

Wrzesinski, Marcel; Riechert, Patrick Urs; Dubois, Frédéric; Katzenbach, Christian

Working Paper

Working with publication technology to make open access journals sustainable

HIIG Discussion Paper Series, No. 2021-2

Suggested Citation: Wrzesinski, Marcel; Riechert, Patrick Urs; Dubois, Frédéric; Katzenbach, Christian (2021) : Working with publication technology to make open access journals sustainable, HIIG Discussion Paper Series, No. 2021-2, Alexander von Humboldt Institut für Internet und Gesellschaft (HIIG), Berlin, <http://dx.doi.org/10.5281/zenodo.4558781>

This Version is available at:

<http://hdl.handle.net/10419/231355>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



MARCEL WRZESINSKI, PATRICK URS RIECHERT, FRÉDÉRIC DUBOIS,
CHRISTIAN KATZENBACH

Working with publication technology to make open access journals sustainable

ABSTRACT

Over the last 25 years, scholars around the world have used electronic publishing to open up their work, share it with interested publics instantly or even become publishers themselves. This white paper explores in what ways advances in publication technology in the journal sector (e.g. the widespread use of content management and editorial systems) contributes to a more inclusive and sustainable open access ecosystem. Drawing on a study we did in Germany in 2019–2021, and for which we tested technical solutions together with the international, peer-reviewed diamond open access journal *Internet Policy Review*, we present and discuss publishing solutions based on software, workflows, and collaborations with regard to their practicability and scalability. The paper finds that scholar-led publishing is a force to be reckoned with when it comes to technical solutions tending towards increased bibliodiversity (i.e., variety of content, publication formats and publishing institutions).

KEYWORDS

Open access, Scholar-led, Publication technology, Sustainability, Small sciences

CITATION

Wrzesinski, M., Riechert, P. U., Dubois, F., & Katzenbach, C. (2021). Working with publication technology to make open access journals sustainable. HIIG Discussion Paper Series 2021–2. 27 pages. <https://doi.org/10.5281/zenodo.4558781>

LICENCE

This work is distributed under the terms of the Creative Commons Attribution 4.0 Licence (International) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<https://creativecommons.org/licenses/by/4.0/>). Copyright remains with the authors.

AUTHOR INFO / FUNDING

Marcel Wrzesinski is Open Access Officer at the Humboldt Institute for Internet and Society and a researcher in the project “Innovative Open Access in the Small Sciences”. He worked with several Open Access Gold journals in order to create and sustain structures of scholar-led publishing.

Patrick Urs Riechert is Student Assistant at the Humboldt Institute for Internet and Society where he works as editorial assistant for e-journal *Internet Policy Review*. He is an MA student in Political science at Freie Universität Berlin.

Frédéric Dubois is Managing Editor of *Internet Policy Review*, an open access journal on internet regulation published by the Humboldt Institute for Internet and Society.

Dr Christian Katzenbach is Senior Researcher and Research Group Leader at the Humboldt Institute for Internet and Society. He co-directs the interdisciplinary research program “Digital Society“ and leads multiple international projects on platform governance, controversies around artificial intelligence, and open access.

The project “Innovative Open Access in the Small Sciences” is a cooperation between the Humboldt Institute for Internet and Society and the ZBW—Leibniz Information Centre for Economics and was funded by the German Research Foundation (DFG) between June 2019 and February 2021.

CONTENT

1 Introduction	5
2 Technical innovations fostering sustainability	6
3 Open access and web accessibility	7
3.1 Web accessibility and accessibility evaluation	7
3.2 Accessible output formats	8
3.3 Key steps	9
3.4 Use case <i>Internet Policy Review</i>	9
4 Archiving, identification, and indexing: Making research available, discoverable, and impactful	11
4.1 Availability through long-term preservation	11
4.2 Identification services for electronic publishing	12
4.3 Indices and metrics for research dissemination	14
4.4 Key steps	17
4.5 Use case <i>Internet Policy Review</i>	17
5 Managing and documenting workflows: Open source (journal) systems	20
5.1 Editorial management systems as tools of sustainability	20
5.2 Key steps	21
5.3 Use case <i>Internet Policy Review</i>	21
6 Recommendations for sustainable open access publishing	24
6.1 Streamlining workflows with professional publication software	24
6.2 Displaying and contextualising altmetrics to showcase bibliometric diversity	24
6.3 Increasing impact through repository-based archiving and indexing	25
6.4 Providing accessible, single-source output formats	25

6.5 Identifying authors and institutions via open databases and standards	25
7 Conclusion	25
8 References	26

1 Introduction

The publishing industry transformed more in the last 25 years than it has in the previous three hundred, largely due to the digital revolution: new business models arose (Björk and Korkeamäki, 2020) and an innovative culture of writing and reading emerged. Suddenly, electronic publishing enabled researchers around the world to open up their work, share it with an interested public instantly, or become publishers themselves (e.g., by running open access journals, university presses, or other publishing initiatives). This coincided with the formation of the open access movement, which questioned the toll access model and fostered an equitable, open system of science.¹ And besides the many good political reasons for publishing research open access,² there are also pragmatic and infrastructural advantages of electronic publishing and the technological revolution behind it: knowledge is more powerful if created together, it can be disseminated widely and more quickly, and its scientometric impact can be measured more efficiently.³

Against this backdrop, this white paper explores in what ways advances in publication technology and electronic publishing in the journal sector (e.g., the widespread use of content management systems and the rise of editorial management systems like OJS) can contribute to a more inclusive and sustainable open access ecosystem, therefore increasing independence, productivity, and quality. As such and acknowledging the diversity of stakeholders within that ecosystem, this paper addresses not only those actively involved in publishing academic content but also funders and readers. This wide audience—including libraries, journals, digital data infrastructures, and funding organisations—means this paper is broad in scope and adopts a somewhat more generic approach.

This paper is situated in the wider context of the quest for sustainable scholar-led publishing and bibliodiversity,⁴ wherein we understand ‘scholar-led’ as having scholars actively leading publishing projects independently of large and commercial publishing houses. In this understanding, ‘scholar-led’ includes the important work of small and non-profit publishing houses, which are indicative of the diversity especially of small and interdisciplinary fields of research. Within the DFG funded project ‘Innovative Open Access in Small Sciences’ (June 2019 to February 2021), we addressed questions of sustainable publishing with regard to both business models and technical solutions—building on previous efforts to contribute to innovations within open access publishing. The project tested these solutions together with the international, peer-reviewed diamond open access⁵ journal *Internet Policy Review* (cooperatively published by the Alexander von Humboldt Institute for Internet and Society), which has already been using public and private funding to improve its technical infrastructure and to introduce new features facilitating a more timely and diverse

¹ Innovations in the publishing industry on the infrastructural, strategic and actor levels (see Ponte et al., 2017, p. 97).

² Editorial management systems like OJS allow people (editors, reviewers, authors) from around the world to be involved in the publication process. To his extent, scholar-led publishing might further attempts to decolonise open access publishing (Willinsky and Mendis, 2007; Piron, 2018).

³ See for different open science ‘schools of thought’ and their implications Fecher and Friesike (2013).

⁴ Bibliodiversity refers to the idea that—as in nature—there needs to be a variety of content, publication formats, and publishing institutions in order to maintain a sustainable publishing ecosystem, particularly with regard to open access ambitions (see the Jussieu Call for Open Science and Bibliodiversity; <https://jussieucall.org/>).

⁵ Different types of open access are coded with different colours. The most prominent ones are: ‘gold’, which requires immediate access, ‘green’, which refers to self-archiving, ‘hybrid’, which is a mix of models, and ‘diamond’, which is immediate open access without any author-facing charges.

way of disseminating the research of its peer community. This work was guided by the idea to provide a better practice in publishing, incentivise reuse of these improvements, and to realise *Internet Policy Review's* open access mission.⁶ The same approach motivates this white paper: we present and discuss publishing solutions found and tested during the project—either based on software, workflows, or collaborations—with regard to their practicability (How does it work for us?) and scalability (How can it work for others?). This is a work in progress, and we welcome all feedback on how to ensure continuous technical innovation to further strengthen scholar-led journals.

The white paper is organised in five sections. This structure is motivated by our understanding of sustainability and therefore lays out arguments and measures to improve the (inter-)operability of open access journals. We start by connecting the question of technology to a workable concept of sustainability in journal publishing (section 1). Then we explore crucial services that contribute to certain levels of sustainability in journal publishing, addressing accessibility issues regarding the journal's web environment and output formats (section 2); appropriate long-term preservation, identification, and indexing of research articles (section 3); and streamlining workflows with editorial management systems (Ch. 4). We do so to provide recommendations at the end that hopefully will empower journals to sustain their work with the technologies and networks at hand.

2 Technical innovations fostering sustainability

The widely proclaimed and desired open access transformation seeks to establish a sustainable journal ecosystem without access fees. In that regard, the main goals include keeping journals independent, productive, and quality-controlled in the long-run. Yet, we can observe that—while flipping toll-based journals to OA is widely subsidised—the day-to-day operation of scholar-led journals often rests on rather precarious and uncertain conditions, with editorial teams working on low or even zero budgets (see InnOAccess white paper on business models).

In addition to sound budgets, the sustainable operation of OA journals is contingent on the installation and routine usage of up-to-date publication infrastructure and technology. Backend systems facilitate smooth submission and review processes and ease the publication of articles for everyone involved. In the traditional model, commercial publishers usually provide this function to academic editors. In scholar-led OA contexts, academic editorial teams usually take care of this themselves, often building up piecemeal editorial set-ups resting on open source systems such as Open Journal Systems, open source editing tools such as Etherpad or Collabora, and commercial yet free-of-cost products such as Google Docs. Their business models refer to fair and equitable open access principles⁷ and in many cases they are aiming to transform the ways we process and disseminate research results as a whole. In this regard, many journals already fulfil the 'basic mandatory conditions' outlined by the cOAlition S in their 'technical guidance and requirements':⁸ the use of persistent identifiers for articles, ensuring long-term preservation and archiving of

⁶ This mission is aligned with the 'Principles of Transparency and Best Practice in Scholarly Publishing' issued (vers. 3 in January 2018) by the Committee on Publication Ethics (COPE), the Directory of Open Access Journals (DOAJ), the Open Access Scholarly Publishers Association (OASPA), and the World Association of Medical Editors (WAME) (<https://oaspa.org/principles-of-transparency-and-best-practice-in-scholarly-publishing-3/>)

⁷ See the FAIR OA Principles (<https://www.fairopenaccess.org/the-fair-open-access-principles/>).

⁸ See the Plans S 'Technical Guidance and Requirements', especially section 1 on the 'Requirements for Publication

their content, providing high-quality and open metadata to all relevant parties. This white paper follows-up on the guidelines and hopes to incentivise a wider adoption of those and further publishing practices.

As we will see later on, it is worth asking: how much of this independence (and sometimes DIY spirit) should someone be willing to give up sustaining a publishing project? Given the limitations of standardised editorial management systems (see section 4), is their use more a constraint than a publishing benefit? Can the attempt at inclusion by measures of web accessibility legitimise the tremendous additional workload that comes with adjusting your website and PDFs to accessibility standards? Do the advantages of interoperable and aggregated metadata, identifiers, and indices outweigh the legitimate concerns of privacy advocates in the age of platform societies? Who is responsible for providing the necessary ‘digital literacy’ for editors to keep them up to date? Addressing these questions often leads to intensive restructuring and rethinking on the part of editors, the staff dealing with technical issues, and the publishing institutions. Yet the workload for each measure is unequally distributed: something that is beneficial from an editor’s perspective and with regard to the journal outreach and reputation might come with extraordinary challenges for the programmers or coders. This needs to be acknowledged for every step along the road to a more sustainable publishing approach.

3 Open access and web accessibility

Open access not only refers to the ability to access the preprint, article, or book without reader-facing fees or charges but also concerns its dissemination and impact. The criteria of ‘accessibility’ here—with respect to the full possibilities of electronic publishing—includes the notion that research is in essence discoverable and readable for everyone regardless of their abilities or the technology they are equipped with. In this way, truly open access literature is available to people with disabilities, to people with below-average technological equipment, and to people and regions with limited access to the internet.⁹ With regard to academic publishing, necessary measures may include screen-readable and moderately sized output formats, responsive and tagged websites, and an awareness of responsibility towards the journal’s audience.

3.1 Web accessibility and accessibility evaluation

But what is an accessible publication? Rothberg (2018) proposes to ‘think of a publication as having a structure like an onion, with layers serving different purposes [...]. At the centre is the content itself – the text, images, or interactive widgets that make up the publication. These are surrounded by a structure that puts that content in chapters, sections, and so on. A delivery format wraps around the structure, and then catalogue information is added to enable online sale, rent, or loan of the publication’ (p. 45). Subsequently she outlines how each layer of additional information beyond the mere words or code can and should be designed using accessible technologies, markup languages, and standards. In this context, the most common references for actual web requirements are the Web Content Accessibility Guidelines (WCAG version 2.0) by the Web Accessibility Initiative (WAI),¹⁰ accessibility compliant output standards such as EPUB (version

Venues’ (https://www.coalition-s.org/technical-guidance_and_requirements/).

⁹ See the discussion of web accessibility standards and disability (Lewthwaite, 2014) and issues of electronic accessibility in the context of the Global North/South hierarchy (King et al., 2018).

¹⁰ They define the ‘Four Principles of Accessibility’ with regard to web content being ‘perceivable,’ ‘operable,’ ‘understandable,’ and ‘robust’ (<https://www.w3.org/TR/UNDERSTANDING-WCAG20/intro.html>; see Andrae et al.,

3.1) and PDF (PDF/UA), accessible and standardized metadata, and finally the general consensus to exempt publication from copyright law ‘to encourage the sharing of accessible document formats between countries.’¹¹

Attempts to improve the accessibility of an electronic resource should ideally start with an accessibility evaluation. For journal publishing, this typically requires using online evaluation tools to assess the publishing frontend. The most widespread, open source content management systems for journal publishing like Open Journal Systems, or in general like Wordpress or Drupal, have their own web accessibility standards already implemented, although the grade of actual compliance depends on how much customisation has been done and which additional plugins and tools are utilised.¹² Still, a brief check with automated testing tools about non-compliant features and design elements is recommended in order to ensure a maximum of inclusivity.¹³ Online tools and browser extensions like WebAIM or HTML Code Sniffer identify issues such as: ‘Presence of alternative text for images, presence of form labels, presence of page title, language indicator, presence of landmarks, contrast ratios, proper table structure and tags, proper semantic structure, and appropriately nested markup’ (Waecker et al., 2019, p. 148). Dealing with these web-based issues is rather time consuming and should be done in accordance with both the journal’s and the publishing institution’s design/identity. The next step in this regard is manual testing using common assistive technologies such as screen readers, either by simulating a case environment or working with native users of assistive technologies. Finally, if possible, it is fruitful to reach out to disability representatives at your institution, region, or the national disability/accessibility organisation and request a live test of the modified content.¹⁴

3.2 Accessible output formats

Alongside this core web accessibility, the output format and layout are important factors for journals, as are the way the content is structured, pictures are used, and charts are implemented. Here, users must be able to navigate easily through the document, and assistive technologies must be given certain orientation points within the document’s structure (e.g., semantic tags and semantic markups including alternative descriptions). Yet in some cases, questions of accessibility might conflict with prior or current standards of the journal’s design identity: muted colours or contrast issues, no clear structure of (sub)headings,

2020, paragraph 7).

¹¹ Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired, or Otherwise Print Disabled (MVT) (2013); exemptions and exchange only between participating countries or countries that have ratified the treaty.

¹² The Radical Open Access Collective has published an extensive, commented list of open access publishing tools (last update in May 2019): <https://radicaloa.disruptivemedia.org.uk/resources/publishing-tools/>

¹³ Kuykendall (2017) points out that most automated testing tools can only catch 10–30% of errors, and suggests workarounds. <https://www.microassist.com/digital-accessibility/role-web-accessibility-testing-tools/>

¹⁴ Public institutions in Germany are bound by the German disability act (*Gesetz zur Gleichstellung von Menschen mit Behinderungen*; BGG) to ensure accessibility wherever possible. The German Federal Office for Accessibility (<https://www.bundesfachstelle-barrierefreiheit.de/>) and the Informationstechnik Zentrum Bund (https://www.itzbund.de/DE/Leistungsangebot/Beratung/BGG/bgg_node.html) offer counselling, training, and workflow examples on all related matters, as well as on a European level, this is clarified by the ‘Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies’ (<http://data.europa.eu/eli/dir/2016/2102/oj>).

reflowable texts, adjustable font size and colours, searchable images and others (see Waecker et al., 2019, p. 150). This applies in particular to the most used Portable Document Format (PDF; for long-term archiving it is the substandard PDF/A¹⁵), where a minimum accessibility requirement is to comply with the latest ISO standard, based on the WCAG guidelines, and therefore create PDF with ‘universal accessibility’ (PDF/UA).¹⁶ To avoid any confusion: these accessibility issues and possible solutions are complex and, in many cases, institutions cannot comply based on their available resources. It is imperative to understand, though, that there is no web environment that is free of barriers and therefore universally accessible; reducing barriers and creating inclusive web/publishing content is an ongoing process with no definite end or overarching solution. Consequently, content creators and journal publishers should educate themselves and try to understand key element of web accessibility, familiarise themselves with the main accessibility testing tools (including WCAG), be a contact point for inquiries about accessibility issues, release an accessibility statement indicating current issues and solutions, follow-up on accessibility requests diligently, and finally choose ‘born accessible’ (like open access first) wherever it is possible.¹⁷ As Waecker et al. (2019) and many others point out, creating an accessible publishing ecosystem, especially in the realm of scholar-led open access, is a joint effort by authors, publishers, librarians, and aggregators alike.

In general, accessible documents are based on structured tagging, markups, and alt-text options. Therefore, any html and xml output formats are a more sensible and sustainable solution from an accessibility standpoint, since they can be universally read, reused, and exported in different formats. Yet, setting up fully automated and functional xml workflows is time consuming, and dedicated jats/xml reader software is rare (see chapter below).¹⁸

3.3 Key steps

- Schedule an accessibility evaluation and inform institutional representatives
- Run testing tools (WebAIM and HTML Code Swiffer) for an overview of your website
- Check if your output formats comply with accessibility standards (e.g., ISO)
- Plan further testing of your measures with assistive technology and native users
- Consult with disability representatives and organisations to start a discussion

3.4 Use case *Internet Policy Review*

As an open access journal with a focus on digitalisation, *Internet Policy Review* is committed to providing its content as accessibly as possible. Within the project context, we adjusted the web appearance as well as the

¹⁵ See explanation by the PDF Association (<https://www.pdfa.org/accessibility-what-pdf-a-1a-really-means/>).

¹⁶ See ‘Document management applications - Electronic document file format enhancement for accessibility - Part 1: Use of ISO 32000-1’ (PDF/UA-1) (ISO 14289-1:2014). ‘Accessibility NGOs provide checklists and guidelines on how to produce accessible PDFs’ (<https://www.einfach-fuer-alle.de/artikel/checkliste-barrierefreie-pdf/>)

¹⁷ This list of practical steps follows in some part a list of ‘five desirable skills for library staff who work with online publishing platforms’ (Waecker et al., 2019, p. 152). It was amended with the call for an accessibility statement.

¹⁸ We focused on the PDF output and web environment, while briefly touching upon *Internet Policy Review*’s XML workflow. ePub could not be addressed in this paper, yet there are extensive guidelines for how to accessibly mark-up this publication format provided by the W3 consortium (<https://www.w3.org/Submission/epub-a11y/>). See also the Accessible Publishing Knowledge Database on how to implement the WCAG 2.0 standard with ePub and HTML (<http://kb.daisy.org/publishing/docs/>).

PDF output, which was the main source of delivering designed, structured content to our readers. After testing the frontend of *Internet Policy Review* (<https://policyreview.info/>) using the online evaluation tool webAIM Wave, we were able to identify certain accessibility issues that indicated non-compliance with WCAG 2.0 Level AA, mostly related to alternative text options, colour contrast, font size, and redundant/empty link information. Based on the workflow and workload of the core editorial team and in cooperation with the web design agency, we drafted an implementation plan, assigned milestones, and secured support from the publishing institution for long-term solutions. Fundamental accessibility issues were outlined in the *Internet Policy Review* accessibility statement, created in accordance with the WCAG and using the statement generator by the W3C Web Accessibility Initiative.¹⁹

With regard to the output formats, *Internet Policy Review* was looking into ways to automate the production of PDFs, ideally through promoting a single-source publishing workflow. After initial funding to establish this single-source workflow and provide XML galleys, *Internet Policy Review* rerouted to creating PDFs from the HTML galleys published via its content management system (CMS) Drupal. Because no open source applications for automated PDF creation met the demands, especially with regard to widely accessible PDFs (PDF/UA), the journal reached out to for-profit software developers. Emphasising the accessible and non-profit nature of its publication concept, the most promising avenue for open access journals was to approach software developers and negotiate fair terms and conditions for the use of their products. The solution ended up being a partnership with YesLogic Pty Ltd, makers of the Prince PDF formatter. This software's unique feature lies in its ability to create widely accessible PDFs (i.e., PDF/UA; ISO 14289) directly and automatically from HTML or XML sources without requiring a labour-intensive manual tagging process for each article. The final workflow involves the following steps:

1. [manual] Humans prepare the article: proofreading and formatting it as well as adding alt-texts for images
2. [automatic] A script converts a .docx file into an html file
3. [automatic & manual] References are parsed with Anystyle.io, imported into and corrected in Zotero, and then exported in human-readable APA 7 and machine-readable formats.
4. [manual] A human inserts the article and author metadata into the CMS
5. [automatic] A custom CMS module produces a JATS XML file containing the article content as well as its metadata
6. [automatic] A second custom CMS module sends the HTML version of the article and the JATS-formatted metadata to an external server. This server runs Prince to produce a widely accessible PDF into which it also embeds author, title, keyword, and licensing metadata. The CMS module can also be configured to archive PDFs with a repository via email.

This workflow minimises the labour time expended per article: the responsible person only needs to insert metadata once, and as much of the process as possible is handled in an automated fashion. That said, the

¹⁹ The generator is only in the W3C-WAI section (<https://www.w3.org/WAI/planning/statements/>).

Here we indicate the conformance status, a point of contact and complaint, technical specifications, and all known accessibility issues including alternatives (<https://policyreview.info/accessibility>).

implementation requires coding proficiency and should not be underestimated.

Finally, inspired by the example of the Open Library of Humanities, *Internet Policy Review* added a ‘dyslexia button’ that adjusts article formatting to be more legible for people with dyslexia. Specifically, it changes the background colour and adjusts line, letter, and word spacing; it also replaces italics with highlights and modifies the appearance of links. This constitutes a relatively low-effort investment that offers certain readers valuable advantages.

4 Archiving, identification, and indexing: Making research available, discoverable, and impactful

Most common notions of (libre) open access emphasise the accessibility and reusability of academic literature. Taking into account that open access articles are published electronically, the question of how to discover such literature (and their respective authors) online is of equal importance. This can be sustainably realised by using existing infrastructures and initiatives that deal with the (1) depositing and long-term archiving of publication output, the (2) persistent and interoperable identification of literature and authors, and the (3) meaningful indexing and measuring of research dissemination and citations.²⁰

4.1 Availability through long-term preservation

Any journal output format needs to be stored and deposited in order to comply with the basic requirements of academic research and ensure long-term availability. And while the most commonly used way of creating journal output is by delivering the research article in the form of a PDF, other formats such as XML or HTML are more suitable for archiving. Regardless of the choice of format, journals are advised to use a ‘dedicated, digital archiving and preservation service’ (see DOAJ guidelines; <https://doaj.org/publishers#advice>). For these procedures and services, sometimes referred to as ‘Long Term Preservation and Archiving (LTPA),’ open repositories are the first point of contact: using standardised workflows, assigning interoperable metadata, and providing crucial persistent identifiers in many cases, open repositories address the huge challenge of digital preservation in times of digital overflow in a transparent way and provide the gateways for the vast electronic archives.²¹ Yet, digital preservation is a complex field involving expertise from different areas such as library and information science, digital humanities, and electronic publishing. Here, libraries, repositories, and other infrastructural facilities cooperate in the best interest of the research community and provide tailored, sustainable solutions for journals.

The necessity for long-term archiving aside, having a solid and recognisable archiving policy in place and therefore being indexed in certain databases may attract conscious authors and increase the credibility of an academic journal. Listed as one of the seven key criteria for their official seal, the Directory of Open Access Journal (DOAJ) refers to a variety of options for long-term archiving, while maintaining strictly that the

²⁰ A comprehensive overview structured according to topics such as visibility, impact, continuity, and ethics can be found at PKP’s docs section ‘Getting Found, Staying Found’ (<https://docs.pkp.sfu.ca/getting-found-staying-found/en/>).

²¹ There are institutional and disciplinary repositories, addressing either the internal community of researchers or a thematic field of research. Databases like the Directory of Open Access Repositories (<https://v2.sherpa.ac.uk/opensoar/>) provide an overview.

‘journal content must be continuously deposited’ in a proper digital archive.²²

4.2 Identification services for electronic publishing

Clearly identifying research and making it permanently citable is one of the most important tasks of electronic publishing, especially in consideration of the vast quantity of digital items available. For journal publications this concerns first and foremost the article level itself, while journals can apply for an International Standard Serial Number (ISSN) as a unique identifier. Assuming that the aforementioned long-term archiving succeeds in securing the availability of research (either the article text or data), assigning a persistent identifier (like a Digital Object Identifier, DOI) makes a digital object permanently findable and prevents empty or broken links when referencing. This can be done manually by subscribing to fee-based DOI distributors (e.g., CrossRef preferably for research literature and DataCite for research data); by depositing literature or data in an open or toll access repository (see above) or open registries like Zenodo. Alternatively, scholar-led publishers can utilise editorial management tools like OJS (and its DOI registration plugin) in cooperation with infrastructural institutions like libraries to obtain DOIs.

Identification services for authors work in a different way but with a similar purpose.²³ In a globalised world with an international research community, the variety *and* similarity of human names makes it difficult to attribute consistently research output to the research originator. The situation is compounded by the introduction of other symbols or scripts from e.g., Chinese, Arabic, or Kyrillic; there are also problems when people have names based in languages with a heavy use of diacritical marks or just when they change their names. Even with a correct affiliation, this problem arises when institutions are not cited consistently or the institution’s county changes its name (see Mirčić and Đokić, 2018, p. 83). Yet connecting research and researchers is of utmost importance, since the present academic system demands a visible and consistent research profile, with implications when applying for research grants or tenured positions. Many institutions and research funders are already committed to certain author identification, and therefore journals are advised to take this opportunity of supplying bibliographic metadata via author identification services into account.

A widely accepted solution is to use mostly proprietary identification services and aggregating databases,²⁴ which assign a persistent ID to researchers. But in terms of sustainability, transparency, and the mission of open access, it is recommended to use open services and registries like the Open Researcher and Contributor ID (ORCID) introduced in 2012.²⁵ ORCID is a free, non-proprietary 16-digit alphanumeric, unique code identifying an author, fully compatible with certain international standards (e.g., ISO), connected to a large number of established information systems (Web of Knowledge, Scopus, Crossref

²² All seven criteria can be found on the DOAJ website (<https://doaj.org/apply/seal/>).

²³ Besides author identification, there is a community-led, open registry project for unique institutional IDs: the Research Organization Registry (ROR) is based on a GRID set of data, assigns institutional ID and aggregates corresponding metadata (<https://ror.org/>).

²⁴ Well-known examples include the Scopus Author ID, Thomson Reuter’s ResearcherID, ORCID identifier, and ISNI (International Standard Name Identifier); smaller initiatives include RePEc Author, LATTES, VIAF, NARCIS, arXiv Author ID, Names Project Mimas, Chinese NSL Author Service and Author Claim (see Memon and Azim, 2019, p. 890)

²⁵ For further information see the international ORCID Website (<https://orcid.org/>) and the German project partner ‘ORCID DE’ (<https://www.orcid-de.org/>).

etc.), large publishing houses, editorial management systems, and funding organisations. Additionally, it is interoperable in a sense that allows journals to share metadata (import/export) automatically and therefore enhance the visibility of its output.²⁶

For journal publishers and publishing institutions, the advantages and strength of this service are clear: using interoperable, standardised metadata including an ORCID enables the mapping of research output, allows for the automated update and retrieval of information between the editorial management system and aggregating services, and eventually facilitates monitoring of an article's impact and whether it has been appropriately tracked or is compliant with open access policies. Added to this, there is a certain marketing quality: using innovative, open source identification services like ORCID can attract technologically interested authors and serve as a unique selling point within a competitive journal market.

Of course, aggregating data and sharing it within a wide array of information systems and databases may raise concerns about privacy rights and malpractice. As a non-proprietary open source system, ORCID aims at protecting the user's identity and data, e.g. by letting the author customise which information is public/private or which secondary database or API is granted access to such information. In sum, as a recent study (Schallaböck and von Grafenstein, 2017) shows, there are no major privacy issues apparent and the advantages outweigh the disadvantages. Yet problems remain: ORCID is not and does not aim to be a repository, so no digital object can be stored, archived, or persistently identified. And as with any self-curated profile, the user is responsible for keeping it up-to-date and applying rigorous research ethics.

Identification service	Description	Website / contact
Open Researcher and Contributor ID (ORCID)	ORCID is a nonproprietary, alphanumeric code to uniquely identify authors and contributors.	orcid.org
Research Organization Registry (ROR)	ROR is a community-led project to develop an open, sustainable, usable, and unique identifier for every research organisation in the world.	ror.org
Global Research Identifier Database (GRID)	GRID is an openly accessible database of educational and research organisations worldwide, where each organisation is assigned a unique GRID ID and a corresponding web address and page in the database.	www.grid.ac
Scopus Author Identifier	Scopus Author Identifier provides unique author identification by assigning each author in the proprietary, subscription-based Scopus infrastructure a unique number and grouping all of the documents written by that author.	www.scopus.com
Thomson Reuter's Researcher	Researcher ID is a proprietary, unique and	www.researcherid.com

²⁶ Memon and Azim outline a few other interoperability measures and contextualise it for Pakistan (Memon and Azim, 2019, p. 889), where the convention of using multiple names and in different orders question the Western notion of a first and last name.

Identifier	persistent identifier for authors and assists with correct attribution of work. The ID is connected to platforms such as Publons, Web of Science, and ORCID.	
arXiv Author Identifiers	arXiv public author IDs connect researchers and their papers within the arXiv depositing infrastructure. They are syntactically not bound to the arXiv user ID and can be connected to an active ORCID (including API).	arxiv.org

Table 1: Popular identification services for authors and institutions

4.3 Indices and metrics for research dissemination

Very much connected to any attempt at increasing visibility for research through archiving and proper identification is the need to be included in research indices that provide quality assessment through a set of admission criteria—something that may determine an author's decision of where to publish. For open access journal publishing, the Directory of Open Access Journals (DOAJ) is to this day the most pertinent database. As it indicates on its website, the DOAJ is a ‘community-curated online directory that indexes and provides access to high quality, open access, peer-reviewed journals. DOAJ is independent. All funding is via donations, 18% of which comes from sponsors and 82% from members and publisher members. All DOAJ services, including being indexed in DOAJ, are free of charge. All data is freely available. DOAJ operates an education and outreach programme across the globe, focusing on improving the quality of applications submitted.’ (DOAJ.org, 2020) Based on a set of 58 questions and the respective guidelines²⁷ a rigorous evaluation and review process takes place in order to determine the quality and accuracy of suitable archiving, fair financing models, reuse of (meta)data, identification, and independence of the editorial and review process. The strict criteria imposed by the DOAJ in many cases worked as an incentive for journals to further implement open access aspects that contribute to a more professional and sustainable approach of their scholar-led ambitions: either by familiarising and eventually complying with present digital archiving policies (like CLOCKSS or LOCKSS), rethinking their take on APCs (by establishing a waiver option or removing them completely), or deciding to use more open licensing. As studies have shown (Marchitelli et al., 2017) ‘there is a noticeable quality improvement of the journals indexed in DOAJ.’²⁸

Within social sciences and humanities, the ERIH PLUS (formerly known as ERIH) index collects bibliographic information for journals that comply with yet another set of criteria.²⁹ The index is not focused on open access journals but remains a credible source for establishing that a journal has a transparent approach and neutral editorial policy. Established by the European Science Foundation and

²⁷ See the DOAJ application criteria and guidelines (<https://doaj.org/application/new>).

²⁸ A discussion of the new criteria established by the DOAJ in 2015 is provided by Olijhoek, Mitchell and Bjørnshauge (2015).

²⁹ See the complete list of criteria by the Norwegian Centre for Research Data (https://dbh.nsd.uib.no/publiseringskanaler/erihplus/about/criteria_for_inclusion).

currently led by the NSD – Norwegian Centre for Research Data, it functions as a database for scholarly communication and exchange in Europe and beyond, while enabling researchers and publishers to promote and disseminate their outputs in a national and international setting.

Given that commercial platforms such as Google Scholar or SSRN have over the last decade become key resources for finding available literature, it is worth considering them. Journals should not necessarily invest resources in optimising their articles for these platforms, but they should at least make sure that, in situations where authors want to work with these players, they know how to use these platforms.

The standard for evaluating the relevance and scientometric impact of research across disciplines is to create **metrics** by aggregating data on citation counts and calculating a factor that indicates how much this research contributes to scientific progress or prompts discussions on a subject matter. Conversely, the previously described reputation system is built upon the complex question of ‘impact,’ and bibliometric systems aim at providing answers to this question. Scholar-led journals are advised to familiarise themselves with those standards of measuring impact and therefore assess whether it is relevant to the academic community they are servicing; at the same time, they can advance alternative ways and metrics (‘altmetrics’; see below). Either way, without a database of committed (and paying) subscribers, open access journals have to address the question of scientific impact.³⁰

The field of biblio- or scientometrics is diverse and their methods are highly complex. Therefore we would like to only give a brief overview of current practices for measuring relevance and recent developments that are connected to open science policies. In general, metrics are provided on a journal-level, an article-level, and an author-level. Originally, metrics were developed to measure the impact of research at the journal-level: the widely known ‘Journal Impact Factor’ (JIF) as well as the Science Citation Index (SCI)/Science Citation Index Expanded (SCIE) are used as a basis for bibliometric analyses. Author-level metrics can be based on journal-level metrics but have developed into a metrical instrument of their own indicating the authors ‘productivity’ and ‘impact’ within the research community (e.g., the ‘h-index’). Lastly, article-level metrics indicate citation and usage of an individual scholarly article within their respective communities as a basic citation count. And while these metrics serve as a factual account of the number of citations and are therefore proxies of an assumed popularity at a given time in a given community, multiple studies point out that they are not suitable for evaluating a researcher’s productivity or the quality of work, especially not in performance review processes.³¹ At the same time, there are plenty of discussions about whether open access diamond journals have a general citation advantage due to their unrestricted availability—at least in terms of journal-level metrics, there is no evidence for this (Dorta-González and Santana-Jiménez, 2018).

A relatively new phenomenon emerging in close connection to the open access movement and the increased usage of social media by academics are alternative citation indicators utilised to determine the

³⁰ In this white paper we will not address the question of how to attract readers through commercially acting, scholarly networking platforms (e.g. Academia.edu, ResearchGate) or the dedicated use of social media (e.g., via CMS plugins or involvement on Twitter).

³¹ The project ‘*metrics’ (<https://metrics-project.net/>) tried to provide a more meaningful understanding of conventional metrics. A citation factor can by no means indicate a certain level of academic quality. And while we will not discuss the issues of quality and research ethics in this paper, there is strong evidence that qualitatively questionable publications (in the wake of ‘predatory publishing’) are not acknowledged quantitatively anyway (Björk et al., 2020).

short term influence of a journal, author, or article.³² These ‘altmetrics’ aggregate certain sets of non-traditional data measuring the conversations around research outputs on social media platforms like Twitter, with regard to the process of social bookmarking (or social cataloguing and social tagging) and citations in new or mass media formats (such as Wikipedia and the press). Contrary to established bibliometric approaches and databases that rely on citations in academic publications, altmetrical statistics therefore are immediately available and significant, avoiding a ‘citation lag’ (the time lag between when an article is first published and when it is cited in a subsequent publication) but still being applicable to traditional and non-traditional publishing formats. Most tools regarding altmetrics use persistent identifiers and public APIs to aggregate data in a wide range of platforms, depending on the scope and proprietary or non-proprietary nature. An extensive toolkit to better understand and use metrics, especially for researchers and evaluators, is the ‘Metrics Toolkit,’ edited by professionals in the field of bibliometrics (<https://www.metrics-toolkit.org/>). As pointed out by several studies and research projects, these altmetrics need to be further contextualised, e.g., by asking why a certain article is ‘trending’ on Twitter, considering whether certain buzzwords informed that trend and what it means for the state of scholarly communication with the public.³³ Still, ‘altmetrics’ are often seen as mere community tools, therefore reflecting a scepticism towards the usage of social media in science or metrics in general. Consequently, many journals still rely on the impact factor as the lead indicator and established (proprietary) databases.³⁴

All these kinds of databases and indices obviously contribute to a questionable focus of the academic publishing system on reputation, excellence, and so-called impact. Since they are (mis)used to qualify a researcher’s publications within job applications processes, they unfortunately add to the widespread ‘publish or perish’ principle, therefore simply quantifying article publication in assumed ‘high’ quality journals instead of assessing the suitability of the candidate in an in-depth selection process. Still, being included in those indices creates visibility in many ways: most bibliometric or information studies analysis draw their samples from such databases, which in reverse means that being indexed leads to the status of being empirically relevant—‘those things which are measured tend to come to matter.’ (Neylon, 2020, p. 109). Secondly, in a highly dynamic and fast-paced academic environment, first glance labels of credibility and excellence enable other researchers, institutions, and funders to assess the publication medium without going through the often long and incomprehensible editorial policy sections of the respective journal.

Every take on academic impact, excellence, and the wide dissemination of knowledge has to be weighed against questions of cultural dominance and a subtle neo-colonial agenda: the scientific lingua franca is (still) English, the open access discourse and roadmap is led by stakeholders in the Global North, and the reputational databases are based on standardisation, not diversity, and favour the journals and topics of the Global North tremendously (Alperin, 2013). There are some authors (Alperin, 2013; Memon, 2019; Seo, 2018) arguing that, for instance, the large databases and indices marginalise journals from ‘developing’ countries, applying criteria hard to fulfil without the means of a large academic institution, and therefore

³² In an activist fashion, the use of alternative metrics was promoted 2010 in the form of a manifesto (Priem et al., 2010).

³³ In the film and arts research community, many researchers criticise metrics more fundamentally, as they argue that these just take into account ways of measuring buzz on certain texts, when in fact research output can be images, art works, movies, etc.

³⁴ Lemke et al. (2019) identified key inhibitors for the adoption of altmetrics in social sciences. Some of their concerns—lack of time for and knowledge about metrics (p. 14)—can hopefully be addressed via this white paper.

calling for more regional or local initiatives.³⁵ Others advocate for a plurilingual publishing system and more ‘cognitive justice and greater fairness between the visible and accessible knowledge from the North and from the Global South’ (Piron, 2018, p. 5). Or they see a chance for empowerment through the thoughtful and contextualised use of ‘altmetrics’ (for ‘alternative scholars’; Alperin 2013, p. 19).

Indexing service	Description	Website
Directory of Open Access Journals (DOAJ)	DOAJ is a community-curated list providing metadata, information on publishing cost and awarding the DOAJ seal for best practice in OA publishing.	doaj.org/
European Reference Index for the Humanities (ERIH Plus)	ERIH Plus is a curated list of research journals by the European Science Foundation and currently hosted by the Norwegian Centre for Research Data; it aims at increasing the visibility of SSH.	dbh.nsd.uib.no/publisering/gskanaler/erihplus/
Scopus	Scopus is a subscription-based database by Elsevier covering mostly peer-reviewed literature from journals of all disciplines.	www.scopus.com/
Web of Science	Web of Science is a subscription-based service by Clarivate Analytics providing access to multiple databases with broad citation data across all disciplines.	www.webofknowledge.com

Table 2: Indexing services and referencing databases

4.4 Key steps

- Select a suitable open access repository (<https://v2.sherpa.ac.uk/opensoar/>) and set up a customised depositing agreement.
- Use open source identification systems for articles, authors, and institutions to increase outreach and minimise attribution error.
- Familiarise yourself with open and proprietary metrical systems in order to access and assess the aggregated data about your journal.
- Reach out to your (or friendly nearby) library and ask for support on issues related to long-term preservation, archiving, and bibliometrics.

4.5 Use case *Internet Policy Review*

Over the course of the InnOAccess research project, *Internet Policy Review* established or transitioned to a

³⁵ The arguments deal with these issues more broadly (see Alperin, 2013, p. 20; Memon, 2019, p. 1351; Seo, 2018). Yet there are more and more initiatives providing OA infrastructures, preprint servers, repositories, and digital archives for Central and South America (<http://amela.org/>), South Asia (Seo, 2018), and Africa (see the preprint server project ‘Africa Arxiv’ [<https://info.africarxiv.org/>]).

set of stable practices, workflows, and technical solutions for archiving, identification, indexing, and evaluating.

For long-term archiving of its scholarly materials, the journal entered into a partnership with the repository EconStor. Following an initial mass-submission of all existing articles in PDF form to EconStor, an automatic emailing system was integrated with the PDF generator (see chapter 3.4) in order to keep the repository up to date without having to rely on their sporadic, manual efforts. As a side effect, since Google Scholar also seems to index large repositories more frequently and comprehensively than small journals, this has also helped with the ranking problem articulated below.

While falling outside of the scope of traditional ‘archiving’ practices, the transition to Crossref and the concomitant improvements to the XML export workflow (see chapter 5.3) also enabled depositing of structured reference data with Crossref via the JATS upload. In this way, ‘scholarly metadata’ is given to the ecosystem for both immediate and long-term use.

In terms of identification, *Internet Policy Review* has now integrated with the two primary service providers in the scholarly publishing community: Crossref for scholarly output and ORCID for the scholars themselves. While a change in usage conditions at the previous DOI registration agency, DataCite (via dalra), necessitated the switch to Crossref, this transition to the *de facto* default for articles came with further benefits, including the aforementioned submission of reference data and also integration with the other leg of identification, ORCID, as well as advantages in metrics (discussed below).³⁶

When article DOIs are registered with Crossref via its JATS (or NLM) upload service, the system automatically credits the authors work on their ORCID pages, provided they have a profile at the service and have given the journal the ID. This automatic attribution encompasses the journal’s integration with ORCID. While membership of ORCID—which requires a significant technical investment—could effectively turn the data exchange with the service into a two-way street (e.g., allowing authors to log in to *Internet Policy Review*’s systems with their ORCID identifier and import their data from the service), the costs of such tighter integration outweigh the benefits. By ensuring that JATS-XML files contain ORCIDs (where provided), *Internet Policy Review* picked the most effective and significant pathway. ORCID identifiers are visually linked from author profiles as well and thus also visible from the journal frontend.

One aspect related to conventional indices is that of presence and ranking in Google Scholar, since it not only matters how often you are cited, but in what order articles are displayed. This search engine constitutes one of the primary discovery mechanisms for scholarly works. Therefore, the *Internet Policy Review* journal invested some efforts into boosting its rankings in Google Scholar—conforming to its Inclusion Guidelines for Webmasters³⁷ where possible, for instance, through the comprehensive adoption of the recommended metadata tags—yet numerous rounds of changes did not render observable improvements. Such difficulty in reliably being indexed by Google Scholar represents an infrastructural problem for small science journals. The search engine is one of the key gatekeepers in accessing scientific

³⁶ Benefitting from its institutional membership at Open Access Scholarly Publishers Association, *Internet Policy Review* can register up to 50 DOIs per year for free.

³⁷ <https://scholar.google.com/intl/en-us/scholar/inclusion.html>

research, yet this ‘black box’ character means that small journals may end up wasting valuable technical resources in an attempt to obtain uncertain payoffs. It should be noted, however, that more standardised systems such as OJS seem to fare better in this regard, benefitting from more widely distributed incentives to optimise the search engine.³⁸

In the case of *Internet Policy Review*, the strategy was that the editors (those closest to the content) would approach two types of databases: the curated journal databases such as DOAJ, ERIH Plus, and Free Journal Network, and the impact measurement databases, including Scopus and Clarivate Analytics (Web of Science Emerging Sources Citation Index). This happened over a period of four years. Practically, the applications to these indices are relatively simple and can be done within a few hours. Scopus, for example, has a straightforward application submission system made up of a simple form to fill out and where the applicant can track the evaluation evolution. Some indices require specific anti-plagiarism policies, a minimum number of peer-reviewed articles per year, and an authoritative/international editorial board. The requirements are generally useful and provide, in addition to resources like the COPE guidelines,³⁹ guidance to early-stage journals in figuring out governance and quality-insurance policies. The Scopus and Clarivate Analytics applications took about nine months to get through. The ERIH Plus evaluation was quicker. It is advisable to apply to the DOAJ early, and to the other indices once a journal has reached a certain cruising speed. In the case of the larger indices, journal editors need to select the sub-index to which they want to submit. For instance, it was impossible for *Internet Policy Review* to directly land in the main Web of Science index. The sub-index ESCI was therefore selected. After some 2–3 years in that sub-index, a journal might get a positive evaluation and enter the main index. This is determined internally and is mainly based on citation impact factors.

In this context, altmetrics constitute another area in which *Internet Policy Review* benefits from Crossref integration. Whereas the journal had previously attempted to track metrics with its own setup of the Lagotto article-level metrics software (<https://github.com/lagotto/lagotto>), visualised through almviz (<https://github.com/jalperin/almviz>), both the technical maintenance work associated with this approach and as well as changes in source APIs made this untenable, at least for a small journal team. The switch in DOI registration agency enabled collaboration with the ROSI project’s development of the ImpactViz prototype (<https://github.com/TIBHannover/impactViz>). While embedding this prototype did require some technical customisation, it represents a marked improvement over the previous setup. The project’s sought-after goal of easy customisability in metrics sources and display while offloading much of the technical debt aligns well with *Internet Policy Review*—and may be appropriate for small science journals in general. The prototype lists four dimensions of impact, visualised as clickable icons. The *scientific impact* is classically determined by citation rate; the *societal impact* draws from social media metrics; the *scholarly community impact* measures the level of conversation about an article in academic fora. The fourth dimension is the *open access* one, which mainly provides context on the specific open access strategy of the journal. Although these categories are still crude, *Internet Policy Review* is working with the project to

³⁸ This attempt is often referred to as ‘academic search engine optimisation’ or ASEO (See Rovira et al., 2019). PKP provided OJS specific guidelines on this: <https://docs.pkp.sfu.ca/getting-found-staying-found/en/getting-found-appendix-2-google-scholar>

³⁹ Provided by the Committee on Publication Ethics: guidelines (<https://publicationethics.org/guidance/Guidelines>)

further reach a level of visualisation that paints a convincing picture of an article's context.

5 Managing and documenting workflows: Open source (journal) systems

High quality academic publishing requires managing complex workflows, usually starting with the abstract or article submission, tracking the submission, managing the review process, recording versions, providing copy-editing and language checks, and eventually publishing in a format of choice including appropriate metadata and other measures of dissemination. Added to that, especially in academic publishing, the task requires oversight and management of a diverse editorial team in multiple geographical locations and often without any prior experience in journal editing or publication technology. To ensure a smooth workflow and consistent publication output, every journal needs to implement a procedure for acquiring new editors and replacing parting ones, including thoughtful succession planning for managing editors.

5.1 Editorial management systems as tools of sustainability

To take on all of these tasks in a sustainable way, the use of editorial management systems is essential. The publishing sector provides a large number of proprietary and non-proprietary solutions, some of them as all-in-one platforms and others addressing certain stages of the editorial process. Given the increased financial pressure open access journals are exposed to, these kinds of platforms and tools need to be free of charge and preferably open source. Open access initiatives like the 'Radical Open Access Collective' provide a non-comprehensive list, which includes leading open source solutions like Open Monographs Press, Open Journal Systems, and Janeway.⁴⁰ Comprehensive reports (Cyzyk and Choudhury, 2008) and guidelines on how to set up and use open access publishing solutions (Beringer and Arning, 2020; with regards to requirements of Plan S and DINI) shed further light on the multitude of options. This white paper limits itself to the community-based Open Journal Systems (OJS), developed and maintained by the non-profit research initiative 'Public Knowledge Project' (PKP; <https://pkp.sfu.ca/>). Systems like OJS revitalise and energise scholar-led publishing and facilitate a substantial change in scholarly communication by defying major publishing houses and large society publishers (Edgar and Willinsky, 2010).

Open Journal Systems⁴¹ provides a platform to manage and publish journal articles based upon issues or annual volumes. With its three core areas—frontend to access published content, backend to manage the editorial workflow, database to administrate and export metadata—it provides a collaborative, multilingual, and highly customisable work environment for editorial teams. An added communication feature allows outreach to the registered readers by publishing notifications and further news. In the end, '[t]he goal is to reduce the energy spent on finding email addresses, tracking manuscripts and preparing emails, so that more time can be spent, especially by editors, on working with authors to improve the quality of the articles' (Willinsky and Mendis, 2007). PKP and other institutions offer fee-based hosting services, but many university libraries and data centres cover those services for their members without charge.⁴² A large

⁴⁰ The commented list was last updated in May 2019, yet remains a valuable source for OA publishing tools (<https://radicaloa.disruptivemedia.org.uk/resources/publishing-tools/>; last retrieved Sept. 14, 2020).

⁴¹ The OJS main page offer basic information (<https://pkp.sfu.ca/ojs/>), extensive documentation for editors, IT administrators, developers, and others are available in the OJS handbook (<https://docs.pkp.sfu.ca/#appojs3>).

⁴² Every journal is advised to check for institutional resources in order to reduce costs. Comprehensive information on the state of OJS in Germany is provided by the OJS-de network (<https://ojs-de.net/>).

and active community continuously works on further tools and plugins in order to comply with the ever-changing standards of professional scholarly publishing. They encourage reuse, transparency, and cooperation in this regard by publishing code and recent developments via GitHub (<https://github.com/pkp/ojs>). Despite the many advantages of OJS, using such standardised publishing solutions or platforms sometimes conflicts with the demands of publishing projects—for instance, by requiring other output formats (e.g., videos) or design elements in the frontend. Especially for already existing journals, the transformation or switch to OJS can be difficult because individual workflows, tools, and journals archives might not easily be migrated to the OJS infrastructure. In the end, journals should rise to these challenges, since OJS or similar infrastructures provide publication services and plugins that are essential for professional and sustainable publishing, but often beyond what academic editors are able (or willing) to learn. Compared to that, gaining a basic understanding of a structured and thoroughly documented system like OJS is relatively easy and at the same time enables scholars to participate and advance in innovative publishing environments.

5.2 Key steps

- Evaluate your editorial workflow and check available open source publishing tools for a match.
- Consider standardised over individual solutions: maintenance and succession planning are much easier that way.

5.3 Use case *Internet Policy Review*

As a young but relatively established open access journal, *Internet Policy Review* with its Drupal-based content management system was able to integrate OJS in a way that makes its editorial workflow for authors, editors, and reviewers more transparent, cooperative, and better documented, while keeping the Drupal frontend, its branding, and therefore user experience the same.

While, over the course of the project, *Internet Policy Review* was able to augment and streamline its workflow, many of the components and decisions stem from long-term experiences or commitments made early on. Thus, for instance, the open peer review process has always made Google Docs a natural fit for the journal, allowing simultaneous editing and organic conversations to develop directly in the article text. Drupal as the technical infrastructure meant that the journal could evolve into its present state naturally; it furthermore allows for a great deal of flexibility with presentation formats even as it incurs a technical cost. In light of these more flexible solutions, the recent decision to migrate the editorial ‘backend’ work—the administration, automation, and tracking of the submissions process—to OJS represents an effort to lighten the editorial burden of the increased number of submissions, improve the user experience for authors and readers, and enhance transparency of the entire editorial workflow. In a similar vein, the journal has automated aspects of the publication process as well. The resulting workflow is a patchwork of established solutions and custom scripts:

- **Authoring:** application agnostic
- **Submission [OJS]:** Authors submit papers through Open Journal Systems. This provides the editorial team with an overview of the journal work and clarifies the steps for authors.
- **Review & Revision [OJS/Google Docs]:**

- Open Journal Systems and Google Docs constitute the two legs of *Internet Policy Review's* editorial management system. OJS takes work off of the editors' hands by tracking the submissions; at the same time, its establishment in the broader scholarly publishing environment means that there is a broadly established familiarity with its interface and workflow for authors. This saves the editors time, removing the overhead of manually tracking the status of submissions and articles-in-progress through their emails. It also disambiguates the process for authors by reminding them of the necessary requirements at each step of the way.
- Where OJS does fall short, though, is in parallel work on one document. Here, *Internet Policy Review's* experience with running its 'open peer review' process on Google Docs makes up for the deficits of this system. By allowing reviewers and authors to interact on the document, Google Docs complements OJS's structure.
- **Publishing [scripts]:**
 - Before a finished article is placed in the Drupal content management system, a custom script automates much of the conformance to *Internet Policy Review's* formatting guidelines. This process removes a good chunk of the work involved in the final publication process, converting the .docx files exported from Google Docs into HTML. The script uses Mammoth.js to convert the .docx files to .html and Cheerio.js to manipulate the resulting HTML.
 - In addition to this human-readable version, the journal also uses JATS, an XML standard for describing journal articles. By offering a uniform and interchangeable description set for journal articles' metadata and content, it promotes both compatibility between different publishing software packages as well as archiving efforts. *Internet Policy Review* began adopting JATS in a previous effort, but in the course of the InnOAccess project it was able to streamline both the workflow and aspects of the technical setup.
 - Some of the aforementioned infrastructural improvements, such as the submission of ORCID IDs as well as the depositing of machine-readable references, depend on the 'universal language' of JATS.
 - The depositing of machine-readable references via their description in the JATS file corresponding to each article merits special mention.
 - Both manually correcting author-submitted bibliographies and parsing such plain-text references into a machine-readable form constitute labour-intensive tasks. Indeed, the former has been a consistent bottleneck in the article publication process for *Internet Policy Review*, and the addition of the latter—while evincing significant benefits for the ecosystem—threatened to increase this temporal burden.
 - Within the context of the InnOAccess project, *Internet Policy Review* set up a process that at once combines and streamlines both of these tasks. A blog post by the editors of *Entangled Religions* (<https://belter.hypotheses.org/45>) provided an important impetus in this regard (see table 3):
 - First, Anystyle, an open source web app (as well as a command-line

- program and Ruby library) parses the submitted plain text references
 - Then, the resulting structured references are imported into Zotero, an open source reference manager
 - In Zotero, both manual and automatic correction occur
 - The finalised, correct references are exported twice:
 - a. once in APA 7, for the HTML and PDF versions of the articles provided by *Internet Policy Review*;
 - b. once in JATS XML format, via an improved version of JATS.csl
 - These are both imported into the version of the article on Drupal. The custom Drupal JATS module combines the JATS references with the remainder of the article content and metadata. This can then be submitted to e.g., Crossref.
- **Community building:**
 - In order to credit the integral work that reviewers do for the journal and its community, *Internet Policy Review* began crediting them on its website in 2016. The task of compiling this list, maintaining its accuracy, and asking for consent to display this information previously relied on a Google Apps script tied to a Podio database; this stopgap solution still involved a lot of manual curation. Integrating this process as a checkbox in OJS allows the journal to do it with far fewer moving parts.

Anystyle	A webapp, command-line application, and Ruby library that parses plain-text bibliographies into machine-readable structured bibliographic formats.	https://anystyle.io
Zotero	An open source reference manager	https://zotero.org
[Zotero plugin] DOI Manager	A plugin for Zotero that fetches DOIs for articles, thus aiding in the final submission of complete and unambiguous metadata.	https://github.com/bwiernik/zotero-shortdoi
[Zotero service] Internet Policy Review shared library	A shared Zotero library opens references to further research in a more convenient and accessible way	https://www.zotero.org/groups/2535817/internet_policy_review/library
JATS.csl	A citation style language description of the Journal Article Tag Suite's description of references. Based on Martin Fenner and Martin Paul Eve's work.	https://github.com/InternetPolicyReview/JATS-csl
Drupal module	A custom Drupal module that converts <i>Internet Policy Review</i> 's HTML conventions to JATS. It also integrates the JATS references derived from the above workflow.	

Table 3: Internet Policy Review's JATS and structured reference stack

As mentioned, this workflow combines a patchwork of established solutions and programs with some custom hinges. Other offerings, such as those of Redalyc, seem to have evolved in parallel and cover similar but slightly divergent workflows.

Ultimately, the holy grail of an efficient scholarly publishing workflow—with a tight integration between authoring, reviewing, editing, formatting, and publishing—would involve a collaborative, (JATS-)XML-based WYSIWYG editor. Substance's Texture constituted one attempt to produce such a piece of software. However, it seems that this project—backed by a few large publishing outfits—stopped development due to diverging needs. Yet, properly backed, such an application—perhaps organised and developed in a modular and open source manner in the spirit of Open Journal Systems, thus integrating with various aspects of the OA environment—would be a significant boon for scholarly publishing as a whole, but especially small science endeavours in particular: it would absorb significant aspects of technical and procedural overheads and ease labour-intensive bottlenecks in the workflow while contributing to (re)usable and portable data in the ecosystem.⁴³

6 Recommendations for sustainable open access publishing

There is a wide array of literature, guidelines, handbooks, and even white papers on workable methods for state-of-the-art open access publishing. Given the precarious nature of many scholar-led journals, the emergence of new yet less scalable diamond open access models and a confusing number of networks and initiatives, this white paper aimed at shedding light on a few issues that constitute sustainable publishing—that is enabling journals to publish constantly, with forethought, agilely, and efficiently.

As pointed out in the very beginning, there is an inherent connection between publications technologies and business models, especially when including the community effort and extensive networks that have always been a strength of the open access movement. Using these collaborative efforts and support of the community, we would like to outline five core elements where we feel sustainable publication services might be connected to and encourage open access journals to consider them.

6.1 Streamlining workflows with professional publication software

Journal editing and publishing is a multi-level, complex process involving stakeholders in many places. Using professional and dedicated open source publication software (e.g., OJS or Janeway) to manage these processes is key for stable and reliable workflows. The software fosters cooperation because it allows remote work; it empowers users, since processes and steps are kept transparent and therefore easy to learn and digest; and it provides the journal with a structured editorial memory, in that way enabling effective succession planning.

6.2 Displaying and contextualising altmetrics to showcase bibliometric diversity

What is counted counts! Despite profound critique, open access journals should still familiarise themselves with aspects of citation counts and databases aggregating those information (at least regarding the

⁴³ See a project cooperation of the Technische Universität Hamburg (TUHH) and the Staats- und Universitätsbibliothek (SUB) on developing an OJS/XML workflow (<https://oa-pub.hos.tuhh.de/en/>).

aforementioned journal-level metrics). As the industry standard of calculating the ‘impact’ of research and evaluating the degree of professionalism, it is still a valuable asset in order to sustain a journal’s future. Yet, scholar-led open access journals can use their relative independence to push for the use of altmetrics and take advantage of the opportunities of electronic publishing: defying the ‘citation lag’ of standard metrics, diamond open access can include community-based solutions and publication outlets as relevant factors of assessing research quality and outreach (e.g., as ‘sharing factor’). For many this is considered to be a building block for a more biblio-diverse culture of publishing.

6.3 Increasing impact through repository-based archiving and indexing

Another essential aspect of professional open access publishing is achieved by repository-based archiving and indexing. Using either institutional or disciplinary resources at hand and depositing all relevant output formats in a state-of-the-art repository ensures permanent availability, the assignment of persistent identifiers, and the distribution of all relevant metadata (e.g., via CrossRef standards and APIs). Additionally, journals should make the effort to research and apply for (community-)curated lists, which increase their visibility and make it easier for potential contributors to select a high-quality outlet.

6.4 Providing accessible, single-source output formats

The open access movement is dedicated to providing universal access to research literature, which should include barrier-free websites and output formats. Therefore, journal publishers need to educate themselves on basic accessibility requirements (web accessibility, PDF accessibility, markup format) or seek out trusted accessibility organisations and initiatives that provide assistance in this regard. Again, there is an active community in which journals should consider participating; in this context, they can help develop new modes of publishing, e.g., by offering developers the journal infrastructure as a ‘playground’ or first phase implementation stage for more timely single-source output formats.

6.5 Identifying authors and institutions via open databases and standards

Keeping adequate track of author identifications and the institutions they are affiliated with is meticulous work, even more so since the academic research community is global and their records are manifold. Instead of solely relying on correctly filled-out author profiles, journals should consider using open source identification services as it seems to be the most pragmatic thing to do. And while the measures we described in this white paper may, at first glance, seem to add more work, a fully implemented conceptual framework for author and institutional IDs in the journal’s editorial system does save a lot of time and effort in the long run. Established consortia and initiatives like ORCID and the ROR database offer support on many levels (from API description to advocacy guidelines) to build upon and make their approaches worthwhile.

7 Conclusion

The core endeavour of this white paper—outlined and exemplified in the individual sections—was to provide a summarising guide for scholar-led open access journals and assist them in making informed decisions about better practice in scholarly publishing. Focusing on publication technologies and networking opportunities, we wanted to highlight how this independent part of the publishing sector can

be sustained and empowered while reducing costs and fostering innovative workflows. A rich set of community-based know-how and non-profit cooperative practices (e.g., on GitHub, the PKP forum, the ROAC website etc.) enables journals these days not only to publish meaningful content but also to contribute to the free distribution of knowledge, with transparent opportunities about reuse and a culture of sharing—both about what and how we publish.

8 References

- Akbaritabar, A., & Stahlschmidt, S. (2019). Merits and Limits: Applying open data to monitor open access publications in bibliometric databases. *Digital Libraries*, arXiv:1902.03937 [cs.DL]. <https://arxiv.org/abs/1902.03937v2>
- Alperin, J. P. (2013). Ask not what altmetrics can do for you, but what altmetrics can do for developing countries. *Bulletin of the American Society for Information Science and Technology*, 39(4), 18–21. <https://doi.org/10.1002/bult.2013.1720390407>
- Andrae, M., Blumesberger, S., Edler, S., Ernst, J., Fiedler, S., Haslinger, D., Neustätter, G., & Trieb, D. (2020). Barrierefreiheit für Repositorien. Ein Überblick über technische und rechtliche Voraussetzungen. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, 73(2), 259–277. <https://doi.org/10.31263/voebm.v73i2.3640>
- Beringer, C., & Arning, U. (2020). *Guide for establishing Gold and Green Open Access Publishing Platforms*. Köln: ZB MED – Information Centre for Life Sciences. <http://dx.doi.org/10.4126/FRL01-006421133>
- Bierman, J. (2019). Discovering Open Access Engineering Journals: An Examination of Indexing Rates for DOAJ and Non-DOAJ Content. *Issues in Science and Technology Librarianship*, Spring(91). <https://doi.org/10.29173/ist1>
- Björk, B.-C. (2017). Scholarly journal publishing in transition- from restricted to open access. *Electronic Markets*, 27, 101–109. <https://doi.org/10.1007/s12525-017-0249-2>
- Björk, B.-C., Kanto-Karvonen, S., & Harviainen, J. T. (2020). How Frequently Are Articles in Predatory Open Access Journals Cited. *Publications*, 8(2), 17. <https://doi.org/10.3390/publications8020017>
- Björk, B.-C., & Korkeamäki, T. (2020). Adoption of the open access business model in scientific journal publishing: A cross-disciplinary study. *Digital Libraries*, arXiv:2005.01008 [cs.DL]. <https://arxiv.org/abs/2005.01008>
- Czyzyk, M., & Choudhury, S. (2008). *A Survey and Evaluation of Open-Source Electronic Publishing Systems*. Collections: Sheridan Libraries Staff Research. <http://jhir.library.jhu.edu/handle/1774.2/32737>
- Dorta-González, P., & Santana-Jiménez, Y. (2018). Prevalence and citation advantage of gold open access in the subject areas of the Scopus database. *Research Evaluation*, 27(1), 1–15. <https://doi.org/10.1093/reseval/rvx035>
- Edgar, B. D., & Willinsky, J. (2010). A Survey of the Scholarly Journals Using Open Journal Systems. *OJS På Dansk*, 1(1). <https://doi.org/10.7146/ojssb.v1i1.2707>
- Fecher, B., & Friesike, S. (2013, May). Open Science: One Term, Five Schools of Thought. *RatSWD Working Paper Series*. <http://dx.doi.org/10.2139/ssrn.2272036>
- Foley, M. J., & Kochalko, D. L. (2010, November 3–6). *Open Researcher and Contributor Identification (ORCID)*. Proceedings of the Charleston Library Conference, Charleston, South Carolina, United States. <http://dx.doi.org/10.5703/1288284314850>
- Green, T. (2019). Is open access affordable? Why current models do not work and why we need internet-era transformation of scholarly communications. *Learned Publishing*, 32(1), 13–25. <https://doi.org/10.1002/leap.1219>
- King, M., Pegrum, M., & Forsey, M. (2018). MOOCs and OER in the Global South: Problems and Potential. *The International Review of Research in Open and Distributed Learning*, 19(5). <https://doi.org/10.19173/irrodl.v19i5.3742>
- Kuykendall, H. (2017, July). The Practical Role of Automated Web Accessibility Testing Tools. *Digital Accessibility Digest*. <https://www.microassist.com/digital-accessibility/role-web-accessibility-testing-tools/> (Accessed 10 September 2020).

- Kwok, R. (2013). Altmetrics make their mark. *Nature*, 50, 491–493. <https://www.nature.com/articles/nj7463-491a.pdf?origin=ppub>
- Larivière V., Haustein S., & Mongeon P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE*, 10(6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>
- Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2019). ‘When You Use Social Media You Are Not Working’: Barriers for the Use of Metrics in Social Sciences. *Frontiers in Research Metrics and Analytics*, 3(39). <https://doi.org/10.3389/frma.2018.00039>
- Lewthwaite, S. (2014). Web accessibility standards and disability: developing critical perspectives on accessibility. *Disability and Rehabilitation*, 36(16), 1375–1383. <https://doi.org/10.3109/09638288.2014.938178>
- Marchitelli, A., Galimberti, P., Bollini, A., & Mitchell, D. (2017). Improvement of editorial quality of journals indexed in DOAJ: a data analysis. *Italian Journal of Library, Archives and Information Science*, 8(1), 1–21. <http://dx.doi.org/10.4403/jlis.it-12052> <https://www.jlis.it/article/view/12052>
- Memon, A. R., & Azim, M. E. (2019). Open Researcher and Contributor Identifier and other author identifiers: Perspective from Pakistan. *Journal of the Pakistan Medical Association*, 69(6), 888–891. <https://europepmc.org/article/med/31201398>
- Memon, A. R. (2019). Scholarly publishing and research dissemination in South Asia: some exemplary initiatives and the way forward. *Journal of the Pakistan Medical Association*, 69(9), 1348–1353. <https://pubmed.ncbi.nlm.nih.gov/31511723/>
- Mering, M. (2017). Correctly Linking Researchers to Their Journal Articles: An Overview of Unique Author Identifiers. *Serials Review*, 43(3–4), 265–267. <https://doi.org/10.1080/00987913.2017.1386056>
- Mirčić, M., & Đokić, Z. (2018). Trends in academic media – open researcher and contributor identifier, concept and significance for the scientific research community. *Medicinski Pregled*, 71(3–4), 83–87. <https://doi.org/10.2298/MPNS1804083M>
- Moore, S. A. (2020). Revisiting ‘the 1990s debutante’: Scholar-led publishing and the prehistory of the open access movement. *JASIST*, 71(7), 856–866. <https://doi.org/10.1002/asi.24306>
- Narayan, B., & Luca, E. (2017). Issues and challenges in researchers’ adoption of open access and institutional repositories: a contextual study of a university repository. *Information Research: an international electronic journal*, 22(4). <http://hdl.handle.net/10453/121438>
- Neylon, C. (2020): Research excellence is a neo-colonial agenda (and what might be done about it). In E. Kraemer-Mbula, R. Tijssen, M. Wallace & R. McLean (Eds.), *Transforming Research Excellence* (pp. 92–118). Cape Town: African Minds.
- Olijhoek, T., Mitchell, D., & Bjørnshauge, L. (2015). Criteria for open access and publishing. *ScienceOpen Research*, 1–8. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AMHUHV.v1>
- Pierson, C. A. (2016). ORCID: Not a flower but an Open Researcher and Contributor ID, *Journal of the American Association of Nurse Practitioners*, 28(3), 124. <https://doi.org/10.1002/2327-6924.12356>
- Piron, F. (2018). Postcolonial Open Access. In U. Herb & J. Schöpfel (Eds.), *Open Divide: Critical Studies on Open Access*. Sacramento, California: Litwin Books.
- Ponte, D., Mierzejewska, B. I., & Klein, S. (2017). The transformation of the academic publishing market: multiple perspectives on innovation. *Electronic Markets*, 27, 97–100. <https://doi.org/10.1007/s12525-017-0250-9>
- Priem, J., Taraborelli, D., & Groth, P. (2010, October). *Altmetrics: A manifesto*. <http://altmetrics.org/manifesto>
- Rosa, C. A., Craveiro, O., & Domingues, P. (2017). Open Source Software for Digital Preservation Repositories: a Survey. *International Journal of Computer Science & Engineering Survey (IJCES)*, 8(3), 21–39. <https://arxiv.org/abs/1707.06336>
- Rothberg, M. (2018). Publishing with accessibility standards from the inside out. *Learned Publishing*, 31(1), 45–47. <https://doi.org/10.1002/leap.1149>
- Rovira, C., Codina, L., Guerrero-Solé, F., & Lopezosa, C. (2019). Ranking by Relevance and Citation Counts, a Comparative Study: Google Scholar, Microsoft Academic, WoS and Scopus. *Future Internet*, 11(9), 202.

<https://doi.org/10.3390/fi11090202>

- Schallaböck, J., & von Grafenstein, M. (2017). *ORCID aus datenschutzrechtlicher Sicht: 'Gutachten im Auftrag des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projektes ORCID DE zur Förderung der Open Researcher and Contributor ID in Deutschland'*. Berlin: iRights.Law Rechtsanwälte. <http://doi.org/10.2312/lis.17.02>
- Seo, T.-S. (2018). Open access full-text databases in Asian countries. *Science Editing*, 5(1), 26–31. <https://doi.org/10.6087/kcse.114>
- Suber, P. (2012). *Open access*. Cambridge: MIT Press.
- Waecker, E., Fulkerson, M., Power, J., & Ku, J. (2019). From Content Creation to Content Delivery: Partnering to Improve E-Book Accessibility. *The Serials Librarian*, 76(1–4), 147–155. <https://doi.org/10.1080/0361526X.2019.1565512>
- Willinsky, J., & Mendis, R. (2007). Open access on a zero budget: a case study of Postcolonial Text. *Information Research* 12(3), Paper 308. <http://informationr.net/ir/12-3/paper308.html>
- Zhu, Y. (2017). Who supports open access publishing? Gender, discipline, seniority and other factors associated with academics' OA practice. *Scientometrics*, 111, 557–579. <https://doi.org/10.1007/s11192-017-2316-z>