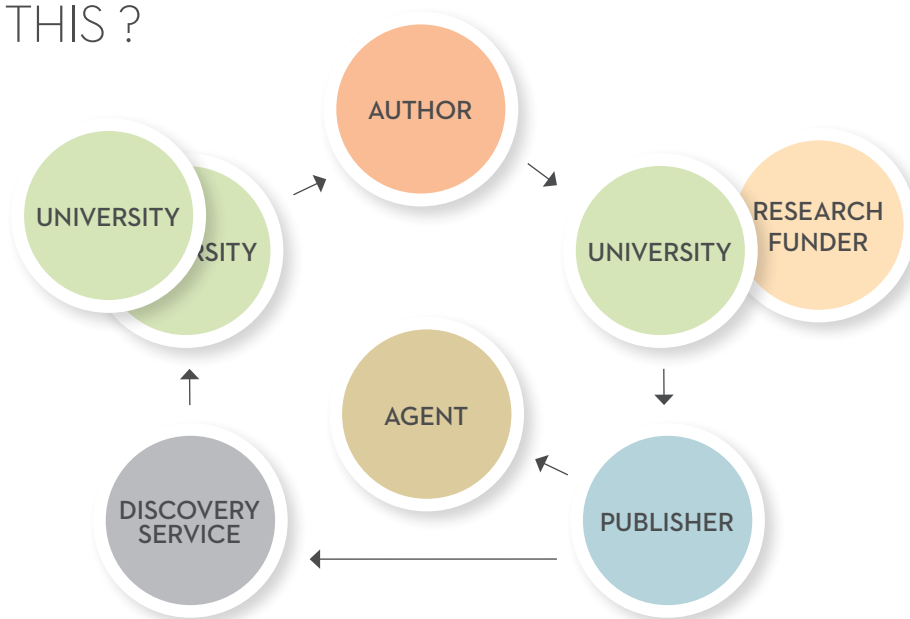


Figure 1. The Changing Roles of Stakeholders in the Information Chain  
 (Reproduced courtesy of Neil Jacobs, Head of Scholarly Communications Support, Jisc)



every paper written in a UK university is deposited in the relevant institutional repository. The first step toward that is obviously that an institution is alerted to when a paper has been accepted and this can only be done automatically if there are unique, globally recognized standards for identifying higher education institutions—perhaps even down to departmental level. While commercial initiatives such as Ringgold and nascent collaborations such as the Consortia Advancing Standards in Research Administration Information (CASRAI) may provide part of the solution, coordination is required by all stakeholders. The new ISO International Standard Name Identifier (ISNI) standard (ISO 27730) is promising, but is still in the early stages of recognition and adoption. Though author identifiers such as ORCID may one day be part of the solution, it is generally accepted it will be many years until a critical mass of the research community is indexed and registration becomes the norm.

This is only one example of the myriad of new interactions that will be required to develop global, industry-wide scalable and interoperable systems. There will be numerous new interfaces between publishers, authors, institutions, funders, intermediaries, and third-party vendors. Figure 1 illustrates how the relationships between the various stakeholders in the scholarly communication system are changing. It describes movement from a relatively stable environment with well-defined roles for all in the chain—publishers, institutions, authors, agents, discovery agents—to something that is much less clear. There is a new role for funders and several new roles for institutions, not just through paying for open access but also in monitoring and compliance. All these new interactions will require universal standards and identifiers in order for workable APIs to be developed.

There are many initiatives under way that form part of the picture. In addition to those mentioned above, NISO itself has created the Open Access Metadata and Indicators Working Group (renamed the Access and License Indicators Working Group), which is currently finalizing its recommendations following public consultations (see article on page 35). Jisc has started exploring the development of a managed shared service, Jisc Monitor, which might support UK institutions, central to which will be the adoption of standards to enable the interoperability required.

Legacy systems currently in use by all the stakeholders complicate matters further. Publishers' internal systems often have difficulty interfacing

with one another and the complexity of enabling these to interface with an entirely new set of external systems is not to be underestimated. Consider also that publishers generally depend on a variety of third-party vendors for key parts of their workflow. Manuscript processing systems and hosting platforms are generally contracted out and these have their own limitations including being generally built for journal-level workflows. These vendors have multiple customers and multiple, sometimes conflicting, demands, which means that they are not always the most flexible or swift at adapting.

### One Size Does Not Fit All

The open access publishing market as driven by APCs has predominately developed in the biomedical market, but it has always been accepted that humanities and social sciences (HSS) would present their own challenges. A recent study commissioned by the British Academy, *Open Access Journals in the Humanities and Social Science*, went further and suggested that the market should not so much be viewed as STM vs. HSS, but rather biomedicine vs. the rest. Certainly OA in the humanities, where the monograph is the main conveyer of information, faces its own challenges, but many other social science disciplines, where the research article is still the main vehicle, face their own particular problems, mainly due to the relative paucity of funding. In many HSS disciplines, the journal itself serves a very different purpose than in STM. Selection mechanisms are different and the necessity for expert opinion to confer authority on scholarly work that deals with concepts and ideas rather than empirical data requires different approaches.

This last point is made to illustrate that the development of the open access market is not evenly distributed, by geography or subject. For large global publishers that cover a range of disciplines this presents yet another level of complexity as they attempt to find viable long-term solutions that satisfy all the requirements of all major stakeholders: authors, editors, societies, universities, and funders.

IP | doi:10.3789/isqv26no2.2014.04

---

**DAVID ROSS** ([david.ross@sagepub.co.uk](mailto:david.ross@sagepub.co.uk)) is Executive Publisher of Open Access for SAGE Publishing ([www.sagepub.com](http://www.sagepub.com)) where he oversees a team that manages SAGE's flagship Open Access journals and develops systems, process, and policies to support their OA publishing.

CONTINUED »

**CAS [Chinese Academy of Science] Issues Open Access Policy**  
[http://english.cas.cn/Ne/CASE/201405/t20140516\\_121037.shtml](http://english.cas.cn/Ne/CASE/201405/t20140516_121037.shtml)

**CHORUS (Clearinghouse for the Open Research of the United States)**  
<http://chorusaccess.org/>

**Consortia Advancing Standards in Research Administration Information (CASRAI)**  
<http://casrai.org/>

**Copyright Clearance Center Launches RightsLink for Open Access**  
<http://goodereader.com/blog/e-book-news/copyright-clearance-center-launches-rightslink-for-open-access>

**Creative Commons**  
<http://creativecommons.org/>

**CrossRef's FundRef**  
<http://www.crossref.org/fundref/>

**Darley, Rebecca, Daniel Reynolds, and Chris Wickham. Open Access Journals in the Humanities and Social Science. The British Academy, 2014.**  
<http://www.britac.ac.uk/templates/asset-relay.cfm?frmAssetFileID=13584>

**Fair Access to Science and Technology Research (FASTR) Act**  
<https://beta.congress.gov/bill/113th-congress/house-bill/708>

**Frontiers, Innovation, Research Science, and Technology (FIRST) Act**  
<https://beta.congress.gov/bill/113th-congress/house-bill/4186>

**Higher Education Funding Council for England (HEFCE)**  
<http://www.hefce.ac.uk>

**International Standard Name Identifier (ISNI)**  
<http://www.isni.org>

**Jisc Monitor**  
<https://www.jisc-collections.ac.uk/Jisc-Monitor/>

**New Policy for open access in the post-2014 Research Excellence Framework: HEFCE Press Release, March 2014**  
<http://www.hefce.ac.uk/news/newsarchive/2014/news86805.html>

**NISO Access and License Indicators Working Group**  
<http://www.niso.org/workrooms/ali/>

**Office of Science and Technology Policy (OSTP) Memorandum: Increasing Access to the Results of Federally Funded Scientific Research**  
[http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

**Open Access and research funding by the DFG (German Research Foundation)**  
[http://www.dfg.de/en/magazine/spotlight/open\\_access/](http://www.dfg.de/en/magazine/spotlight/open_access/)

**Open Access Key (OAK)**  
<https://www.openaccesskey.com>

**Open where possible, protected where needed: NWO and Open Access. Netherlands Organisation for Scientific Research, November 2012.**  
<http://www.nwo.nl/binaries/content/assets/nwo/documents/nwo/open-access-flyer-2012-def-eng-scherm.pdf>

**ORCID**  
<http://orcid.org/>

**PubMed Central (PMC)**  
<http://www.ncbi.nlm.nih.gov/pmc/>

**The Research & Innovation Performance of the G20. Thomson Reuters, March 2014.**  
<http://sciencewatch.com/sites/sw/files/images/basic/research-innovation-g20.pdf>

**RCUK (Research Council UK) Policy on Open Access**  
<http://www.rcuk.ac.uk/research/openaccess/policy/>

**Ringgold**  
<http://www.ringgold.com>

**ROARMAP (Registry of Open Access Mandatory Archiving Policies)**  
<http://roarmap.eprints.org/>

**SAGE research funder mandate compliance information for authors**  
<http://www.sagepub.com/author-info.sp>

**Scopus**  
<http://www.elsevier.com/online-tools/scopus>

**SHERPA/FACT (Funders & Authors Compliance Tool)**  
<http://www.sherpa.ac.uk/fact/>

**Wellcome Trust Open Access Policy**  
<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/>



RELEVANT  
LINKS



# The Need for Research Data Inventories and the Vision for SHARE

CLIFFORD LYNCH

*Disclaimer & Disclosure: I am a member of the SHARE steering group. SHARE's design is still actively evolving and undergoing prototyping and validation, and what I describe here are a mixture of my own ideas about SHARE and the broader enterprise of research data management, as well as fundamental functions that have already been adopted explicitly into plans for the SHARE system.*

There is a major movement calling for public access<sup>1</sup> to the results of funded research, both in the US and globally. These results include both publications (most notably journal articles) and underlying observational or experimental data. In the US, the funders include federal agencies (where the White House Office of Science and Technology Policy is coordinating a government-wide effort to open up federally funded research), state governments, and private foundations.

In parallel with these developments has been a growing focus on the importance of research data management across all fields of scholarship. That is to say—essentially the idea that appropriate stewardship of data used in or arising from research is essential to preserving, communicating, and replicating scholarship and that, in fact, great opportunities exist to improve the pace and effectiveness of scholarly inquiry broadly if relevant data can be discovered, reused, recombined, and re-purposed in creative ways. Funders and disciplinary scholarly communities have also taken measures to advance these ideas.

With the broad adoption of these ideas, it has become clear that the research and higher education community needs to better understand and manage the research outputs that it produces. SHARE (SHared Access Research Ecosystem) is a joint project of the Association of Research Libraries (ARL) and the two key higher education presidential associations, the Association of American Universities (AAU) and the

Association of Public and Land-grant Universities (APLU); ARL, with generous grant funding from the Alfred P. Sloan foundation and the US Institute for Museum and Library Services (IMLS), is leading the implementation effort. My own organization, the Coalition for Networked Information (CNI), with its deep expertise in both research data management and emerging developments in scholarly practice and scholarly communication, is also helping through its participation on the SHARE project steering group.

What I want to do here is to briefly summarize the potential role of SHARE in the overall scheme of managing research data, with some emphasis on the importance of standards (both existing and to be developed) for making this vision a reality. Note that there are parallel efforts within the SHARE development to address research publications, but I won't discuss those further here.

Most fundamentally, SHARE functions as an *inventory* of research data that is produced by scholars within the higher

CONTINUED »

# SHARE

Great opportunities exist to improve the pace and effectiveness of scholarly inquiry broadly if relevant data can be discovered, reused, recombined, and re-purposed in creative ways.

education community. The system would ultimately include data elements such as what the data is; who created it and their affiliations; what organization and what program or grant funded its creation or capture (if any); where it is currently stored; who is funding the management of the data and how long that funding is guaranteed; and some notes on any access or use restrictions (e.g., embargoes, human subject constraints) that may apply to the data. Populating all of these data elements will require integration of a substantial number of different data sources, and in the early days of SHARE will be sparse; this will improve over time, as both the data gathering system and the sources it gathers from evolve.

SHARE is not itself a repository for data, but simply a place to record deposits and associated metadata. It is agnostic to the use of any specific repository and indeed seeks to span as many repositories as possible. These will include disciplinary, institutional, and funder-provided repository services.

Note that while this sounds simple, it is rife with scoping challenges that will need to be sorted out. Only a modest part of research data is “files” or “datasets” coming from individual investigators; often investigators contribute to very complex shared or pooled community scientific information systems (e.g., Genbank, the Protein Data Bank, the Astrophysics Data System, etc.) and how to reflect these contributions is unclear— as is how to reflect the ongoing stewardship of such data, which depends on the assurance of sustained support for these complex community data systems more broadly. There is also observational or cultural data that is collected and stewarded by a great assortment of entities (including research libraries on behalf of one or more scholarly communities), or that may even support a multiplicity of scholarly, commercial, and broader public uses: synoptic sky surveys, Web crawls, weather, geospatial and remote sensing data, and the Twitter archive. Projects and collaborations span institutional and national boundaries: scholarship is a global undertaking.

Contributors or co-contributors of data include not just academics but government, research, and even commercial groups (consider the pooling of information now occurring between major drug companies and academic researchers, for example). Exactly what should be represented in the inventory?

In the SHARE architectural model, this inventory is stored in a component called the registry. The registry is “fed” by a series of services that make up the notification component, which gathers data from many sources and can also redistribute that data to other interested “subscribers” besides the registry. As data is fed from the notification system into the registry, efforts are made to normalize and consolidate data, which will be an ongoing challenge. It is very likely that there will be functions within the registry, as well, that try to continue to improve the quality of data normalization and consolidation.

Data picked up by the notification system can come either from external events occurring in environments that have been modified to post these events to SHARE, or from software that harvests metadata from the catalogs associated with existing repositories, for example. Events of interest might include the award of grants; the submission of progress reports to funders or achievement reports to host institutions; deposits of data to various repositories or scholarly information systems; the acceptance of a data management plan (hopefully with some of that data being in structured form that can allow the identification of intent to create and deposit data as part of a funded project—imagine building this into widely deployed tools like DMPTool); citation of deposited data in the literature; and reappraisal events and transfers of stewardship responsibilities. Clearly, the system relies upon a mass of standards (existing, under development, and/or as yet undefined) for harvesting, for structuring data, and for “vocabularies” for purposes like the identification



of organizations and funding sources. Simply enumerating relevant current standards and standards efforts would take an article longer than this one.

Personally, I am convinced that in the emerging world of international research data management, we are going to see more movement of data from one repository to another, and transfer of stewardship responsibility or funding sources to underwrite ongoing management—much more often than we are accustomed to as we have managed the traditional base of research publications. It is already common to make research data available for limited time periods through pre-funding built into grant budgets, setting up the need for periodic re-appraisal, and transfer of stewardship, though it is unclear who will conduct this or how it will be done. But the type of inventory envisioned as a core part of SHARE will be essential to managing these processes on a multi-disciplinary and multi-institutional large-scale basis.

Complementing the notification and registry components of the system are the discovery services; many of these services will simply incorporate data extracted from the registry into other discovery services within the research data management ecosystem. Because SHARE is so fundamentally and broadly multi-disciplinary in its coverage, I suspect that most researchers working in one or two specific disciplines will gravitate towards discovery tools (perhaps, for example, associated with specific disciplinary repositories or clusters of such repositories) that are optimized to understand

the knowledge organization practices, ontologies, and vocabularies of specific disciplines. There will need to be at least some basic query interfaces to the registry itself, of course, to allow the most precise searching feasible on some structured data elements, such as funding sources.

A system like SHARE will be useful for many purposes. First and foremost, it will give researchers new tools to manage and reuse vital research data. It will help funders to understand the impact and outcome of their funding programs. It will help those responsible for the stewardship of scholarship to manage processes like reappraisal and transfer of stewardship. It will also provide visibility in the scale of current investment and future obligations related to the management of research results and outcomes, and help to clarify the rate of growth of these obligations.

PI | doi: 10.3789/isqv26no2.2014.05

---

**CLIFFORD LYNCH** (clifford@cni.org) is Executive Director at the Coalition for Networked Information (CNI), an adjunct professor at Berkeley's School of Information, co-chair of the National Academies Board on Research Data and Information (BRDI), and a member of the Steering Committee of the SHared Access Research Ecosystem (SHARE).

*My thanks to Elliott Shore and Eric Celeste for very helpful comments on an earlier draft of this; Diane Goldenberg-Hart helped immensely with the final version.*

---

<sup>1</sup> At least in the United States, federal funders in particular have used the term "public access" rather than the related "open access" to describe their goals. The distinction and ambiguities here are important but beyond the scope of this short article. Note, as discussed later, that the SHARE system is agnostic to access limitations.

<sup>2</sup> Data registries similar to SHARE are under consideration in several other nations at present, and one urgent open question to be explored is how these systems should best interconnect or interoperate.

**Association of American Universities (AAU)**

<https://www.aau.edu/>

**Association of Public and Land-grant Universities (APLU)**

<http://www.aplu.org/>

**Association of Research Libraries (ARL)**

<http://www.arl.org/>

**Astrophysics Data System**

<http://adswww.harvard.edu/>

**Coalition for Networked Information (CNI)**

<http://www.cni.org>

**DMPTool**

<https://dmp.cdlib.org/>

**Genbank**

<http://www.ncbi.nlm.nih.gov/genbank>

**Protein Data Bank**

<http://www.rcsb.org/pdb/home/home.do>  
SHARE (SHared Access Research Ecosystem)  
<http://arl.org/share>

**White House Office of Science and Technology Policy (OSTP)  
Memorandum: Increasing Access to the Results of Federally  
Funded Scientific Research**

[http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)



RELEVANT  
LINKS



Alice Meadows

Howard Ratner

# CHORUS Helps Drive Public Access

ALICE MEADOWS AND HOWARD RATNER



Public access to published research is growing quickly. Initiatives like CHORUS—which went into production in July 2014—and the US Department of Energy’s Public Access plan—announced in August 2014—are just two examples of the accelerating projects in the public access space. The continuing move to online publication of research, exponential increase in data collection, and expansion of public access mandates globally are just some of the factors that make scholarly communications so much more complex today than they were 20 years ago.

As a result, publishers have invested significantly in tools and services to help make the research process more efficient and effective—from manuscript submission systems to CrossRef DOIs to ORCID unique identifiers for researchers. CHORUS—the Clearinghouse for the Open Research of the United States—is a new, not-for-profit, publisher-led initiative designed to help implement public access to the articles resulting from US federally-funded research, as required by the Office of Science and Technology Policy’s February 2013 memo.

The CHORUS pilot project launched in Fall 2013. On July 31, 2014, CHORUS went into production and a few days later, on August 4, the US Department of Energy (DOE), announced that it would “collaborate with CHORUS on our implementation of public access to scholarly publications resulting from DOE-funded research.”

CHORUS supports public access to federally funded research by acting as an information bridge, linking the public to freely accessible journal articles directly on publisher platforms, where the articles can be read and preserved in their scholarly context. Its open technology platform leverages publishers' existing infrastructure, avoids duplication of effort, minimizes cost to the government, taxpayer, and grantee institutions, and ensures the continued availability of the research literature.

### CHORUS Provides Five Core Functions:

- **Identification** - One of the major challenges for funders is simply identifying which published articles have resulted from their funded research. Most agencies don't have good systems for tracking this information themselves and, until recently, researchers have not had an easy way to provide this information. Now though, thanks to the introduction of CrossRef's FundRef service, solutions exist for both of these problems. As of the time of writing, FundRef includes information on well over 7,000 funders globally. CHORUS incorporates the relevant data on US funders. All the researchers need to do is to name their funding source during the submission process. This adds the relevant metadata, which can trigger public access to the article.
- **Discovery** - CHORUS has been designed to facilitate discovery of the latest research articles via agency portals and common search engines, as well as through its own Search application. To date, CHORUS has taken advantage of existing open application programming interfaces (APIs) from organizations like CrossRef and ORCID. Plans are underway to issue CHORUS-specific APIs to help innovators create new tools and functionality that further support public access. CHORUS is leveraging existing tools, like CrossRef's new Text and Data Mining services, to eliminate much of the time and effort that has been spent in the past to set up machine access to scholarly content. CHORUS directly commented on the draft of the NISO Open Access and Metadata Indicators Working Group (recently renamed to Access License and Indicators; see separate article on page 35) and is eager to make best use of the forthcoming Recommended Practice within CHORUS.

# CHORUS

CHORUS supports public access to federally funded research by acting as an information bridge, linking the public to freely accessible journal articles directly on publisher platforms, where the articles can be read and preserved in their scholarly context.

- **Access** - Even when readers have found the article(s) they need, it's not always easy to find the best available version. Multiple versions may exist online, some of which may be pre-publication prints and could contain errors or misinformation that were corrected in the published or later version. CHORUS points users to the best available version (accepted author manuscript or Version of Record) of articles on their publisher publication sites, where essential context, tools, and information, either immediately at publication or after an embargo period, are openly accessible.
- **Preservation** - Everyone involved in scholarly communications recognizes the importance of preserving access to research. CHORUS ensures the integrity and sustainability of the scholarly record through partnerships with CLOCKSS, Portico, and other services that archive and preserve research articles in perpetuity. These agreements have been set up to ensure that, irrespective of journal transfers, evolving publisher policies, or other future changes, readers will continue to have access to the articles arising from publicly funded research.
- **Compliance** - Tracking compliance with funder requirements for authors to identify funded articles when published is another critical issue for funders and institutions alike. By integrating FundRef data from CrossRef, CHORUS makes it easy for authors to comply with these requirements, while the CHORUS Dashboard service facilitates monitoring and reporting by funders and institutions, without adding unnecessary costs and administrative overhead.

CONTINUED »

## Next Steps

Even though CHORUS is now officially in production, there is still much work to be done. For example, the CHORUS Search tool is currently in beta. Results returned identify articles that report on agency funded research from our growing database, but these articles may or may not be publicly accessible at this time as the implementation of various agency policies is still underway. In the near future, CHORUS will introduce a method to clearly identify publicly accessible articles in search results. Other areas of focus for development include work on a standard way to surface publishers' article reuse terms, improved dashboards for monitoring the status of articles in the CHORUS system, integration with the SHARE notification system (see separate article on SHARE on page 29), and better integration with our dark archive partners.

The new CHORUS website will feature improved navigation, more information, and seamless integration with the dashboard and search services—all branded with the new visual identity. Very importantly, it will support CHORUS' membership marketing and activities. As a new organization, CHORUS has benefited—and continues to benefit—from the generous support of a number of publishers and publishing organizations. We are now introducing a range of membership options to help ensure the future sustainability of CHORUS

and its parent organization, CHOR, Inc. These include sliding-scale, fee-based Publisher Membership and Affiliate Membership (for non-publisher organizations and businesses involved in scholarly publishing), as well as gratis Funder Partner and Academic Supporter options.

Continuing to collaborate over the coming months with funders and with other organizations such as SHARE—with whom we are already working on common standards and technologies for persistent identifiers and metrics—will also be critical to CHORUS's future success. Looking further ahead, while CHORUS is focused on providing a neutral, stable, and effective platform to help increase public access to peer-reviewed publications arising out of US Government-funded research, addressing connections to publicly accessible data and international concerns are also being investigated.

PI | doi: 10.3789/isqv26no2.2014.06

---

**ALICE MEADOWS** (ameadows@wiley.com, ORCID: 0000-0003-2161-3781) is Director of Communications, Global Research, Wiley and Chair, CHORUS Communications Working Group. **HOWARD RATNER** (hratner@chorusaccess.org, ORCID: 0000-0002-2123-6317) is Executive Director, CHOR, Inc., the non-profit organization that provides the CHORUS service.

### CHORUS

<http://chorusaccess.org/>

### CrossRef

<http://www.crossref.org/>

### FundRef

<http://www.crossref.org/fundref/index.html>

### NISO Access License and Indicators Working Group

<http://www.niso.org/workrooms/oami/>

### ORCID

<http://orcid.org/>

### Office of Science and Technology Policy's February 2013 memo

[http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

### US Department of Energy Public Access plan

<http://www.energy.gov/downloads/doe-public-access-plan>

### US DOE and CHORUS collaboration announcement

<http://chorusaccess.org/july-2014-chorus-to-work-with-DOE/>

### CLOCKSS

<http://www.clockss.org/clockss/Home>

### Portico

<http://www.portico.org/digital-preservation/>

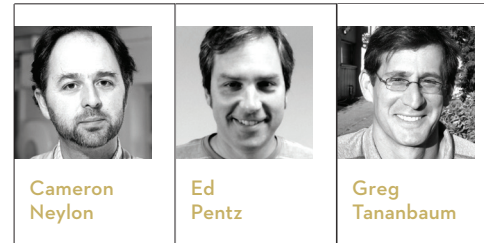
### SHARE

<http://www.arl.org/focus-areas/shared-access-research-ecosystem-share>



RELEVANT  
LINKS





CAMERON NEYLON, ED PENTZ, AND GREG TANANBAUM

# Standardized Metadata Elements to Identify Access and License Information

Many journal articles are available from publishers under the banner of Open Access (OA), Public Access, or similar names. The meanings of these terms vary both between publishers and within publishers by journal—and in some cases, based on the funder. Adding to the potential confusion, a number of publishers also offer hybrid options, in which one or more articles in a journal are freely accessible, while the rest of the content in that journal remains under subscription control.

The guide *HowOpenIsIt?* from SPARC, PLOS, and OASPA depicts a continuum of openness that also varies by the rights accorded to readers, reuse rights, copyrights, author posting rights, automatic posting, and machine readability. Clearly, as the Guide points out, “not all Open Access is created equal.” Currently, there is no standard metadata in use that succinctly defines these various levels of openness and licensing. As a result, readers are often unaware of the free-to-read status of specific articles and downstream users are unsure of the reuse rights, if any. Authors have difficulty determining what rights they will retain and whether they are compliant with a given funder policy. Aggregators and service providers have no machine-readable mechanism for identifying articles that can be legitimately harvested.

In January 2013, NISO Voting Members approved a new work item proposal to develop a Recommended Practice on Open Access Metadata and Indicators (later re-named Access and Licensing Indicators) to address this gap. The goal of the project was to identify a standardized set of metadata elements to describe both the accessibility of a specific article and the available reuse rights.

CONTINUED »



The Working Group specifically decided against proposing metadata items that were labeled or named “Open Access” due to the many different definitions of this term.

A draft for comments Recommended Practice was issued in January 2014 proposing the adoption of two core pieces of metadata that can be transmitted through existing channels:

» **Free-to-read** (<free\_to\_read>) -

A simple status that defines whether the work is accessible, without charge or other restriction (such as registration), to read online. This tag has two defined attributes that should be used, if applicable, to indicate start and end dates. Start and end dates would accommodate delayed access models (embargoes) and special offers where content was free-to-read for a period of time or after a particular date. The absence of both a start and end date would mean a permanent state of free-to-read access.

» **License reference** (<license\_ref>) -

A reference to a URI that carries the license terms specifying how a work may be used. There are no limitations on the license specified or on the terms contained within the license. Multiple license reference elements can be provided. Each of these may have a different start dates to address embargoes or how usage rights change over time. There is no corresponding end date attribute for the <license\_ref> element, because including end dates could introduce ambiguities. The data within this tag should be a stable identifier expressed as an HTTP URI, the maintenance of which would be the responsibility of the platform making the content available.

The Working Group specifically decided against proposing metadata items that were labeled or named “Open Access” due to the many different definitions of this term, as discussed above. Instead, the chosen approach was to provide factual metadata to be disseminated to enable people and machines to make decisions about how they can use the content. With widespread implementation of these recommended metadata tags, humans and machines will be able to assess the accessibility and reuse rights associated with a given article.

The Working Group considered and rejected the expression of reuse rights in the actual metadata. These rights could vary depending on who the user is and it could be difficult to fully and accurately express them in metadata, possibly creating a conflict or inconsistency with the actual license. Therefore,

the agreed approach was to have a reference in the metadata to the license that would be posted separately and linked from the metadata reference.

It is the view of the working group that these two metadata elements can cover most current use cases of delayed access and of license terms that activate at a particular time post publication. Use cases fully addressed include:

- » End user seeks to discover, identify, and access free-to-read items
- » End user seeks to know the readability status of an item
- » End user seeks to know reuse permissions of an item
- » End user seeks to know reuse permissions of a sub-component of an item
- » Repositories seek to expose free-to-read items

Use cases that are at least partially addressed by the new elements are:

- » End user seeks to text mine content
- » Ensure author/publisher rights assertions align with license statements
- » Funding agency seeks to track compliance of research outputs to open access mandates
- » Institution seeks to report on open access compliance of research outputs

While it was outside the scope of this Recommended Practice to determine how components of works (e.g., figures, images, datasets, etc.) should be identified, where such components are separately identified, the <free\_to\_read> and <license\_ref> tags can be applied separately to those components.

Wherever possible, creation and population of these elements should become part of standard editorial/production workflows. The metadata should be made an integral part of the feeds to CrossRef and other DOI registration agencies, included alongside (or within) article/chapter content on hosting websites, and delivered in content feeds to third parties. The metadata should be embedded in the content itself along with other metadata; for example, in HTML META tags and in PDF files where bibliographic and other metadata are being included.

The Working Group is also recommending that the “free-to-read” and “license reference” metadata be encoded in XML and included in existing metadata

distribution channels and with the content itself, where appropriate. Thus the <free\_to\_read> and <license\_ref> tags would need to be added to existing schemas and workflows. Publisher or aggregator systems could be programmed to read the tags and display appropriate status icons to users.

It may also be worthwhile for content providers to consider including the metadata elements within other alerting channels, such as e-ToCs and RSS subscription feeds as well as information provided directly to abstracting and indexing services. Whatever channel is used, wider distribution of this (and other) article, chapter, or book metadata is likely to be helpful in driving discovery and usage for the materials concerned.

The Working Group is currently finalizing the Recommended Practice to address issues identified during the public comment period. The final document is expected to be published in the fall of 2014.

The Group recognizes that if the recommendations are adopted, there will need to be further work on implementation and an analysis done on the best way to incorporate the <free\_to\_read> and <license\_ref> metadata into existing formats, such as ONIX, RDF, OAI-PMH, and Dublin Core (DC). NISO will be looking into the need for a Standing Committee to work on these follow-up items. | NR | doi: 10.3789/isqv26no2.2014.07

**CAMERON NEYLON** (cneylon@plos.org) is Advocacy Director for the Public Library of Science (PLOS).

**ED PENTZ** (epentz@crossref.org) is Executive Director, CrossRef.

**GREG TANANBAUM** (greg@scholarnext.com) is Consultant, SPARC Scholarly Publishing & Academic Resources Coalition

*The three are the co-chairs of the NISO Access and License Indicators Working Group (formerly Open Access and Metadata Indicators).*

#### CrossRef

<http://www.crossref.org>

**HowOpenIsIt? SPARC, PLOS, and OSPA, 2013.**

<http://www.plos.org/open-access/howopenisit/>

**NISO Access & Licensing Indicators Working Group webpage**

<http://www.niso.org/workrooms/ali/>



**RELEVANT  
LINKS**



## NISO Publishes Three Recommended Practices on Knowledge Bases, Demand Driven Acquisition of Monographs, and Library Discovery Services

*The National Information Standards Organization (NISO) has published two new and one revised Recommended Practices.*

The revision is to the *Knowledge Base and Related Tools (KBART) Recommended Practice* (NISO RP-9-2014). The original recommended practice, issued in 2010, provided all parties in the information supply chain with straightforward guidance about metadata formatting—focused mainly on journal resources—to ensure the exchange of accurate metadata between content providers and knowledge base developers. Building on the initial recommendations, the revision includes the more granular, complex issues that cause problems in metadata supply, including consortia-specific metadata and metadata transfer for open access publications, e-books, and conference proceedings.

A new recommended practice, *Demand Driven Acquisition (DDA) of Monographs* (NISO RP-20-2014) discusses the DDA method—also referred to as patron-driven acquisition—used by libraries for collection development where monographs are purchased at their point of need when selected by users from a pool of potential titles. NISO's Recommended Practice discusses and makes recommendations for publishers, vendors, aggregators, and libraries about key aspects of DDA, goals and objectives of a DDA program, choosing parameters of the program, profiling options, managing MARC records for DDA, removing materials from the consideration pool, assessment of the program, providing long-term access to un-owned content, consortial considerations for DDA, and public library DDA. Although DDA is more commonly used for e-books, the method can also be applied to print publications and these recommendations provide a single set of best practices for both formats, with articulation of differences where they occur.

The second new recommended practice is *Open Discovery Initiative: Promoting Transparency in Discovery* (NISO RP-19-2014), which provides specific guidelines on participation in the new generation of library discovery services. The goal of the NISO Open Discovery Initiative (ODI) was to develop recommendations that would increase transparency across all aspects of indexed discovery that use an aggregated central index to enable searching across a wide range of library related resources. The Recommended Practice includes guidelines to content providers on disclosure of level of participation, the minimum set of metadata elements provided for indexing, linking practices, and technical formats. Recommendations for discovery service providers address content listings, linking practices, file formats and methods of transfer to be supported, and usage statistics. The document also provides background information on the evolution of discovery and delivery technology and a standard set of terminology and definitions for this technology area. ■

🔗 KBART Recommended Practice available at: [www.niso.org/workrooms/kbart](http://www.niso.org/workrooms/kbart)

Demand Driven Acquisition of Monographs Recommended Practice available at: [www.niso.org/workrooms/dda/](http://www.niso.org/workrooms/dda/)

Open Discovery Initiative Recommended Practice available at: [www.niso.org/workrooms/odi/](http://www.niso.org/workrooms/odi/)





## NISO and OAI Publish American National Standard on ResourceSync Framework Specification

The National Information Standards Organization (NISO) and the Open Archives Initiative (OAI) have published the *ResourceSync Framework Specification* (ANSI/NISO Z39.99-2014)—a new American National Standard for the Web detailing various capabilities that a server can implement to allow third-party systems to remain synchronized with its evolving resources. The ResourceSync joint project, funded with support from the Alfred P. Sloan Foundation and Jisc, was initiated to develop a new open standard on the real-time synchronization of web resources.

Increasingly, large-scale digital collections are available from multiple hosting locations, are cached at multiple servers, and leveraged by several services. Since Web resources are continually changing, this proliferation of content yields the challenging problem of keeping services that leverage a server's evolving content synchronized in a timely and accurate manner.

The ResourceSync specification introduces a range of easy to implement capabilities that a server may support to enable remote systems to remain more tightly in step with its evolving resources. It also describes how a server can advertise the capabilities it supports. Remote systems can inspect this information to determine how best to remain aligned with the evolving data. All capabilities are implemented on the basis of the document formats introduced by the Sitemap protocol. Capabilities can be combined to achieve varying levels of functionality and hence meet different local or community requirements. As a modular specification grounded in protocols that are already widely adopted, ResourceSync can be used to meet a wide variety of use cases. ■

- 🔗 The ResourceSync specification and video tutorials on using the standard are available on the NISO website at [www.niso.org/workrooms/resourcesync/](http://www.niso.org/workrooms/resourcesync/).

## EPUB 3.0.1 Issued by International Digital Publishing Forum

The International Digital Publishing Forum members have approved the updated version 3.0.1 of the EPUB standard as a final Recommended Specification. EPUB 3.0.1 is a minor revision, focusing primarily on bug fixes and errata for the 3.0 specification, together with several minor backwards-compatible additions.

Included among the changes is the unbinding of the EPUB Structural Semantics Vocabulary updating from the EPUB specifications revision cycle. The EPUB Working Group can now vet new property requests and make additions on an ongoing basis. Many new additions were also made to the vocabulary, most notably a new section dedicated to educational properties.

The EPUBCheck validation tool is expected to include complete support for EPUB 3.0.1 by August of this year, in conjunction with which samples for the new features of 3.0.1 are being developed. The EPUB Reading System Test Suite will be updated to include tests for 3.0.1 during the same timeframe. And the Radium open source implementation of EPUB rendering already supports several EPUB 3.0.1 features and expects to deliver full support during this year. ■

- 🔗 Announcement of EPUB 3.0.1: [idpf.org/news/epub-301-approved-as-final-recommended-specification](http://idpf.org/news/epub-301-approved-as-final-recommended-specification)

EPUB 3.0.1 specification: [www.idpf.org/epub/301/spec/epub-overview.html](http://www.idpf.org/epub/301/spec/epub-overview.html)

Summary of changes from version 3.0: [www.idpf.org/epub/301/spec/epub-changes.html](http://www.idpf.org/epub/301/spec/epub-changes.html)

## Library of Congress Identifies Recommended Formats for Long-Term Preservation

The Library of Congress released a set of recommended formats for a broad spectrum of creative works, ranging from books to digital music, to inform the Library's acquisition practices. The format recommendations will help ensure the Library's collections processes are considering and maximizing the long-term preservation potential of its large and varied collections.

### Six categories of creative output are addressed:

- » Textual Works and Musical Compositions - including print, digital, electronic serials, and score-based representations
- » Still Image Works - including print and digital photographs, print and digital other graphic images, and microforms
- » Audio Works - including digital or analog audio on tangible medium and digital audio that is media independent



- » Moving Image Works - including digital and physical media motion pictures and file-based and physical media video
- » Software and Electronic Gaming and Learning
- » Datasets/Databases

The recommendations will enable the Library to identify the preferred format for acquisition when a work is offered in more than one format, but will not result in the exclusion of other formats from consideration for the Library's permanent collections. In addition to informing internal processes, the Library is also making the recommended formats public to inform the creative and library communities of best practices for ensuring the preservation of and long-term access to creative output. ■

🔗 LC Recommended Format Specifications 2014-2015:  
[www.loc.gov/preservation/resources/rfs/TOC.html](http://www.loc.gov/preservation/resources/rfs/TOC.html)

## UKSG Transfer Working Group Announces Improvements to the Code of Practice with Release of Version 3.0

The UKSG Transfer Working Group announced the release of the *Transfer Code of Practice* Version 3.0. The new version has a number of key updates dealing with new content types, clarification of subscriber types, journal URLs and redirects, nomenclature, and the timing and content of communications. Over a period of 18 months the Transfer Working Group revised and improved upon the previous version of the Code (released in September 2008) and sought feedback from the community through a public review process. Publishers will now be encouraged to follow the new version. Those publishers who endorse Transfer's principles by agreeing to align their procedures with the Code, and to apply them in practice, will be considered 'Transfer Compliant'.

The *Transfer Code of Practice* is a set of voluntary guidelines for publishers involved in any journal transfer. It covers difficult issues including ongoing provision of access to online content, exchange of subscriber lists, DOI and URL transfer, as well as perpetual access rights to journal content. Transfer and the Enhanced Transfer Alerting Service were developed in response to the expressed needs of the scholarly journal community for consistent guidelines to help publishers ensure that journal content remains easily accessible by librarians and readers when there is a transfer between parties, and to ensure that the transfer process occurs with minimum disruption. ■

🔗 UKSG Transfer Code of Practice: [www.uksg.org/transfer](http://www.uksg.org/transfer)

| NW | doi:10.3789/isqv26no2.2014.08



# SD [ STANDARDS IN DEVELOPMENT: June 30, 2014 ]

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/](http://www.niso.org/workrooms/)) and the *Newsline* quarterly supplements, *Working Group Connection* ([www.niso.org/publications/newsline/](http://www.niso.org/publications/newsline/)), for updates on the working group activities.

Note: DSFTU stands for Draft Standard for Trial Use.

WORKING GROUP	STATUS
<b>Access and License Indicators</b> (formerly <b>Open Access Metadata and Indicators</b> ) Co-chairs: Cameron Neylon, Ed Pentz, Greg Tananbaum	Recommended Practice (NISO RP-22-201x) being finalized for publication following the public comment period.
<b>Demand Driven Acquisition of Monographs</b> Co-chairs: Michael Levine-Clark, Barbara Kawecki	Recommended Practice <i>Demand Driven Acquisition of Monographs</i> (NISO RP-20-2014) published.
<b>Journal Article Versions (JAV) Addendum</b> Chair: open	Revised Recommended Practice (NISO RP-9-201x) in development.
<b>Knowledge Base and Related Tools (KBART) Phase II</b> Co-chairs: Magaly Bascones, Chad Hutchens	Revised Recommended Practice <i>Knowledge Base And Related Tools (KBART)</i> (NISO RP-17-2014) published.
<b>Open Discovery Initiative</b> Co-chairs: Marshall Breeding, Jenny Walker	Recommended Practice <i>Open Discovery Initiative: Promoting Transparency in Discovery</i> (NISO RP-19-201x) published.
<b>Protocol for Exchanging Serial Content</b> Co-chairs: Leslie Johnston, Kimberly Tryka	Recommended Practice (NISO RP-23-201x) in development.
<b>Resource Synchronization</b> Co-chairs: Herbert Van de Sompel, Todd Carpenter	<i>ResourceSync Framework Specification</i> (ANSI/NISO Z39.99-2014) published.
<b>Standard Interchange Protocol (SIP)</b> Co-chairs: John Bodfish, Ted Koppel	Standard (NISO Z39.100-201x) in development.
<b>SUSHI Lite</b> Co-chairs: Paul Needham, Oliver Pesch	Technical Report (NISO TR-06-201x) in development
<b>SUSHI Standing Committee</b> Co-chairs: Marie Kennedy, Oliver Pesch	Revision of the SUSHI Protocol standard (Z39.93-201x) in development.
<b>US Profile of ISO 3166 Country Codes</b> Chair: TBD	Working group being formed to develop standard (Z39.101-x)



# NISO

NISO helps you and your organization connect with the information world.

NISO crosses over the library, publishing, and information systems communities, providing an excellent forum for these groups to agree on solutions.



**JOIN NISO**  
and play an  
active role in  
standards  
solutions!

[www.niso.org](http://www.niso.org)