Open Infrastructure and the Threat of "Vanishing" Journals: Leveraging Open Knowledge Commons, Open Source Software, and DIY Solutions to Preserve Humanities and Social Sciences Research

GRAHAM JENSEN, SAJIB GHOSH, ARCHIE TO, AND RAY SIEMENS

Abstract: Academic journals, institutional repositories, and emerging digital technologies have played a crucial role in providing access to scholarship. However, free and unfettered access to research is not a given—nor are the digital infrastructures through which open research is published and made accessible immune to commercial enclosure or obsolescence. The threat of "vanishing" digital publications also remains a real threat, and open access and humanities and social science (HSS) journals are particularly at risk of disappearing. In this article, we aim to address the related issues of access to, and preservation of, HSS research by examining our own experiments with open methods and tools for the (re)publication of open access scholarship via open infrastructure. As part of this process of self-examination, we focus on one infrastructural initiative that is equipped to support this work: the Canadian-based HSS Commons.

In the process, we also invite consideration of how low-budget, DIY-style innovation and experimentation in the realm of digital research software constitute valid, crucial forms of humanistic intervention and activity. To do so, we discuss a project that emerged from the HSS Commons' collaborative partnership with Iter Canada: a large-scale migration of open access back issues from scholarly journals or book series operated by Iter.

In conclusion, we reflect on the larger significance, potential wider application, and limitations of such interventions. Indeed, while there are many possible benefits to the workflow we developed—which resulted in the publication of over 6,000 publications in the HSS Commons repository and which we hope will serve as a model for other groups or journals interested in backing up and increasing the discoverability of their own research—our work on this project also highlighted the many methodological, infrastructural, and institutional challenges that still face those who may be interested in pursuing open scholarship of this kind.

https://doi.org/10.3998/jep.7860

Academic journals, institutional repositories, and emerging digital technologies have played a crucial role in providing new and previously unimaginable forms of access to scholarship. However, free and unfettered access to research is not a given—nor are the digital infrastructures through which open research is published and made accessible immune to commercial enclosure or obsolescence. As several critics have noted in recent years, the threat of "vanishing" digital publications also remains a very real threat (Lightfoot 2016; Laakso et al. 2021; Eve 2024). What is more, open access and humanities and social sciences journals are particularly at risk of disappearing (Laakso et al. 2021).

In this article, we aim to address the related issues of access to, and preservation of, humanities and social sciences research by reflecting on our experiments with open methods and tools for the (re)publication of open access scholarship via open infrastructure. As part of this process of self-examination, we focus on one of our own infrastructural initiatives that is equipped to support this work: the Canadian-based Humanities and Social Sciences (HSS) Commons.

The HSS Commons (https://hsscommons.ca) is an in-development not-for-profit and multilingual community space for academics, research partners and stakeholders, students, and interested members of the public.[1] Hosted on Canadian servers but open to members from around the world, the HSS Commons is an initiative led by the Implementing New Knowledge Environments (INKE) Partnership (https://inke.ca) and its partners.[2] It is based on HUBzero, an open source platform originally developed at Purdue University and now managed at the University of California San Diego. Serving as a hub for open social scholarship, it combines elements of social networking sites, tools for collaboration, and institutional repositories, allowing researchers to freely share, access, re-purpose, and develop scholarly projects, publications, preprints, educational resources, data, and tools.

After introducing the issue of vanishing or vulnerable HSS research and the other theoretical contexts that motivate this work, we will examine how—by actively republishing and increasing the visibility of resources via open repositories—digital knowledge commons such as the HSS Commons can shape the academic landscape in material ways. We will also invite consideration of how free or low-budget DIY (do-it-yourself) innovation and experimentation in the realm of digital research software constitute valid, crucial forms of humanistic intervention and activity.

To do so, we will discuss a project that emerged from the HSS Commons' collaborative partnership with Iter Canada (https://www.itergateway.org): a large-scale

1. For more information about the HSS Commons, see, for example, Jensen et al. (2022), Tracy and Jensen (2025), and Winter et al. (2020).
2. The INKE Partnership "is a North American-based research network with the goal of fostering open social scholarship: academic practice that enables the creation, sharing, and engagement of open research by specialists and non-specialists in accessible and significant ways" (https://inke.ca).

migration of open access back issues from scholarly journals or book series operated by Iter Gateway, including *Renaissance and Reformation*, *Confraternitas*, *Early Modern Digital Review*, *Quaderni d'Italianistica*, and others.[3] The article's detailed case study of our collaboration with Iter describes how we used OpenRefine (https://openrefine.org), an open source program for cleaning and transforming data, to:

1. scrape and clean publication metadata;
2. transform the metadata using the General Refine Expression Language (GREL) and XML;
3. batch download publication PDFs using a custom Python script; and
4. import the publications and their metadata into the HSS Commons' repository.

This process and our case study's discussion of it both employ mixed methods, combining a practice-based and experimental methodology informed by environmental scans in the fields of critical infrastructure and platform studies, archival and library information studies, and open scholarship. Building on the preceding overview and analysis sections, our Iter case study is offered as an example of how, as in STEM fields and discussions of open science, "open research" in HSS can be both theoretical and practical in nature.

In conclusion, we reflect on the larger significance, potential wider application, and limitations of such interventions. Indeed, while there are many possible benefits to the workflow we developed—which resulted in the publication of over 6,000 journal articles and reviews in the HSS Commons repository via the "Iter Community" group and its sub-projects (https://hsscommons.ca/groups/itercommunity) and which we hope will serve as a model for other groups or journals interested in backing up and increasing the discoverability of their own research—our work on this project also highlights the many methodological, infrastructural, and institutional challenges that still face those who may be interested in pursuing open scholarship of this kind.

Relevant Contexts and Motivating Factors

Many theoretical and practical considerations underpin our work with Iter Canada. However, it would be misleading to suggest that all of these considerations motivated

3. As noted on their website, "Iter Canada facilitates the study and teaching of the Middle Ages and Renaissance (400 to 1700) through the investigation, promotion and use of digital practices in collaborative, community-based environments. We engage individuals, informal groups and organizations within Canada and beyond in order to share best practices and to develop resources. Our initiatives and projects are offered to the public through Iter Community" (Iter Gateway, n.d.).

our efforts from the start. In truth, the process of attending to the practical demands of our relatively small-scale, boutique preservation project often forced us to step back and reflect on theoretical implications beyond those identified at the outset.

Having carried out this work, we can now identify at least three pressing issues identified by scholarly communication critics, librarians, and other information professionals to which our project—even in its preliminary, experimental stage—responds in concrete, generative ways. These interrelated issues include (1) the threat of vanishing HSS resources, despite the valuable role played by preservation schemes; (2) the importance of improving both the quantity and discoverability of open access HSS research via centralized subject repositories; and (3) the need for free, open, and DIY solutions as part of larger, concerted responses to vanishing HSS resources or to any number of contemporary crises posing a threat to the preservation of research and cultural heritage materials.

The Threat of "Vanishing" HSS Resources

The first issue we aim to counter is the threat of "vanishing" or "disappearing" HSS resources, particularly those that are open access. This threat persists despite the important work that is being done by preservation networks or schemes, as discussed below. Additionally, though, this threat speaks to—and is an example of—the more general issue of the preservation of knowledge over time and across media. How might researchers leverage digital research infrastructure and open source software to better preserve the knowledge they produce? And why is this work of preservation especially imperative when it comes to open access research in the humanities and social sciences?

Although preservation in the broader sense is not the subject of the current study, it is worth nothing what this term entails in our own (academic) context: as Martin Paul Eve suggests, "Preservation activities cover everything from running server infrastructures to storing extra copies to ensuring that material is sent to archives at the time of publication" (2024, 4). Of note here, both digital infrastructure and the storage of extra copies of research materials by these infrastructures are important elements of what the Digital Preservation Coalition defines as "long-term preservation" (quoted in Eve 2024, 4). And while the objectives and means of preservation and archiving may overlap—frequently enough that the two terms are sometimes used interchangeably—we understand preservation to mean a wide range of activities (e.g., creating backups, migrating materials to new storage sites) that may or may not lead to archiving. Both are concerned with safeguarding physical or digital materials, but we use "preservation" to refer to activities geared towards safeguarding materials while also keeping them accessible for active research in the present; by contrast, we use "archiving" to refer to

activities that are comparatively less interested in providing ready access but more interested in long-term preservation and accurately preserving the original contexts and/or physical media associated with an object (e.g., through detailed description, recording provenance).

Admittedly, there may not always be a clear line delineating preservation and archiving. Our provisional definitions here are meant primarily to distinguish our own activities—which were focused on duplicating digital research materials while providing ready access to them—from those of trained archivists better able to preserve these materials "in perpetuity" and in line with exacting cataloging rules or archival description standards. However, because *digital* preservation aims to provide easy, sustainable access to data, it can also be exacting, and have a high bar for participation, in the sense that it must grapple with challenges posed by variables such as media degradation and differences in data formats and operating systems (Johnston 2020). Additionally, and more relevant in the case of our own preservation project, studies by Katrina Fenlon et al. (2025) and Nick Ruest et al. (2021) discuss how digital preservation often requires collaboration with user communities or collaborators to figure out their data needs and build reciprocal relationships to support sustainability.

For its part, as Canadian digital research infrastructure, the HSS Commons aims to participate in long-term preservation for open access resources by storing additional copies of the resources, whether or not they are already backed up elsewhere. And there are obvious, well-documented benefits—including discoverability, community building, among many others—to preserving or indexing materials in multiple places, beyond just the benefits we had in mind for our project: to build community around specific sets of scholarly resources and to promote serendipitous discovery of open access research in the HSS Commons environment. Indeed, the importance of preserving research by backing it up in multiple places is the driving idea behind so-called preservation systems or networks such as LOCKSS ("Lots of Copies Keep Stuff Safe").

Preservation schemes or systems such as LOCKSS (https://www.lockss.org), Controlled LOCKSS (CLOCKSS; https://clockss.org), Portico (https://www.portico.org), and the Public Knowledge Project (PKP) Preservation Network (https://pkp.sfu.ca/pkp-pn/) serve a vital role: they ensure that digital content is stored in duplicate and safeguarded in the event that, for example, a publisher folds; a natural disaster destroys or incapacitates a server or other digital infrastructure; data are compromised, lost, or intentionally destroyed as a result of war or geopolitical tensions; or files become inaccessible due to technological obsolescence.[4] Both LOCKSS and the Directory of Open Access Journals (DOAJ) recommend journals or other entities archive their content

4. For a helpful comparison of major preservation schemes, see Laakso et al. (2021, 1100ff.).

"in more than one place, ideally, in at least three" (DOAJ, n.d.; LOCKSS Program, n.d.). However, preservation networks do not necessarily address two concerns we flagged during the course of our own research into options for preserving Iter journals and other HSS materials: discoverability and possible costs.

For starters, most preservation networks do not address the issue of discoverability, since they actually operate as "dark archives" (i.e., archives that cannot be accessed by the public and are kept primarily as a backup solution in case the original archive becomes inaccessible or other "trigger events" occur).[5] One might reasonably ask, then, *How does one address preservation while also improving discoverability?* This is the question we asked when working with Iter, since our intention was not simply to preserve its journals but to do so in a way that would increase the visibility of, and prospective new readers' engagement with, the journals.

Moreover, there is the possible issue of costs associated with preservation activities. Some preservation networks (PNs) have membership requirements and fees, which can be a major obstacle for journals. To mitigate this concern, the DOAJ has reduced costs through Project JASPER, launched in 2020 (DOAJ, n.d.). And there are undoubtedly other ways to mitigate the larger problem of PN membership fees or costs associated with preservation efforts and infrastructures.

These possible limitations of PNs take us back to the issue of so-called vanishing or disappearing resources. Several critics rightly observe that the phenomenon of vanishing digital publications remains a very real threat (Lightfoot 2016; Laakso et al. 2021; Eve 2024). Eve (2024) reports, "A significant portion, approximately 28%, of academic journal articles with DOIs appear entirely unpreserved in the archives [. . .], endangering both persistent identifier systems and the chain of verifiable citation that they are meant to underwrite" (17). But this figure is for articles across all disciplines. HSS and open access journals are particularly at risk of disappearing: in "Open Is Not Forever: A Study of Vanished Open Access Journals," Mikael Laakso and co-authors remark that "social sciences and humanities (SSH) journals represent the largest share of vanished journals in our sample (52.3%)" (2021, 1106). Compounding the problem, "OA journals are enrolled in preservation schemes at an alarmingly low rate, with 4 in 10 journals indexed in the DOAJ reporting enrollment in at least one preservation or archiving scheme" (Laakso et al. 2021, 1100). Taken together, these numbers underscore that open access HSS research is disproportionately at risk of disappearance, occupying an uncomfortable position in a Venn diagram of academic journal vulnerability. The question of why HSS and open access journals are at particular risk of disappearance deserves further study, including in our own national context, but some of the

5. See, for example, https://pkp.sfu.ca/pkp-pn/.

likeliest explanations—such as the lack of significant, stable, or long-term funding for HSS, which we discuss below—are so well documented that they have almost become truisms. More helpfully, perhaps, one might also point to the technical expertise and time commitments involved in responsible, ethical, community-engaged preservation and archiving of digital research.

Journals hosted or published primarily through PKP's Open Journal Systems (OJS) software and other platforms that facilitate archiving through PNs are not immune, either—which is relevant in the case of our own project, since Iter's journals are hosted through the University of Toronto's Journal Production Services (JPS), an OJS instance.[6] Bronwen Sprout and Mark Jordan observe that "the [Global LOCKSS Network, or GLN] preserves content from around 200 OJS titles (out of approximately 10,000)"; this alarmingly low preservation rate for OJS journals is due to the GLN's prioritization of other content and due to the small size (and presumably limited resources) of many OJS journals (2018, 247–48). Even the PKP PN, which was created to fill this gap for OJS journals, does not provide a foolproof solution; it is up to individual OJS journals to enable an optional LOCKSS plugin to have content backed up in this way.

The fact that individual OJS journals must enable a plugin to be enrolled in a preservation scheme is a small example, but it helps prove a larger point made by many critics: that "there is no consensus over who should be responsible for archiving scholarship in the digital age" (Rick Anderson, quoted in Eve 2024, 4). This lack of consensus has contributed significantly to the problem of vanishing open access resources. Elaborating on this dilemma, Laakso et al. (2021) situate it in relationship to other issues affecting open access journals:

> Although the necessary infrastructure exists, at least to some extent, questions as to what content to preserve and who should be responsible for its preservation remain unresolved. Current practices for selecting content for preservation can disadvantage OA content since aspects such as journal impact factors or the invested cost for content acquisition often drive such decisions. Especially in the case of small and independent OA journals, which face financial and technical barriers to preservation arrangements, it seems that the opposite approach for content selection is needed—one that also includes the most vulnerable journals instead of prioritizing prestige. Indeed, preserving the "long tail" of scholarly literature might be one of the most pressing challenges the scholarly community is facing. (Laakso et al. 2021, 1102)

Yet this "most pressing" concern is amplified further by threats to the long-term survival of preservation networks themselves. Rather ironically, even preservation networks

6. See University of Toronto Libraries (n.d.).

fail or are in danger of shutting down (DOAJ 2018; Laakso et al. 2021, 1109). Clearly, more must be done to preserve not only those resources most at risk of vanishing but also the networks and infrastructures designed to preserve them.

The Need to Improve the Quantity and Discoverability of Open Access HSS Research

Our project also aims to make more open access HSS research discoverable online, specifically through a centralized subject repository. Several popular, discipline-agnostic digital commons host a significant amount of HSS research as well as research from other disciplines, but, as the example of Academia.edu attests, they may do so without regard for robust or standardized metadata (Dingemanse 2016). Many digital research commons of this kind are also "walled gardens": closed, for-profit spaces that exploit user data and operate in a manner largely antithetical to the spirit of the Open Access Movement, in spite of official statements to the contrary.[7] In these cases, discoverability of HSS research, rather than quantity, is sometimes the real issue, since their bespoke metadata is not concerned with interoperability or with conformance to other aspects of the FAIR (Findable, Accessible, Interoperable, Reusable) guidelines for data management.

Of course, a considerable amount of HSS research is also hosted in institutional repositories (IRs), which serve vital preservation, archival, and research-sharing functions. However, some of these distributed IRs are closed to those without institutional affiliation, and, as Brian Clark notes, IRs have been plagued by other issues, including a lack of faculty and student buy-in. Clark writes, "reasons [for this lack of buy-in] range from not knowing the IR exists in the first place, not understanding its purpose or benefits, concerns over copyright and plagiarism, or workflows that require too much time and input from faculty" (2023, 743). For the project outlined in this article, we chose to adopt a more active deposit model to populate the HSS Commons, which is built around a subject repository rather than an institutional one. That is, the deposit model we used is not passive, but active—and also community-engaged, following what Clark calls "mediated deposit models" (743). Initially, though, we had tried other active or hybrid approaches, including inviting researchers in our larger Implementing New Knowledge Environments (INKE) Partnership to add their publications to the HSS Commons repository, offering hands-on assistance from graduate research assistants who were part of the HSS Commons team. While this approach worked, insofar as it

7. See, for example, Winter et al. (2020) and Jensen (2023a, 2023b), which synthesize scholarship critical of Academia.edu and other for-profit knowledge commons.

helped us populate the site with a modest number of new publications, it also took a considerable amount of time and effort—largely because it involved manually copying publication metadata from CVs, personal websites, or existing repositories into corresponding form fields in the front end of the HSS Commons. For these reasons, we began to search for an alternate, semi-automated active deposit model, knowing that well-populated repositories tend to encourage further contributions and thus increase the repository's visibility and potentially prestige (Schlangen 2015; Hwang et al. 2020; Butterfield et al. 2022; Clark 2023).

The Need for Free, Open, and Sometimes DIY Solutions

The third and final issue we identified is that, when funding for publication or long-term preservation is not readily accessible, free, open, and sometimes makeshift or idiosyncratic solutions may be required. This is a known issue facing small journals and libraries (Laakso et al. 2021, 1108), and it constrains the work of publishers and researchers in many parts of the world—even as it drives others to create innovative, equitable, and community-minded publication tools or workflows.

The need for free, open, or DIY solutions as a response to such constraints, or as a response to the issues detailed above, is not just—or not necessarily—a Global North versus Global South issue; it is felt across national, linguistic, and disciplinary boundaries. As Laakso et al. note, for instance, there is a "disproportionately low share of vanished journals from Latin America—where the principles of community and OA are embedded into academic culture" (2021, 1108). In Canada, as elsewhere, funding within and for the humanities and social sciences has been an ongoing concern: despite the promise of open access publishing, barriers to or misalignments in open access policy (CRKN 2025) and cuts to scholarly publishing initiatives or the academic libraries and funding agencies that support such initiatives have been a constant threat, creating or exacerbating existing issues related to the production, accessibility, and preservation of knowledge (Canadian Federation for the Humanities and Social Sciences 2023; CRKN, n.d.).

Still, open access is widely supported across academic disciplines in Canada, and many scholarly communities are working to champion infrastructures and institutional pathways to support open access publication. As the Canadian Research Knowledge Network (CRKN) notes, "Canada has a strong diamond open access ecosystem, particularly in the humanities and social sciences" (2025). In a similarly hopeful vein, Simon van Bellen and Lucía Céspedes observe that, of the 944 active Canadian scholarly journals they identified, only 6% are controlled by "the six major commercial publishers globally, which include RELX-Elsevier, Springer Nature, Wiley, MDPI, Taylor & Francis

and Sage Publishing" (2024, 7). While the tendency towards non-commercial publishers seems to bode well for open access publishing in Canada, there are other elements of Canadian academic culture and publishing undermining open access's widespread adoption—and in some cases these elements disincentivize open access publishing and the long-term preservation of scholarly data. In terms of journals, the CRKN (citing van Bellen and Céspedes) remarks that "the vast majority of Canadian journals operate independently of commercial publishers and have already embraced Diamond OA, that is OA without author-facing fees" (2025). However, even according to the calculations of van Bellen and Céspedes, only 61% of "national" journals are Diamond OA, and these figures do not account for the many "international" journals that—as their summary of Larivière and Warren (2019) acknowledges—continue to hold significant appeal to "Canadian scholars, especially in the SSH" (van Bellen and Céspedes 2024, 2). Moreover, van Bellen and Céspedes nod to the well-documented fragility and instability of not-for-profit journals: they identify "312 periodicals having ceased publication, equivalent to 25% of the total number of journals having been active during this period" (12), noting also that "achieving financial sustainability remains a challenge for many journals, particularly Diamond OA ones" (16). To add to these concerns, policy issues complicate and, in some instances, obstruct the advance of Canadian open access publishing—to say nothing of shortcomings in Canadian funding programs. Van Bellen and Céspedes point out, for example, that "a subscription journal currently stands a better chance of being funded than a Diamond OA journal" (16). Despite Canada's strengths in open access publishing, then, the fact remains that a successful open access model needs support at multiple levels—for example, culture, society, policy, funding, infrastructure—to remain viable as well as resistant to threats of obsolescence and disappearance once it has been implemented.

For the United States and its trading partners, various threats to academic publishing and activity have recently become a painful reality for many in HSS and beyond, as the recent tariff announcements and devastating cuts to the National Endowment for the Humanities both illustrate (Aktorosian 2025; Palmer 2025; Wulf 2025). Moreover, wars, climate disasters, and other violently disruptive events have led to creative solutions to preservation and archiving—such as those employed by Saving Ukrainian Cultural Heritage Online (SUCHO; https://www.sucho.org) and detailed in its members' *DIY Web Archiving* zine (Dombrowski et al., n.d.)—demonstrating how DIY preservation methods might be adapted as part of a preemptive response to any number of crises.

Beyond these more extreme and exemplary examples, though, more mundane factors such as interoperability also impact preservation efforts in vital ways at local, national, or even international levels. For example, Clara Turp et al. observe that in Canada, "There is a need for repositories to move beyond stand-alone systems by promoting interoperability. Systems are slowly evolving away from the stand-alone model,

whether through metadata aggregation services, discovery layers, or linked open data-based approaches" (2020, 1). Interoperability and the need for robust publication metadata linking distributed repositories were relevant in the case of our work with Iter too—work carried out as a research experiment, in which innovation was required less as a response to immediate geopolitical crises than to structural and infrastructural constraints.

But what does "innovation" mean in this context? As the next section elucidates, the workflow we developed involves the use of custom scripts that we wrote, the customization of open source software to extend its out-of-the-box functionality, and also scraping metadata rather than accessing it via application programming interfaces (APIs) due to technical as well as other limitations (such as those discussed below). It is innovative and DIY in part, though, because our team is not trained in preservation work of this kind and therefore was, at the outset, largely ignorant of existing workflows or methods used by other repositories; as it turns out, we were re-inventing a wheel, using whatever tools, materials, and digital skills were at hand. Still, while we later discovered that there are certainly more powerful and efficient ways of handling data migration and mapping than scraping, and that projects such as the SSH Open Marketplace (2025) have used more sophisticated methods to achieve similar metadata-ingestion ends, these often require much greater investment in personnel, infrastructure, and software. By comparison, our method—leveraging OpenRefine's easy-to-use graphical user interface rather than requiring more advanced programmatic methods—may be useful as a customizable DIY solution that nevertheless provides remarkable data cleaning affordances primarily through widely accessible open source software.

Case Study: Iter Journal Migration

Over the last year, the co-authors of this article—all based in the Electronic Textual Cultures Lab (ETCL; https://etcl.uvic.ca) at the University of Victoria—initiated what the four of us have internally been referring to as the "Iter migration project" in consultation with Iter and its founding director, William R. Bowen. Again, our mission was to republish open access back issues of some of Iter's academic journals via the HSS Commons; we took on this project with Iter for preservation purposes, driven by the kinds of pressures and concerns detailed above. But this work was driven equally by our team's research and research-prototyping interests in the areas of community building and open social scholarship. One of our additional goals, then, was to use the HSS Commons as a digital space for building community as well as a tool to support open, collaborative, and experimental research. Indeed, it was in this latter, experimental frame of mind that we first started exploring possibilities for preserving Iter journals.

To make our migration of these Iter journal materials more efficient, we developed a bulk ingestion workflow that scraped article metadata from the journals—all of which are hosted through the University of Toronto's Journal Production Services (JPS), which uses OJS. This workflow, which is described in detail in this section, allowed us to import the articles into the HSS Commons' repository without requiring JPS or the journal editors to provide us with API access (since some concerns were raised about the openness of the API). More than that, though, it forced us to build valuable human oversight into the process while still allowing us to control, customize, and easily reproduce the data cleaning process to account for small but consequential differences between the journals and their metadata-formatting practices.

1. Creating a Target Resource Spreadsheet

Working with Iter and others, we started by identifying all resources (at the journal issue or monograph level) that we wanted to migrate to the HSS Commons repository. We organized this information in a spreadsheet with journal titles, issue numbers or book titles, the URL for each issue on the JPS website, and the publication's Creative Commons license type in separate columns (figure 1). For project management purposes, we also included columns to record the ETCL team member assigned to the migration work for a particular resource as well as to indicate when that resource had been successfully migrated.

2. Creating an OpenRefine Project

For the next steps in our workflow, we used OpenRefine, "a powerful free, open source tool for working with messy data: cleaning it; transforming it from one format into

Journal / Series Title	Issue / Book title	Link	License	Total items	Migration assigned to	☑	Notes (incl. permissions info confirmed by Iter)
Confraternitas	Vol. 23 No. 1 (2012)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1410	CC BY-NC 4.0	6	Sajib	☑	
	Vol. 23 No. 2 (2012)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1434	CC BY-NC 4.0	4	Sajib	☑	
	Vol. 24 No. 1 (2013)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1463	CC BY-NC 4.0	5	Sajib	☑	
	Vol. 24 No. 2 (2013)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1504	CC BY-NC 4.0	6	Sajib	☑	
	Vol. 25 No. 1 (2014)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1570	CC BY-NC 4.0	6	Sajib	☑	
	Vol. 25 No. 2 (2014)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1633	CC BY-NC 4.0	9	Sajib	☑	
	Vol. 26 No. 1 (2015)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1731	CC BY-NC 4.0	5	Sajib	☑	
	Vol. 26 No. 2 (2015)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1796	CC BY-NC 4.0	9	Sajib	☑	
	Vol. 27 No. 1-2 (2016)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1879	CC BY-NC 4.0	9	Sajib	☑	
	Vol. 28 No. 1 (2017)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1901	CC BY-NC 4.0	7	Sajib	☑	
	Vol. 28 No. 2 (2017)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1952	CC BY-NC 4.0	10	Sajib	☑	
	Vol. 29 No. 1 (2018)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/1999	CC BY-NC 4.0	9	Sajib	☑	
	Vol. 29 No. 2 (2018)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/2185	CC BY-NC 4.0	8	Sajib	☑	
	Vol. 30 No. 1-2 (2019)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/2259	CC BY-NC 4.0	10	Sajib	☑	
	Vol. 31 No. 1 (2020)	https://jps.library.utoronto.ca/index.php/confrat/issue/view/2506	CC BY-NC 4.0	7	Sajib	☑	
Early Modern Digital Review	Vol. 1 No. 1 (2018)	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2471	CC BY 4.0	6	Sajib	☑	ALL issues open access (CC BY 4.0)
	Vol. 2 No. 1 (2019)	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2470	CC BY 4.0	5	Sajib	☑	
	Vol. 2 No. 2 (2019)	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2469	CC BY 4.0	6	Sajib	☑	
	Special Issue, Digital	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2468	CC BY 4.0	7	Sajib	☑	
	Special Issue, Womer	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2467	CC BY 4.0	7	Sajib	☑	
	Vol. 3 No. 1 (2020)	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2466	CC BY 4.0	10	Sajib	☑	
	Special Issue, Digital	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2465	CC BY 4.0	11	Sajib	☑	
	Vol. 3 No. 3 (2020)	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2464	CC BY 4.0	10	Sajib	☑	
	Special Issue, Digital	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2478	CC BY 4.0	19	Sajib	☑	
	Vol. 4 No. 1 (2021)	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2528	CC BY 4.0	10	Sajib	☑	
	Special Issue, Digital	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2537	CC BY 4.0	12	Sajib	☑	
	Special Issue, Spatial	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2567	CC BY 4.0	11	Sajib	☑	
	Special Issue, Compu	https://jps.library.utoronto.ca/index.php/emdr/issue/view/2598	CC BY 4.0	14	Sajib	☑	

Figure 1. Screenshot of target resource spreadsheet

another; and extending it with web services and external data." Of note for anyone interested in these kinds of tasks, OpenRefine has active user and development communities as well as excellent official documentation.

In OpenRefine, we first created a new project based on the journal issue URLs from the spreadsheet. We started with URLs, since these were needed to populate the project with metadata obtained through scraping. Since some journals had published material with multiple Creative Commons license types, and since we received journal- and issue-specific notes from Iter, we also decided to create separate OpenRefine projects, grouped according to issues within a journal that used the same Creative Commons license.

3. Scraping Relevant Metadata

OpenRefine does not function exclusively as a scraper for external websites, and many users adopt it primarily for other purposes, but we found that—thanks to its built-in fetching function and additional support for HTML parsing—it can be quite an effective tool for harvesting open data, even though APIs and other programmatic methods for collecting metadata were theoretically available to us (including within OpenRefine, which also works with APIs). Despite the availability of an "OJS Scraper for R" package (gastonbecerra, n.d.), for instance, we preferred OpenRefine so that we could scrape, clean, and export the metadata within a single graphical user interface. This OpenRefine-centric process also allowed us to easily customize the scraping to accommodate any differences specific to JPS or other OJS instances.

Each of the journal issues we migrated from JPS has its own landing page on that journal's JPS website, with additional landing pages for individual publications within each issue. Crucially for our purposes, then, both of these types of landing pages contain predictably structured HTML and metadata that can be scraped in an automated fashion. The predictable structure of the JPS system allowed OpenRefine to use our list of journal issue URLs (from step 2) to scrape the landing pages for each issue and then scrape the landing pages for each of the issues' individual publications to extract metadata.

Within OpenRefine, we accomplished this by scraping the source code for each journal issue's landing page by using the built-in "Add column by fetching URLs . . . " function. Next, we used a General Refine Expression Language (GREL) statement to parse the landing pages' HTML and identify all individual publications within each journal issue so that their metadata could be scraped and used to populate new columns using a series of targeted GREL statements.

The success of this process relies on the predictable structure of an individual publication's metadata within JPS. For example, if one examines the source code for Eva Pelayo Sañudo's "Multicultural Little Italy: A Literary Comparison of Canadian and US

Urban Enclaves" (published in volume 34 of the journal *Italian Canadiana* and accessible at https://jps.library.utoronto.ca/index.php/italiancan/article/view/37450), the metadata information contained in the HTML and scraped by OpenRefine is encoded in hidden meta elements such as the following:

```
<meta name="citation_journal_title" content="Italian
Canadiana"/>
<meta name="citation_journal_abbrev" content="Italcan"/>
<meta name="citation_issn" content="2564-2340"/>
<meta name="citation_author" content="Eva Pelayo Sañudo"/>
<meta name="citation_author_institution" content="University
of Oviedo, Spain"/>
<meta name="citation_title" content="Multicultural Little
Italy: A Literary Comparison of Canadian and US Urban
Enclaves"/>
```

Again, because these HTML elements represent the publication metadata in a predictable, structured, machine-readable way, we were able to use OpenRefine to parse the HTML, extract the relevant metadata, and create separate columns for each of the metadata values we needed for our project. To create a new column and populate it with the title for each publication, for example, we used the following GREL snippet:

```
value.parseHtml().select("meta[name=citation_title]")[0].htmlAttr
("content")
```

However, once we had extracted all of the metadata we required, we then needed to clean and transform the data to ensure consistency and prepare it to be exported to the HSS Commons. This step was the most time-consuming part of the entire process, and the nature of the data cleaning and transformation work was slightly different across the Iter journal datasets. Typically, though, this data cleaning and transformation included using filtering/faceting techniques,[8] GREL, or—in some cases, where the implementation of GREL or other programmatic solutions were prohibitively time-consuming—manual correction to perform the following kinds of tasks:

- Removing whitespaces from the beginning or end of various text-based fields (e.g., authors, publication titles, abstracts);
- Removing instances of multiple spaces within text strings;
- Excluding tables of contents, front matter, back matter, journal covers, and other journal materials we did not intend to republish;

8. In OpenRefine, faceting allows users to isolate and work with specific parts of a larger dataset without necessarily altering the larger dataset permanently. For more about faceting, and for examples, see Smiley (2024).

- Replacing "null" values (e.g., adding a journal title to a publication's list of subject tags, to increase a publication's discoverability and prevent issues in the case of required metadata information in the HSS Commons repository);
- Splitting comma- or semicolon-separated author lists, reversing the given name(s) and last name(s), and creating a semicolon-separated list to be used in citations;
- Prepending the titles of book reviews with "Review of" to distinguish reviews from other publications and prevent confusion regarding authorship;
- Creating a PDF URL column, by identifying JPS galley links, to be used in a later step to batch download all publications; and
- Creating citations of the original JPS publications so that these could be displayed to users in the HSS Commons with an accompanying note acknowledging their provenance and linking back to the journal (to promote the journal, increase discoverability, and provide clear information about the publication of record).

Initially, we performed all data transformations one by one during the raw data cleanup phase through an iterative process, as we learned more about GREL and learned how best to scrape, clean, and transform metadata. Later, though, we created a transformation template, which we could apply all at once using OpenRefine's extremely useful "Apply Operation History" feature (figure 2). This feature allowed us to perform

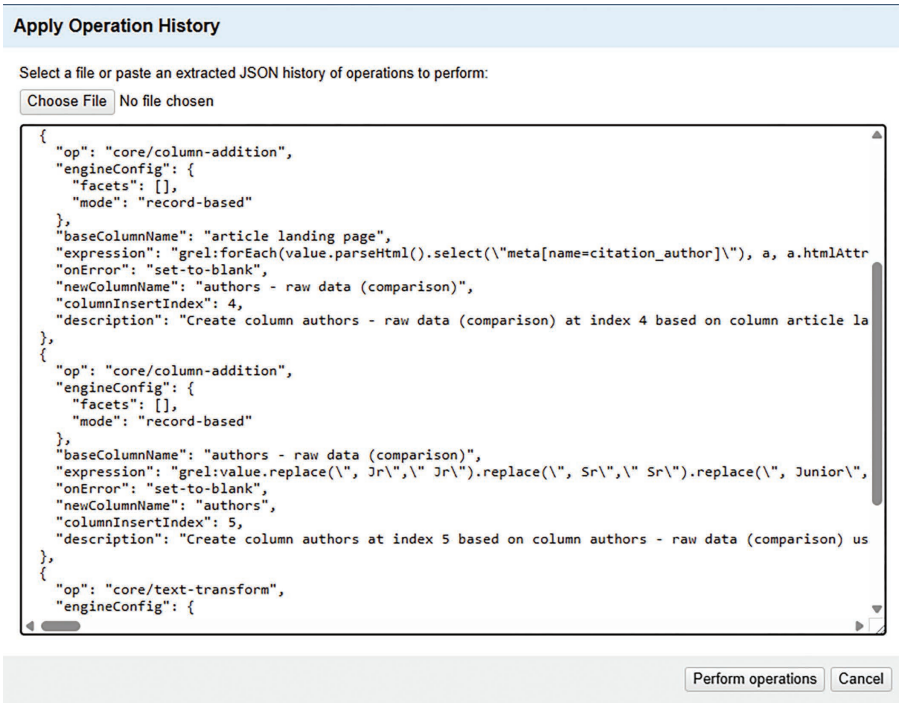

Figure 2. OpenRefine's "Apply Operation History" feature

all of the following actions, which OpenRefine encodes in JavaScript Object Notation (JSON) statements, in a single step: visit and scrape each issue's landing page, visit and scrape any publications' landing pages linked from main issue pages, scrape each publication's metadata, construct a list of publication PDF paths, clean all of the metadata, and create new metadata columns to assist with the export process.

4. Publishing Records on HSS Commons

Publishing the metadata from step 3 involved (1) creating an XML file containing information in a format compatible with HSS Commons' custom schema, (2) downloading the PDF files for each of the records, and (3) uploading and publishing both record metadata and PDFs in the HSS Commons.

To create an XML file containing the metadata for each of the publications, we harnessed OpenRefine's powerful "Templating export" feature (figure 3).[9] By doing this, we were able to export each record as part of an XML file that conformed to the Commons' required schema for batch ingestions. This process involved mapping metadata fields onto the XML schema as well as adding in a custom notes field to be displayed on the HSS Commons. The notes field acknowledges Iter and the journals in question

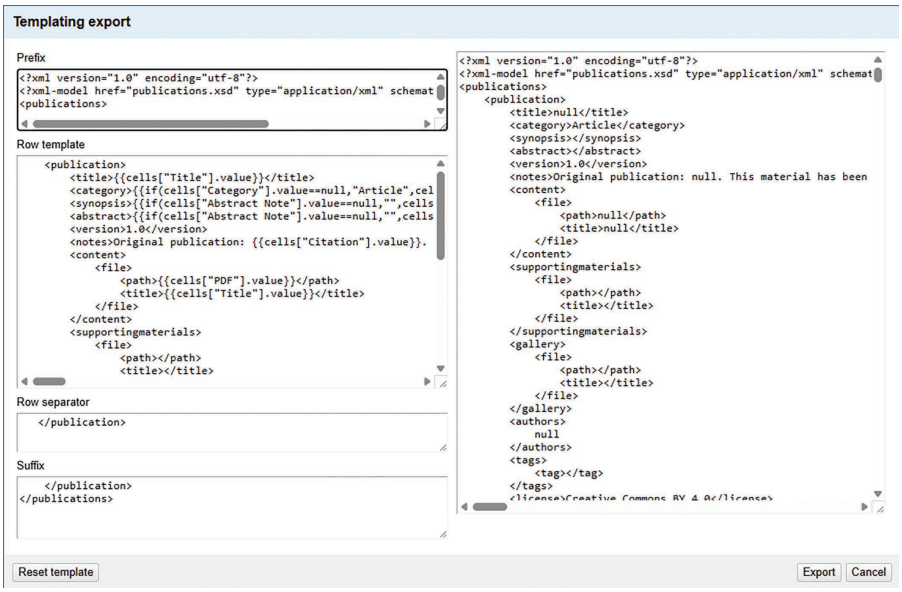

Figure 3. OpenRefine's "Templating export" feature

9. See, for example, Rajput (2024).

for permission to carry out this work and indicates the provenance and permissions for each publication.

Once the XML file was ready, we used a custom Python script, written in XML and GREL by our developer, Archie To, to batch download all of the publication PDFs and insert their file paths into a second version of the XML file.

To upload and publish the records on HSS Commons, we first created an HSS Commons "project" for each journal within the larger Iter Community "group" and logged in with the Iter user account (figure 4). We then uploaded the PDFs for that journal to its project-level repository. Next, we batch imported the metadata by uploading the XML file to the administrative back-end interface of the HSS Commons.

To improve this process in the back end, Archie customized the PHP code governing HUBzero's built-in "Batch Create" feature to automatically publish the records in the HSS Commons repository—a creative technical intervention that saved us from having to manually confirm and publish each of the more than 6,000 records, since the default batch-ingestion function saves items with a status of "Draft" rather than "Published" (figure 5). Additionally, this batch creation and publication process was unique insofar as—according to Clark (2023, 748)—many metadata-ingestion workflows of this kind do not also involve the harvesting and republication of publications'

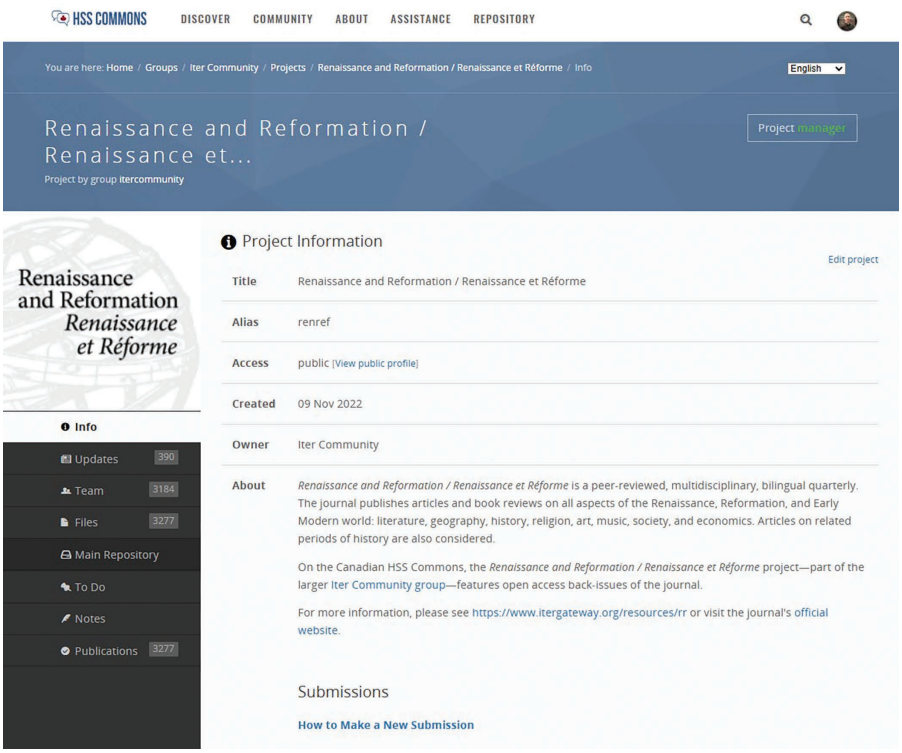

Figure 4. An Iter journal's "project" landing page in the HSS Commons

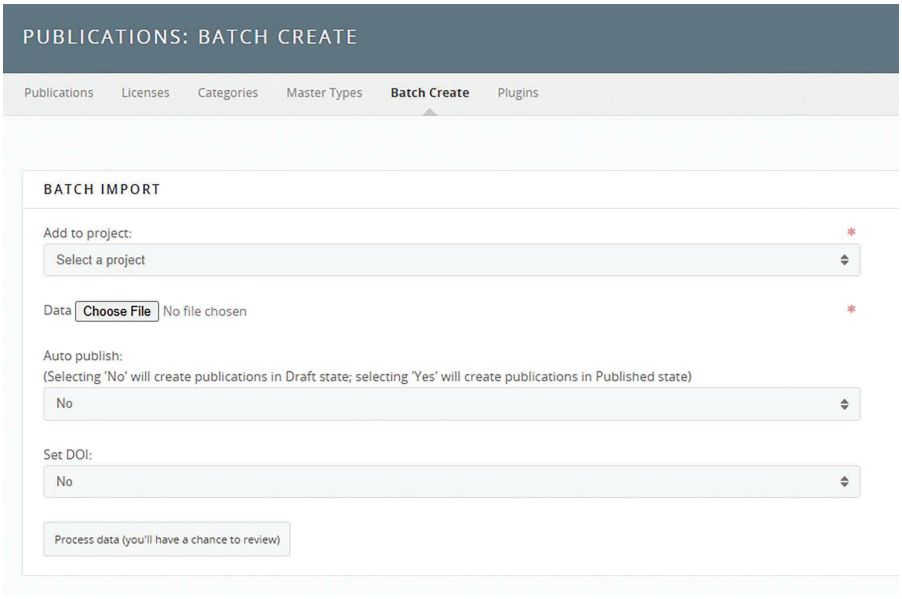

Figure 5. The customized "Batch Create" function in the HSS Commons back end

primary files. When records had been successfully created in the HSS Commons, they became publicly accessible immediately via the site's repository (figure 6).

Discussion: Limitations, Challenges, Theoretical Reflection, and Next Steps

We are incredibly grateful to Iter for its collaboration on this experiment in DIY preservation of open access humanities and social sciences research, which enabled us to migrate 6,100 journal articles and reviews from nine different Iter journals and book series to the HSS Commons. In addition, we developed a comprehensive workflow, which we have since used within our team for similar preservation efforts but are also considering publishing openly for others to adopt and customize. In this way, too, we consider the project a great success, even as we look for new and improved ways—more efficient, more in line with library best practices, and more inclusive in terms of linguistic, disciplinary, or geographical diversity, and so on—to carry out this kind of work in conversation with researchers, publishers, and infrastructural partners.

In equally important but less tangible ways, we also consider this project a success insofar as it constitutes a material expression of our own INKE Partnership's (n.d.) "Connection" Activity Cluster, whose stated goal is "strengthening the open

Figure 6. A public-facing publication landing page in the HSS Commons

communication of research through innovative interaction and engagement among humanities and social sciences researchers, organizations, and the public." The preservation work we did was a form of experimental knowledge mobilization designed to promote the discoverability of open research within and beyond the Iter network, but the repository through which these resources were republished also figured—and

continues to figure—in our thinking as a jumping-off point for new research and academic collaborations.

Limitations and Challenges

Our workflow was deliberately designed to use open source tools or software, to make it more accessible and potentially useful to others. Given our team's own inexperience at the outset of this project, we are optimistic, too, that it can be applied in a variety of open access publishing and open infrastructure contexts. That said, we did document a number of limitations and roadblocks that, in this same spirit of openness, we thought it worthwhile to pass on to others who may be interested in this kind of preservation work.

First, our workflow was only designed to scrape articles from OJS-based journals—though it is, admittedly, a very widely used platform. But, more to the point, our workflow is tailored for ingestion into only one platform, the HSS Commons, and the batch-publication function on our site uses a somewhat idiosyncratic XML schema. Adapting the custom export template for other schemas or file types would require anyone not already familiar with XML to learn some fundamentals (or perhaps use an AI-assisted programming tool).

Second, any customization of our workflow—particularly the web scraping and data cleaning steps—requires some knowledge of OpenRefine, GREL, GREL-supported regex, and the OpenRefine variables used both in GREL expressions and in the templating exporter. Data transformations can be performed using Python/Jython or Clojure instead of GREL, but GREL is the default option. And while excellent OpenRefine tutorials such as the *Programming Historian*'s "Fetching and Parsing Data from the Web with OpenRefine" (Williamson 2022) are freely available, certain cleaning and transformation scenarios may present a challenge to those with minimal GREL or programming experience. During our work with Iter, for example, we struggled to transform a concatenated author list (e.g., "Graham Jensen; Sajib Ghosh") into a list in which the authors' names were inverted (e.g., "Jensen, Graham; Ghosh, Sajib"). After some research, we arrived at the following interim solution using GREL:

```
forEach(value.split("; "), v,
v.match(/(.*)\s(.*)/).reverse().join(", ")).join("; ")
```

However, this transformation was not able to properly handle names with suffixes, such as "Jr." or "III." When doing similar preservation work later, for other open access journals, and after many more failed attempts to transform names such as "Martin Luther King Jr." to "King, Martin Luther, Jr."—as one might expect to see in an

academic paper's references section—we finally decided to return to the problem with the help of Chat GPT-4o. It came up with a solution that, while extremely effective in dealing with the suffixes "Jr.," "Sr.," "Junior," "Senior," and Roman numerals II through XX, we may never have been able to produce on our own:

```
forEach(value.split("; "), v,
   if(
       or(
           v.split("")[v.split("").length() - 1] == "Jr",
           v.split("")[v.split("").length() - 1] == "Sr",
           v.split("")[v.split("").length() - 1] == "Junior",
           v.split("")[v.split("").length() - 1] == "Senior",
           v.split("")[v.split("").length() - 1] == "II",
           v.split("")[v.split("").length() - 1] == "III",
           v.split("")[v.split("").length() - 1] == "IV",
           v.split("")[v.split("").length() - 1] == "V",
           v.split("")[v.split("").length() - 1] == "VI",
           v.split("")[v.split("").length() - 1] == "VII",
           v.split("")[v.split("").length() - 1] == "VIII",
           v.split("")[v.split("").length() - 1] == "IX",
           v.split("")[v.split("").length() - 1] == "X",
           v.split("")[v.split("").length() - 1] == "XI",
           v.split("")[v.split("").length() - 1] == "XII",
           v.split("")[v.split("").length() - 1] == "XIII",
           v.split("")[v.split("").length() - 1] == "XIV",
           v.split("")[v.split("").length() - 1] == "XV",
           v.split("")[v.split("").length() - 1] == "XVI",
           v.split("")[v.split("").length() - 1] == "XVII",
           v.split("")[v.split("").length() - 1] == "XVIII",
           v.split("")[v.split("").length() - 1] == "XIX",
           v.split("")[v.split("").length() - 1] == "XX"
       ),
        v.split("")[v.split("").length() - 2] + ", "+
        join(slice(v.split(""), 0, v.split("").length() - 2),
"") + ", "+
        v.split("")[v.split("").length() - 1],
        v.split("")[v.split("").length() - 1] + ", "+
        join(slice(v.split(""), 0, v.split("").length() - 1),
"")
   )
).join("; ")
```

This seemingly niche example demonstrates not only the potentially surprising complexity of any project dealing with metadata transformation challenges but also the power of GREL and AI-assisted programming tools to come up with DIY solutions to them. Although our experiment also sparked discussion about the ethics and financial,

environmental, or other costs of generative AI (which are well beyond the scope of this article), we know that many information professionals and developers carrying out preservation work of this kind are likely much more well-equipped than we were to deal with such roadblocks without the use of third-party services or tools.

Similarly, we are fully aware that, in some cases, collecting metadata via API instead of scraping may be more efficient, more accurate, or—particularly in the case of much larger scraping projects—the option officially recommended by repositories and their system administrators.

Third, and finally, because we were dealing with resources in English, we did not consider the challenges posed by multilingualism vis-à-vis data repositories. Turp et al. (2020) have summarized some of the challenges involved when attempting to map metadata from one repository onto another, or to reconcile controlled vocabularies between different languages. Drawing on their work on the Canadian Federated Research Data Repository (FRDR), they point out that "[u]sers searching in French, for example, might be interested in searching across all the datasets, regardless of which language is selected in the interface" (Turp et al. 2020, 6). In our case, such considerations were beyond the scope of our project, and we have not yet created bilingual—let alone multilingual—crosswalks.

Further Theoretical Considerations and Next Steps

Returning to the theoretical contexts for this project, I would suggest that, despite the limitations of our approach, we accomplished what we set out to do—and we did so in line with our lab's emphasis on openness and community. Still, the work of preserving HSS as well as open access research must go on. As Laakso et al. note, "[T]his issue [of vanishing OA journals] should be considered as an ongoing process that will continue unless we fully commit to preserving the scholarly record. Successfully solving this issue will require the active involvement of the scholarly community as a whole and solutions as diverse as scholarly research itself. While the current system places the responsibility for preservation mainly on OA journals alone, other actors (e.g., funders, academic institutions, authors) play a vital role in facilitating this process and in mitigating losses" (2021, 1108).

Our efforts to help Iter preserve its journals focused around JPS, an OJS instance that provides optional access to the PKP Preservation Network via a plugin. But our workflow could be adapted quite easily, in the future, to migrate other journals not enrolled in preservation schemes and therefore at even higher risk of disappearance. Tom Cramer et al. echo some of these sentiments; they concur that "digital preservation can never be a solved problem. It is work that does not finish" (2023, 312). Maybe

the point, then, is that we are trying, and we are using open infrastructure to creatively address a shared problem that is everyone's and no one's responsibility.

Moving forward, we intend to migrate other journals to the HSS Commons as we improve both our workflow and the repository's compliance with reusability, interoperability, and other FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Indeed, reproducibility is one of the key benefits of an OpenRefine-centered workflow, as Turp et al. (2020, 4) have also noted. Taking advantage of the "Apply Operation History" feature described above, we have already reused and refined our Iter workflow to preserve other open access journals such as *Canadian Food Studies*.

Other next steps might be to create a similar workflow that uses OJS or OpenAlex API to retrieve open access journal metadata from within the OpenRefine environment; use the Python-based Sickle (n.d.) library to retrieve metadata from Open Archives Initiative sources; and experiment with sunilnatraj's (n.d.) AI/LLM extension for OpenRefine for data transformation and analysis. Once we have improved our repository's support for publication language specifications, as planned, we are also interested in actively preserving more non-English HSS research to improve the diversity of research materials within the repository and its usefulness as a multilingual resource, in line with our ongoing translation of the HSS Commons' interface into languages other than those already available on the site (English, French, Spanish, Bangla, Portuguese, and Ukrainian).

In addition to these ideas, we have already started to act on our initial research into vanishing HSS and open access scholarship by using resources such as the Keepers Registry (https://keepers.issn.org/keepers) to identify at-risk journals that could be backed up in the HSS Commons repository and showcased using the site's group, project, and collection features. As Sprout and Jordan explain, the Keepers Registry serves as a terrific resource for this kind of work by "providing information about which journals are preserved by which archiving services and highlighting those journals for which no arrangements exist" (2018, 252). As we consider what to preserve in the HSS Commons, we would like to leverage resources of this kind to guide our efforts so that we can continue to play a small role in preserving open, HSS scholarship for future generations while also building out our repository in ways that serve the present and always-changing needs of our community.

Finally, as we carry out this ongoing work, we acknowledge the political and critical dimensions of preservation and archiving (which cannot be summarized adequately here but are, nevertheless, worth mentioning as one of the considerations informing this project).[10] As Samantha Cross remarks, summarizing Randall Jimerson, the

10. For a recent sampling of scholarship on the politics of (web) archiving, see part IV of Aasman et al. (2025).

assumption that archiving is a neutral act "deters active archiving and reduces archivists to passive recipients. In reality, archivists have the potential, if not the responsibility, to act and explore other options of collecting and serving their communities" (2017, 1). For us, our project was a deliberate intervention intended to promote open HSS resources originally published by our Canadian research partners. Inevitably, though, it also involved making choices that privileged certain journals, disciplines, or even languages over others, thus reflecting the academic interests and biases of our own research team as well as those of the larger national and international research communities in which we are embedded. When engaging in this preservation work, we have therefore tried to harness its potential for positive transformation of HSS communities and infrastructures; yet we acknowledge the responsibility that accompanies preserving when one interprets it, as we do, as a political act. To engage in what Cross calls "active, deliberate archiving" (2017, 2), one needs to grapple with myriad questions: What gets preserved? Is it open (or should it be)? Which research areas or researchers does this process (de)prioritize? How might it subtly re-direct conversations, emerging methodologies, or disciplinary fields? And so on. In these ways and others besides, we regard even our own modest efforts to republish and increase the visibility of resources through an open digital knowledge commons as acts that inevitably shape the academic landscape—hopefully to its betterment and to the betterment of those who occupy it.

Again, the kind of exploratory, DIY methods outlined in this article are meant only as a stopgap or complementary solution to the kind of large-scale re-evaluation called for by Laakso et al. (2021); Cramer et al. (2023); and other scholars, librarians, and tech experts already active in this area. Nevertheless, while that re-evaluation unfolds—and as the complex and sometimes arduous associated processes involving changes to policy, preservation initiatives, and related infrastructures gradually evolve—we hope that our patchwork project impresses the importance of preservation, including but not necessarily using the repository, software, and methods foregrounded here.

Open Peer Review Reports

Open peer review reports for this article are available at the following location: https://doi.org/10.17613/ds8xv-78y51

Author Biographies

Graham Jensen (https://ghjensen.com) is Assistant Director and Digital Humanities Research Lead in the Electronic Textual Cultures Lab at the University of Vic-

toria. He is also Principal Investigator of the Canadian Modernist Magazines Project (https://www.modernistmags.ca/). Previously at the University of Victoria, he was a Social Sciences and Humanities Research Council Postdoctoral Fellow and Assistant Professor in English (Limited Term). His research interests include twentieth- and twenty-first-century Canadian literatures, modernism, literature and religion, and digital humanities approaches to open publishing, pedagogy, and community-building.

Sajib Ghosh is a PhD candidate in Linguistics at the University of Victoria. He holds a Master of Arts in English and Applied Linguistics and English Language Teaching, and a Bachelor of Arts (Honours) in English, both from Jahangirnagar University in Bangladesh. He is a Graduate Research Assistant and the Communications Coordinator in the ETCL, where he contributes to the lab's communication activities and assists with various lab events.

Archie To is a programmer and consultant. He graduated from the University of Victoria at the end of 2023. He is currently working full-time for Research Computing Services at the University of Victoria as part of the ARCsoft Team, and as a part-time contractor for the Electronic Textual Cultures Lab.

Ray Siemens (PhD, FRSC; https://web.uvic.ca/~siemens/, ORCID 0000-0002-9599-8795) is Distinguished Professor at the University of Victoria (in Humanities and English, with cross-appointment in Computer Science) and previous Canada Research Chair in Humanities Computing (2004–2015). He directs the Electronic Textual Cultures Lab (ETCL) and the SSHRC-funded Implementing New Knowledge Environments (INKE) Partnership, also founding and now co-directing the Digital Humanities Summer Institute (DHSI) – having served as member of the SSHRC Governing Council, Vice-President/Director (Research Dissemination) of the Federation for the Humanities and Social Sciences, Chair of the Alliance of Digital Humanities Organisations steering committee, and President of the Society for Digital Humanities. In 2019–20, he was Leverhulme Visiting Professor at Loughborough U and, 2019–22, Global Innovation Chair in Digital Humanities at U Newcastle.

References

Aasman, Susan, Anat Ben-David, and Niels Brügger, eds. 2025. *The Routledge Companion to Transnational Web Archive Studies*. Routledge.

Aktorosian, Taleen. 2025. "CARL's Response to the Consultation on Proposed Tariffs." Canadian Association of Research Libraries, April 3. https://www.carl-abrc.ca/news/carls-response-to-the-consultation-on-proposed-tariffs/.

Butterfield, Alexandra Carlile, Quinn Galbraith, and McKenna Martin. 2022. "Expanding Your Institutional Repository: Librarians Working with Faculty." *Journal of Academic Librarianship* 48 (6): 102628. https://doi.org/10.1016/j.acalib.2022.102628.

Canadian Federation for the Humanities and Social Sciences. 2023. "Open Access in the Humanities and Social Sciences in Canada: A Conversation." *Social Science Space*, May 10. https://www.socialsciencespace.com/2023/05/open-access-in-the-humanities-and-social-sciences-in-canada-a-conversation/.

Clark, Brian. 2023. "Proactive Institutional Repository Collection Development Techniques: Archiving Gold Open Access Articles and Metadata Retrieved with Web Scraping." *Journal of Library Administration* 63 (6): 743–65. https://doi.org/10.1080/01930826.2023.2240190.

Cramer, Tom, Chip German, Neil Jefferies, and Alicia Wise. 2023. "A Perpetual Motion Machine: The Preserved Digital Scholarly Record." *Learned Publishing* 36 (2): 312–18. https://doi.org/10.1002/leap.1494.

CRKN (Canadian Research Knowledge Network). 2025. "CRKN's Response to Tri-Agency Draft Open Access Policy." May 2. https://www.crkn-rcdr.ca/en/crkns-response-tri-agency-draft-open-access-policy.

CRKN (Canadian Research Knowledge Network). n.d. "CRKN Requests Financial Relief from Publishers." Accessed May 13, 2025. https://www.crkn-rcdr.ca/en/crkn-requests-financial-relief-publishers.

Cross, Samantha. 2017. "Archivists on the Issues: The Neutrality Lie and Archiving in the Now." *Issues & Advocacy*, March 27. https://issuesandadvocacy.wordpress.com/2017/03/27/archivists-on-the-issues-the-neutrality-lie-and-archiving-in-the-now/.

Dingemanse, Mark. 2016. "How Academia.edu Promotes Poor Metadata and Plays to Our Vanity." *Ideophone* (blog), August 25. https://ideophone.org/academia-edu-poor-metadata-vanity/.

DOAJ (Directory of Open Access Journals). 2018. "The Long-Term Preservation of Open Access Journals." *DOAJ Blog*, September 17. https://blog.doaj.org/2018/09/17/the-long-term-preservation-of-open-access-journals/.

DOAJ (Directory of Open Access Journals). n.d. "JASPER Preservation Service: Open Access Journals Must Be Preserved Forever." Accessed April 19, 2024. https://doaj.org/preservation/.

Dombrowski, Quinn, Tessa Walsh, Anna Kijas, Ilya Kreymer, and Amanda Wyatt Visconti. n.d. "DIY Web Archiving." *Zine Bakery*. Accessed May 17, 2025. https://amandavisconti.github.io/zinebakery//homemade-zines/bakeshop-2-diywebarchiving.

Eve, Martin Paul. 2024. "Digital Scholarly Journals Are Poorly Preserved: A Study of 7 Million Articles." *Journal of Librarianship and Scholarly Communication* 12 (1). https://doi.org/10.31274/jlsc.16288.

Fenlon, Katrina, Jessica Grimmer, Alia Reza, and Travis Wagner. 2025. "The Oyster Model: Understanding Community Roles in Sustaining Digital Cultural Knowledge Infrastructures." *Archival Science* 25:37. https://doi.org/10.1007/s10502-025-09510-z.

gastonbecerra. n.d. "gastonbecerra/ojsr." GitHub. Accessed May 15, 2025. https://github.com/gastonbecerra/ojsr.

Hwang, Soo-Yeon, Susan Elkins, Michael Hanson, Trent Shotwell, and Molly Thompson. 2020. "Institutional Repository Promotion: Current Practices and Opinions in Texas Academia." *New Review of Academic Librarianship* 26 (1): 133–50. https://doi.org/10.1080/13614533.2019.1587483.

INKE (Implementing New Knowledge Environments) Partnership. n.d. "Activity Clusters." Accessed May 15, 2025. https://inke.ca/activity-clusters/#connection.

Iter Gateway. n.d. "Iter Canada." Accessed May 15, 2025. https://www.itergateway.org/iter-canada/.

Jensen, Graham. 2023a. "Creating Connections in and Through Knowledge Commons." Open Scholarship Press, January 13. https://openscholarshippress.pubpub.org/pub/uo36o9ut/release/3.

Jensen, Graham. 2023b. "Digital Knowledge Commons, Scholarly Connection, and the Evolution of Open Scholarship." Open Scholarship Press, November 3. https://doi.org/10.21428/47bc126e.0ca461a4.

Jensen, Graham, Alyssa Arbuckle, Caroline Winter, et al. 2022. "Fostering Digital Communities of Care: Safety, Security, and Trust in the Canadian Humanities and Social Sciences Commons." *IDEAH* 3 (2). https://doi.org/10.21428/f1f23564.ed75625f.

Johnston, Leslie. 2020. "Challenges in Preservation and Archiving Digital Materials." *Information Services and Use* 40 (3): 193–99. https://doi.org/10.3233/isu-200090.

Joseph, Ben, Tomasz Neugebauer, and Cole Mash. 2022. "The Bad Batch: Open Refine as a Batch Editing Method in SWALLOW." *SpokenWeb* (blog), June 8. https://spokenweb.ca/the-bad-batch-open-refine-as-a-batch-editing-method-in-swallow/.

The Keepers Registry. n.d. "Information About Archiving Agencies Which Act as Keepers." Accessed April 2, 2024. https://keepers.issn.org/keepers.

Laakso, Mikael, Lisa Matthias, and Najko Jahn. 2021. "Open Is Not Forever: A Study of Vanished Open Access Journals." *JASIST: Journal of the Association for Information Science and Technology* 72 (9): 1099–112. https://doi.org/10.1002/asi.24460.

Larivière, Vincent, and Jean-Philippe Warren. 2019. "Introduction: The Dissemination of National Knowledge in an Internationalized Scientific Community." *Canadian Journal of Sociology* 44 (1): 1–8. https://doi.org/10.29173/cjs29548.

Lightfoot, Elizabeth A. 2016. "The Persistence of Open Access Electronic Journals." *New Library World* 117 (11–12): 746–55. https://doi.org/10.1108/NLW-08-2016-0056.

LOCKSS Program. n.d. "Frequently Asked Questions." Accessed May 15, 2025. https://www.lockss.org/about/frequently-asked-questions.

Palmer, Kathryn. 2025. "'Draconian' Layoffs, Grant Terminations Come for the NEH." *Inside Higher Ed*, April 14. https://www.insidehighered.com/news/government/politics-elections/2025/04/14/draconian-layoffs-grant-terminations-come-neh.

PKP (Public Knowledge Project). n.d. "Copyright and Licensing." PKP Docs. Accessed January 8, 2024. https://docs.pkp.sfu.ca/journal-policies-workflows/en/copyright-licensing.html.

Rajput, Aakash Amod. 2024. "Exporting Your Work." OpenRefine, last modified January 12, 2024. https://openrefine.org/docs/manual/exporting#overview.

Ruest, Nick, Samantha Fritz, Ryan Deschamps, Jimmy Lin, and Ian Milligan. 2021. "From Archive to Analysis: Accessing Web Archives at Scale Through a Cloud-Based Interface." *International Journal of Digital Humanities* 2 (1): 5–24. https://doi.org/10.1007/s42803-020-00029-6.

Schlangen, Maureen. 2015. "Content, Credibility, and Readership: Putting Your Institutional Repository on the Map." *Public Services Quarterly* 11 (3): 217–24. https://doi.org/10.1080/15228959.2015.1060148.

Sickle. n.d. "Sickle: A Lightweight OAI-PMH Client for Python." Accessed May 17, 2025. https://sickle.readthedocs.io/.

Smiley, WR. 2024. "Exploring Facets." OpenRefine, last modified April 29, 2024. https://openrefine.org/docs/manual/facets.

Sprout, Bronwen, and Mark Jordan. 2018. "Distributed Digital Preservation: Preserving Open Journal Systems Content in the PKP PN." *Digital Library Perspectives* 34 (4): 246–61. https://doi.org/10.1108/DLP-11-2017-0043.

SSH Open Marketplace. 2025. "Technical Aspects." Social Sciences and Humanities Open Marketplace, last updated May 7, 2025. https://marketplace.sshopencloud.eu/about/implementation.

sunilnatraj. n.d. "sunilnatraj/llm-extension." AI/LLM Extension for OpenRefine. GitHub. Accessed May 17, 2025. https://github.com/sunilnatraj/llm-extension.

Tracy, Daniel, and Graham Jensen. 2025. "Humanities Scholars' Needs for Open Social Scholarship Platforms as Online Scholarly Information Sharing Infrastructure." *First Monday* 30 (2). https://doi.org/10.5210/fm.v30i2.13742.

Turp, Clara, Lee Wilson, Julienne Pascoe, and Alex Garnett. 2020. "The Fast and the FRDR: Improving Metadata for Data Discovery in Canada." *Publications* 8 (2): 25. https://doi.org/10.3390/publications8020025.

University of Toronto Libraries. n.d. "Journal Production Services." Accessed January 8, 2024. https://jps.library.utoronto.ca.

Van Bellen, Simon, and Lucía Céspedes. 2024. "Diamond Open Access and Open Infrastructures Have Shaped the Canadian Scholarly Journal Landscape Since the Start of the Digital Era." Preprint, arXiv, November 4. https://doi.org/10.48550/arXiv.2411.05942.

Williamson, Evan Peter. 2022. "Fetching and Parsing Data from the Web with OpenRefine." Edited by Jeri Wieringa. *Programming Historian*, last modified November 4, 2022. https://doi.org/10.46430/phen0065.

Winter, Caroline, Tyler Fontenot, Luis Meneses, Alyssa Arbuckle, Ray Siemens, and the ETCL and INKE Research Groups. 2020. "Foundations for the Canadian Humanities and Social Sciences Commons: Exploring the Possibilities of Digital Research Communities." *Pop! Public. Open. Participatory*, no. 2 (October 31). https://popjournal.ca/issue02/winter.

Wulf, Karin. 2025. "The Humanities as Canary: Understanding This Crisis Now." *Scholarly Kitchen*, April 2. https://scholarlykitchen.sspnet.org/2025/04/02/the-humanities-as-canary-understanding-this-crisis-now/.