

Access, preservation and analysis in a consortial journal archive: the evolution of Scholars Portal Journals

This article discusses Scholars Portal Journals (SP Journals), a library consortium-run platform that aggregates and archives licensed scholarly journal content in the province of Ontario, Canada. Born in the early days of e-journals out of a need to provide consistent and long-term access to scholarly materials in the sometimes volatile world of online publishing, SP Journals has evolved into a major digital repository and archive. With over 55 million full-text articles and serving a student population of just under half a million, SP Journals represents a major investment in access to online scholarship. This article explains the lifecycle of content on the platform, from initial publisher negotiations to delivering usage reports, and discusses considerations of running a locally hosted journal platform.

Keywords

Consortium; e-journals; preservation; repository; licensing; e-resources

Introduction: OCUL and Scholars Portal

The Ontario Council of University Libraries (OCUL) is a consortium representing the libraries of all 21 universities in the province of Ontario, Canada. Collectively, these universities have a student population of over 480,000, around one third of the university population of Canada. As the digital services arm of OCUL, Scholars Portal (SP) provides shared infrastructure to OCUL libraries. OCUL's largest member, the University of Toronto, acts as a service provider for SP. Scholars Portal staff are employees of the University of Toronto, and our offices are located in the main library.

A main component of SP's work is preserving and providing access to digital content licensed by OCUL, including e-books, e-journals, microdata and geospatial data.

The platform

Scholars Portal Journals (SP Journals) is a digital repository of academic journal articles providing students, faculty and researchers at Ontario's universities easy access to articles on a wide range of academic subjects through a bilingual search interface. The platform grew out of a locally hosted journal repository first deployed at the University of Toronto Libraries in the 1990s, became an OCUL service in 2002, and was completely redesigned to adopt an early version of the journal archiving tag suite (JATS) metadata standard in 2006.

When SP Journals launched as an OCUL service in 2002, it hosted a modest collection of just over one million articles from a few major publishers. Currently, SP Journals includes about 55 million scholarly articles drawn from 22,549 full-text journals covering every academic discipline. Based on a MarkLogic XML database, SP Journals is designed to provide a single interface for accessing academic journal content that is essential for teaching, learning and research. Moreover, as a trustworthy digital repository (TDR) certified by the Center for Research Libraries, the platform is built to secure the proper preservation of content for the long term.



SABINA PAGOTTO

Client Services &
Assessment Librarian

Scholars Portal
Ontario Council of
University Libraries



WEI ZHAO

Senior Metadata
Librarian

Scholars Portal
Ontario Council of
University Libraries

- 2 SP staff have developed a number of features to improve user experience on the platform. For example, the details page for an article displays a list of other articles on the platform in which it is cited, as well as a list of related articles. Users can save articles to a reading list, export reference lists using QuikBib, or sign up to receive e-mail alerts for their saved searches.

The content

Subscription content

The majority of the journals on SP Journals are subscription-based titles licensed from large publishers. Much of this licensing is done consorcially by OCUL, as SP provides shared infrastructure specifically to support the consortium. Some of the larger journal packages are licensed by the national university library consortium, the Canadian Research Knowledge Network (CRKN), of which all OCUL libraries are also members. Licensed content on SP Journals is currently only available to OCUL institutions, and not the other members of CRKN. Some smaller deals have been negotiated by individual universities for their own usage only.

While many OCUL member libraries prefer to direct their users to SP Journals when possible, the content in these journal packages is also available to authorized users on the publishers' own platforms.

Publishers must agree to specific terms in order to have their content hosted, archived and accessed locally on SP servers (known as 'local loading'). These terms involve granting Scholars Portal several rights including the right to archive and permit authorized users to access the licensed content on our secure network, the right to migrate the content to new formats in response to technological changes, and the right to create derivatives of the licensed content and its metadata as needed to match our technology infrastructure. Local hosting terms are sometimes part of the license agreement, but more often form a separate local load agreement alongside the license agreement. A lighter agreement that involves conferring fewer rights to SP may be a preferable solution for content providers who are hesitant to hand over control of their content, or who do not hold all the rights to the content they are distributing. The content is still discoverable from the SP Journals platform, but it is not preserved as part of our TDR.

Since SP is not directly involved in negotiating licenses for content, local load agreements (or local load clauses in license agreements) are typically negotiated in conjunction with those who license scholarly content for their libraries: OCUL's Information Resources Committee, CRKN's Content Strategy Committee and collections librarians at individual universities. The local load agreement has three signatories: the publisher, the service provider and the licensing body.

'The majority of the journals ... are subscription-based titles licensed from large publishers'

'The local load agreement has three signatories: the publisher, the service provider and the licensing body'

Open access content

SP Journals also hosts a growing number of open access (OA) journals. Bringing this content together onto one platform can help with discoverability and ease of access with a consistent interface, and preserving OA journals is particularly important since the financial model means that the editors or publishers may not be able to do it themselves. Depositing content for long-term digital preservation is one of the criteria for the DOAJ Seal of Approval for Open Access Journals.¹

The total number of OA articles is currently close to 1.7 million. This includes large OA collections from PubMed Central and DeGruyter Open, hybrid OA content from commercial publishers, and independent Canadian OA journals. SP hosts several instances of Open

- 3 Journal Systems (OJS) software on behalf of OCUL member libraries, and many of the journals published there are also ingested into the SP Journals platform for discovery and preservation.

There are also special cases involving open archives or backlists. For example, the archive of PsycCRITIQUES, a database of psychology-related reviews maintained by the American Psychological Association but discontinued in December 2017, was made available on SP Journals as a fully OA resource in spring 2018.

While OA content on SP Journals is open to all, the publisher or rightsholder must still sign a local loading agreement granting us the rights to archive the content and migrate to new formats before we may ingest.

The workflow

Receiving data

As part of the local load discussions, publishers must indicate how and in what format they will convey their content to SP. The publisher can choose to push their content to SP's FTP account or make the content available on their own FTP site for downloading. Our FTP script downloads, decompresses and organizes the data as a data set, and then moves the data from the FTP location to the file system on the storage server. The FTP script is run daily, weekly or monthly depending on the amount of the publisher's data.

'While OA content ... is open to all, the publisher or rightsholder must still sign a local loading agreement'

Data transformation and enrichment

SP Journals ingests data from 40 publishers. In addition to PDF full text, the metadata provided by the publishers are in XML or SGML format using different DTDs or schemas. The publishers' native data is transformed to the JATS format in order to normalize data elements for archiving, display and searching. Customized tags are added and local rules and policies are applied to tag the document in addition to those imposed by JATS.² The XML full text is also transformed when it is available and rendered to users as XHTML.

The data transformation is processed in two steps: mapping and coding. First, the metadata librarian undertakes an intensive analysis of each publisher's source data format from the source DTD, schema and sample data and then develops a crosswalk. The crosswalk includes the mapping of the path from source to target data and the explanation of decisions and compromises. Second, the programmer with coding experience develops a loader according to the crosswalk using Java and XSL transformation. A test environment is set up so the transformations are tested before the data is loaded into production. The metadata librarian inspects the output and the crosswalk can go through several iterations to make sure the data are transformed completely and explicitly. After the loader is directed into the production system, the DTD validation is enforced and the transformation of each data set is logged for any errors. The log files are examined by quality assurance personnel. Any data set with errors is then removed from the production environment and reloaded after the problems have been fixed. The transformed JATS XML files are then stored in the MarkLogic database for display and searching, while the publisher digital objects (including XML, PDF and supplementary materials of figures and tables) reside on the file system for long-term preservation.

The ingest process overview (Figure 1) shows the different aspects of the digital object's journey from the time it is ingested into the repository to the time it is made accessible to the designated community.

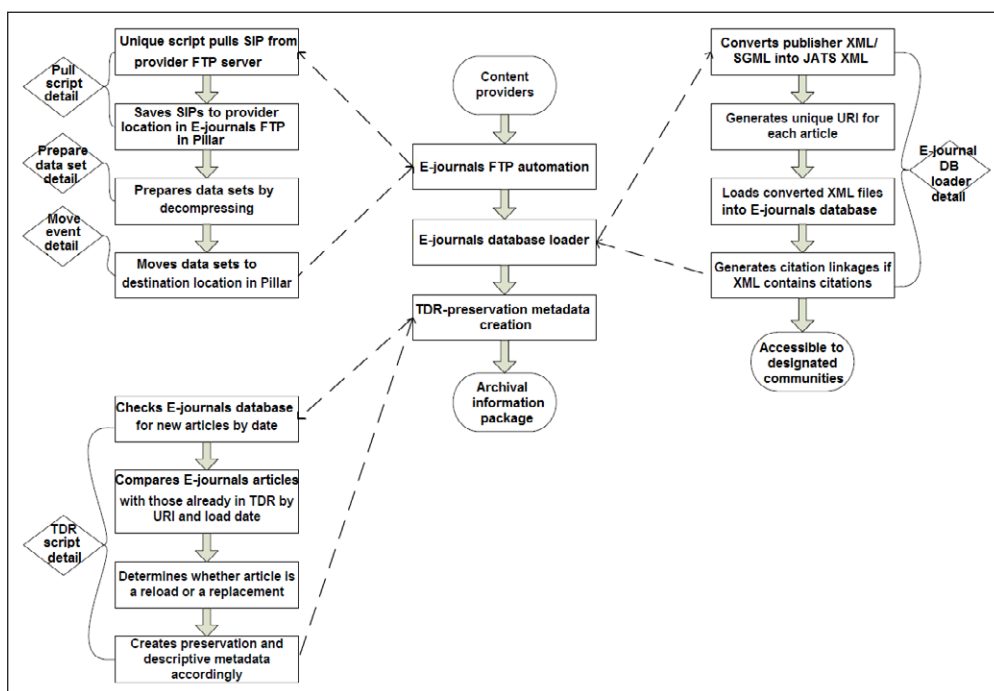


Figure 1. An overview of the Scholars Portal ingest process

Quality assurance

SP has established stringent standards to ensure the high quality of our resources and services. A series of procedures and tools have been implemented throughout the workflow to enable this high quality.³ A team of one metadata librarian and two programmers is responsible for the operation of daily data loading and quality assurance. Additionally, a 'Report a problem' button within the interface allows the end user to report any problems from the point of access, and our client services team responds as quickly as possible so that the user can access the articles they need.

Access management

The SP Journals platform is designed to ensure appropriate full-text access to authorized users, as subscriptions vary among member libraries. There are a number of models for licensing journal content. A popular one, especially during the time of SP Journals' development, is the 'big deal' which was defined by Bergstrom and colleagues as 'contracts for bundled access to a publisher's entire journal list'.⁴ According to their study, the majority of North American libraries have subscribed to bundled contracts with large commercial journal publishers. However, the current e-journal landscape reveals a shift away from big deals. Increasingly, the publishers break their title lists into several subject areas or even allow the libraries to customize their own title list for purchase, which makes tracking entitlements tricky. The merge of journal publishers, the transfer of journal titles from one publisher to another, and post-cancellation perpetual access rights add even more layers of complexity to entitlement management.

'Secure authentication and entitlement processes can restrict access at the level of individual articles'

The entitlement management model in MarkLogic, which combines information about the collection and library subscriptions, allows SP staff to capture information about the variety of purchase models and provides the flexibility to manage the entitlement at both the collection and journal level. The model allows for the creation of customized title lists, for the collection to be added to or removed from entitlement lists, for the journals to be moved from one collection to another, and date ranges to be applied to collections and individual journals. Secure authentication and entitlement processes can restrict access at the level of individual articles.

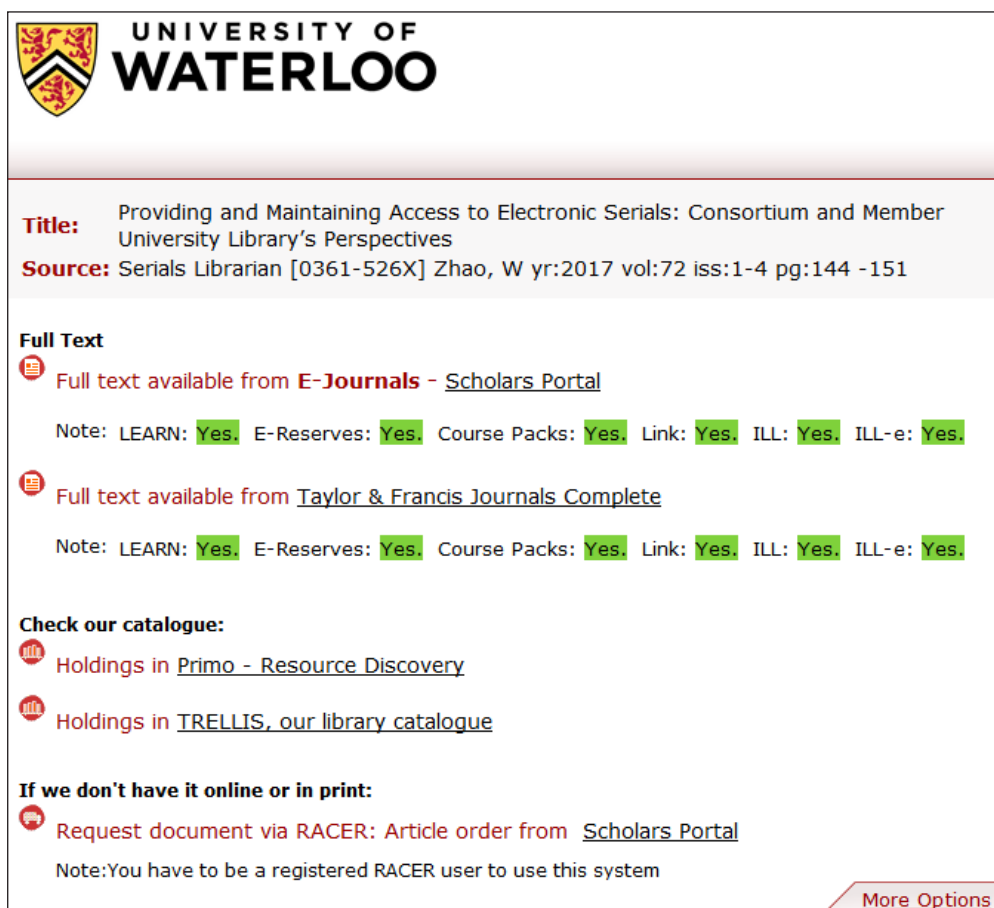
Authentication

SP Journals supports IP-based authentication and maintains current IP ranges for all OCUL institutions. Since the majority of access comes from off campus,⁵ users at unrecognized IP addresses are prompted to authenticate when they attempt to access the full text of an article. Authentication links connect the user either to their institution's off-campus proxy login page, or to a Shibboleth login. SP has worked with the Canadian Access Federation, which co-ordinated federated identity management across Canadian universities, to implement this type of login.⁶

Supporting discovery and usage

SP Journals holdings are exported to SFX, Serial Solutions and EBSCO so the access can be directed from these OpenURL services. The holding records are also contributed to the Keepers Registry and The Print Archives and Preservation Registry (PAPR).

To support discovery and access, SP hosts several instances of Ex Libris's SFX OpenURL link resolver. A central instance contains holdings information for the SP Journals platform and is directly managed by SP staff, while local instances for a number of member libraries are hosted by SP but managed by local staff.⁷



UNIVERSITY OF WATERLOO

Title: Providing and Maintaining Access to Electronic Serials: Consortium and Member University Library's Perspectives
Source: Serials Librarian [0361-526X] Zhao, W yr:2017 vol:72 iss:1-4 pg:144 -151

Full Text

Full text available from **E-Journals** - [Scholars Portal](#)

Note: LEARN: **Yes.** E-Reserves: **Yes.** Course Packs: **Yes.** Link: **Yes.** ILL: **Yes.** ILL-e: **Yes.**

Full text available from [Taylor & Francis Journals Complete](#)

Note: LEARN: **Yes.** E-Reserves: **Yes.** Course Packs: **Yes.** Link: **Yes.** ILL: **Yes.** ILL-e: **Yes.**

Check our catalogue:

Holdings in [Primo - Resource Discovery](#)

Holdings in [TRELIS, our library catalogue](#)

If we don't have it online or in print:

Request document via **RACER: Article order from** [Scholars Portal](#)

Note: You have to be a registered RACER user to use this system

[More Options](#)

Figure 2. Usage terms from the OUR database embedded into the SFX menu

Another service provided by SP to support usage of the Journals platform, as well as other electronic resources, is the OCUL Usage Rights (OUR) database. An implementation of the University of British Columbia's open source Mondo software, OUR displays the usage terms for each licensed product. License information for consortial products is entered by CRKN or OCUL staff, then activated by the individual subscribing library. Libraries can also add records for licenses negotiated locally. SP staff have worked to create linkages between OUR and SFX and other link resolvers, so that the terms of use display in the link resolver menu as the user is about to click through to the full text (Figure 2).

Standards of preservation and reporting

TDR certification

From 2012 SP started to use in-house developed scripts and existing tools such as the file information tool set (FITS) to perform digital preservation activities on journal articles. SP uses a structural metadata scheme based on a simplified metadata encoding and transmission specification (METS) profile as a robust and flexible way to structurally define the content object. It serves as a container for all of the metadata about the object. Within this container, SP uses the preservation metadata implementation strategy (PREMIS) vocabulary, which provides ways of describing objects and processes that are necessary for digital preservation. The preservation metadata is displayed alongside the descriptive metadata to help end-users verify the history and integrity of the digital objects. It also helps researchers, librarians and managers to identify the digital preservation gaps and problems (Figure 3).

In 2013 SP Journals was certified as a TDR by the Center for Research Libraries – the first repository in Canada to achieve this status. This was the culmination of a lengthy process involving extensive documentation and changing processes and protocols to ensure the long-term digital preservation of the content on our platform, and required adding new terms to our local load agreements (i.e. the ability to migrate content to different formats).⁸

'Continuous access by users validates the archive immediately'

Unlike other popular preservation services for scholarly journals, SP's TDR is a light archive, emphasizing long-term access from the point of content loading. Continuous access by users validates the archive immediately, rather than at some later date when issues may be more difficult to fix.⁹

The screenshot shows the 'Preservation Metadata' section of an article page on the SP Journals platform. The article is 'Advanced glycation end products in skin ageing and photoageing: what are the implications for epidermal function?' by Mark D. Farrar, published in December 2016. The page includes a 'Preservation Status' button, a 'Report a problem' button, and a 'Preservation Metadata' section. This section contains an 'Event Log' and a 'PDF Details' table.

Event Type	Event Date	Event Detail
Ingest	2016-12-10	

Property	Value
Identifier :	/U9U66/U6/v25/UU12/94/_agepistatf/pdf_fulltext/148136b6200//
Preservation Level :	bit-level
Format :	Portable Document Format v.1.6
Format Registry Link :	http://www.nationalarchives.gov.uk/pronorm/fmt/20
Size :	201207
MDS :	711768cada702fb6b501a7ebaca8b25b
SHA 1 :	41012844c0e51bd12f96c6431b66ab2c1f8913a0
CRC32 :	1627537372

Event Type	Event Date	Event Detail
Fixity Check - CRC32	2018-08-21	Pass
Fixity Check - CRC32	2017-09-30	Pass
well-formedness check	2016-12-10	Pass
validity check	2016-12-10	Pass

Figure 3. Preservation metadata displayed on the SP Journals platform

Our preservation practices are fully transparent. TDR documentation is available on the Scholars Portal wiki, and the preservation status of each individual article can be viewed from its details page.

COUNTER compliance

Understanding usage is important both for subscribing libraries and for the publishers who own the content. In 2010 SP created the Scholars Portal Usage Data (SPUD) database. Like SP Journals, SPUD is hosted on MarkLogic. Custom scripts take the usage logs from SP Journals and SP Books and convert them into a variety of different usage reports.

In 2011 the Journal Report 1 (JR1) on SPUD, measuring successful full-text article requests, was audited and certified compliant with Project COUNTER Release 3 standard. We have since been compliant with Release 4 and, as of this writing, are working on a COUNTER Release 5 compliant TR_J1 report.¹⁰

'staff provide reports to publishers on the usage of their content'

OCUL member libraries have self-serve access to their usage reports and can log into SPUD to view or download usage for the whole platform or by publisher. SP staff provide reports to publishers on the usage of their content on an annual, semi-annual, or as-needed basis, depending on arrangements with the publisher.

Benefits and challenges

There are many benefits to this shared, locally hosted journals platform, including:

- **Built by libraries, for libraries**
SP Journals is governed by OCUL and developed from a librarian perspective, with the teaching, learning and research workflows of students, instructors and researchers in mind. We are able to be easily responsive to the needs of our users when developing new features or integrating with new applications.¹¹
- **Consistent experience for users**
With 55 million articles on a single platform, students and researchers at our member libraries can access many, if not most, of the scholarly articles they need on SP Journals. This reduces the confusion of different layouts, designs, options and vagaries on various publisher interfaces.
- **Assured access**
At the termination of an agreement, or in the case of a publisher bankruptcy or other failure, our member libraries maintain seamless access to the content to which they have post-termination rights.
- **Long-term preservation**
Our commitment to long-term, large-scale preservation ensures that the articles on our platform will remain reliably valid and accessible into the future, maintaining the scholarly record for future generations.
- **Large corpus for text and data mining (TDM)**
We have accumulated a large corpus of textual content across publishers and subject areas. SP Journals will be a valuable resource for TDM as we move into the next phase of digital research.
- **Canadian governance**
With our servers and our operational team on Canadian soil, our members can be assured that we follow Canadian and Ontario provincial legislation in terms of user privacy, web accessibility requirements, and more.
- **Fully bilingual**
The platform and all functionalities are available in French for our Franco-Ontarian and other French speaking students and researchers. Several of our member universities offer French-language programmes aimed at francophone students.

- **Scalability**

We ingest new content regularly, and intend to do so into the future. Therefore, we make sure that the infrastructure we develop can support large volumes of data and can be easily extended when necessary.

The major challenges are:

- **Reliance on publisher data**

Since we do not create content ourselves, we rely on timely, complete and accurate data from publishers. When an issue is missing, an incorrect PDF is sent, or a publisher is simply late, we do not have that content on our platform.

- **Complicated entitlements**

As we continue to move away from the big deal model of content licensing, both in terms of cancellations (and their subsequent perpetual access implications) and in terms of more granular purchase models, the management of entitlements is increasingly complex. Title transfers also complicate this process.

- **Extensive server and systems requirements**

Between vast amounts of data requiring storage, services to support access and usage, and security risks presented by content harvesters and others, running a platform like SP Journals requires long-term investment in hardware, software and the people and expertise to maintain them.

What next?

Ontario and beyond

Since so much of the licensing occurs at the national level with CRKN, there has been some interest in expanding the reach of SP Journals outside Ontario. The SP Books platform has already started offering access to non-OCUL participants in CRKN purchases, including the Canadian University Presses collection and the Canadian Electronic Library collection, and other Scholars Portal services have successful collaborations with universities across Canada.

Technological innovation

SP continues to innovate around scholarly resources. A new project focuses on implementing linked data (LD) technologies on the SP Journals platform. Modelling our entitlements as LD would make them easier to manage, while LD metadata could be enriched by bringing together information about authors' institutional affiliation, research funding, and more. LD also has the potential to create linkages across our content platforms. For example, a book on the SP Books platform could link directly to a book review article hosted on the SP Journals platform. We see LD as a powerful tool for breaking publisher silos as well as content-type silos.

'We see LD as a powerful tool for breaking publisher silos as well as content-type silos'

Conclusion

Maintaining a locally hosted journal archive is a complex endeavor, bringing together the work of many individuals at SP, OCUL and member libraries. SP Journals is the product of shared infrastructure and long-term investment encompassing content licensing, metadata transformation, data storage, network security, digital preservation, user experience, usage reporting, and more. This collaborative effort provides half a million students, staff and faculty members at Ontario universities with easy, consistent and perpetual access to a large and valuable pool of scholarly resources.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: <http://www.uksg.org/publications#aa>

Competing interests

The authors have declared no competing interests.

References

1. "FAQ," Directory of Open Access Journals: <https://doaj.org/faq#seal> (accessed 18 January 2019).
2. "Journal Archiving and Interchange Tag Suite," NCBI: <https://jats.nlm.nih.gov/archiving/> (accessed 18 January 2019).
3. Wei Zhao, Jayanthi Chengan, and Agnes Bai, "Quality Control Practice for Scholars Portal, an XML-based E-journals Repository" (paper presented at International Symposium on Quality Assurance and Quality Control in XML, Montréal, Canada, August 6, 2012); DOI: <https://doi.org/10.4242/BalisageVol9.Zhao01> (accessed 18 January 2019).
4. Theodore C. Bergstrom, Paul N. Courant, R. Preston McAfee, and Michael A. Williams, "Evaluating big deal journal bundles," *Proceedings of the National Academy of Sciences* 111 (2014): 9425–9430; DOI: <https://doi.org/10.1073/pnas.1403006111> (accessed 18 January 2019).
5. Dana Thomas, Catherine Davidson, Martha Kyriallidou, and Terry Plum, "Measuring use of licensed electronic resources: a second iteration of the MINES for Libraries survey on Scholars Portal and other resources for the Ontario Council of University Libraries," *Library Management* 33 no. 6/7 (2012): 374–388; DOI: <https://doi.org/10.1108/01435121211266195> (accessed 18 January 2019).
6. "Federated Identity Management Service Providers," CANARIE: <https://www.canarie.ca/identity/service-providers/> (accessed 18 January 2019).
7. Wei Zhao, Shuzhen Zhao, and Katie-Scarlett MacGillivray, "Providing and Maintaining Access to Electronic Serials: Consortium and Member University Library's Perspectives," *The Serials Librarian* 72 no. 1–4 (2017): 144–151; DOI: <https://doi.org/10.1080/0361526X.2017.1309831> (accessed 18 January 2019).
8. Scholars Portal, "Trusted Digital Repository Documents," SPOT Docs: <https://spotdocs.scholarsportal.info/display/OAIS/Home> (accessed 18 January 2019).
9. "FAQs," Trustworthy Digital Preservation at Scholars Portal: <https://tdr.scholarsportal.info/> (accessed 18 January 2019).
10. "Scholars Portal," Project COUNTER: <https://www.projectcounter.org/counter-user/scholars-portal/> (accessed 18 January 2019).
11. Mary Curran, "A conversation with Leslie Weir," *The Serials Librarian* 55 no. 4 (2008): 577–597; DOI: <https://doi.org/10.1080/03615260802371063> (accessed 18 January 2019).

Article copyright: © 2019 Sabina Pagotto and Wei Zhao. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author:

Sabina Pagotto

Client Services & Assessment Librarian

Scholars Portal, Ontario Council of University Libraries

7th floor, 130 St. George Street

Toronto ON M5S 1A5, CA

E-mail: sabina@scholarsportal.info

ORCID ID: <http://orcid.org/0000-0001-9130-092X>

Co-author:

Wei Zhao

ORCID ID: <http://orcid.org/0000-0001-5241-1475>

To cite this article:

Pagotto S and Zhao W, Access, preservation and analysis in a consortial journal archive: the evolution of Scholars Portal Journals, *Insights*, 2019, 32: 7, 1–10; DOI: <https://doi.org/10.1629/uksg.458>

Submitted on 20 December 2018

Accepted on 22 January 2019

Published on 21 February 2019

Published by UKSG in association with Ubiquity Press.