



open scholarly communication in the european
research area for social sciences and humanities

OPERAS SIG on Tools for Open Scholarly Communication

White Paper 2021



By Lorenzo Armando, Céline Barthonnat,
Emilie Blotière, Maddalena Briganti,
Luca De Santis, Arnaud Gingold, James
MacGregor, François-Xavier Mas,
Alessandro Pierno, Bernard Pochet, Aleš
Pogačnik, Riccardo Pozzo, Nikola Stanić,
Agnieszka Szulińska

This is the White Paper of the
Tools Special Interest Group.

Starting date: 01.10.2021

Delivered: 18.06.2021

Participants:

CNRS: Céline Barthonnat, Emilie Blotière, Arnaud Gingold,
François-Xavier Mas

CEON/CEES: Nikola Stanić

Univ. Firenze: Alessandro Pierno

IBL-PAN: Agnieszka Szulińska

Lexis: Lorenzo Armando, Maddalena Briganti

Liège University Library: Bernard Pochet

Net7: Luca De Santis

PKP: James MacGregor

Università di Roma Tor Vergata: Riccardo Pozzo

ZRC SAZU: Aleš Pogačnik

The White paper last version has been kindly reviewed and
enriched by:

Jeroen Bosman (Utrecht Univ.)

Bianca Kramer (Utrecht Univ.)

This project has received funding from the
European Union's Horizon 2020 research
and innovation programme under grant
agreement
No 871069



<https://doi.org/10.5281/zenodo.5654319>

Attribution 4.0 International



CONTENT

| | |
|--|-----------|
| SUMMARY | 6 |
| DECLARATION OF INTEREST | 6 |
| INTRODUCTION | 7 |
| LANDSCAPE | 7 |
| SCOPE AND OBJECTIVES | 9 |
| USERS AND USAGES | 10 |
| Researchers | 10 |
| Editors | 11 |
| Providers of publishing services | 11 |
| DEFINITIONS AND CRITERIA | 12 |
| Definitions | 12 |
| Integrated vs independent tools | 12 |
| Open vs closed source | 12 |
| Interoperability | 13 |
| How to select the appropriate tool? | 13 |
| Method | 13 |
| Criteria | 14 |
| AUTHORING | 16 |
| Towards online and collaborative authoring tools | 16 |
| Structured formats in open source tools | 17 |
| Markdown & LaTeX | 17 |
| XML | 18 |
| Proprietary software as a service | 18 |
| Companion tools for authoring | 19 |
| CERTIFICATION | 19 |
| Anti-plagiarism tools | 20 |
| Peer-reviewing overview | 21 |
| Open peer review | 22 |
| Tools for open peer review | 23 |
| Anonymised peer review | 23 |
| Peer review tracking | 24 |
| PUBLISHING | 24 |
| Features of a publishing platform | 25 |
| Software overview | 25 |
| List of software | 26 |
| Integration with third party services | 27 |

| | |
|---|-----------|
| COMMUNICATING | 29 |
| SCHOLARLY PUBLISHING MAIN TRENDS | 30 |
| Preprints | 30 |
| Artificial Intelligence | 31 |
| Data papers | 32 |
| Information: User-centric Criteria | 32 |
| Sustainability: A Tools' Observatory | 33 |
| Training: A Training Material Catalogue | 33 |
| Method: Workflows Guidelines | 33 |
| Community infrastructure: Bringing the pieces together | 34 |
| ANNEXES | 36 |
| Annex I. List of tools mentioned in this paper | 36 |
| Annex II. List of criteria and features for analyzing tools | 39 |
| Basic criteria for all | 39 |
| Authoring tools additional criteria | 40 |
| Publishing tools additional criteria | 40 |
| Annex III. Analytical table of annotation tools | 42 |
| Annex IV. Training material draft | 54 |
| REFERENCES | 55 |

1. SUMMARY

This white paper is the output of the OPERAS Special Interest Group (SIG) Tools and R&D for scholarly communication; it is an updated version of a previous 2018 white paper¹. With a focus on scholarly publishing tools, the objectives of the SIG Tools are to: provide a landscape analysis, identify emerging trends, and list the areas of potential improvements, developments, and collaborations. Since 2018, various studies and initiatives confirmed the necessity to both coordinate the developments of tools and provide guidance to the users. Similarly, OPERAS emphasizes the importance of building the open science scholarly communication infrastructure in Social Sciences and Humanities on community driven tools. The white paper brings information on the existing tools for scholarly publishing, as well as recommendations that will support the building of such an open scholarly communication infrastructure.

The paper first examines tools types, definitions, and criteria that are able to facilitate their description and selection. The tools are then analyzed according to publishing main functions. For authoring, the development of online and collaborative tools represents an interesting perspective, especially when relying on structured formats, but also increases the risk of lock-in within multi-functional proprietary services. In peer reviewing, alongside widely used commercial tools, open peer review represents an innovative area, both in terms of usage and tools. Open source tools for publishing already offer a high level of service, but face interoperability challenges with the integration of an increasing variety of third-party services. A specific section is dedicated to communicating tools allowing for comments and annotations, as such function is transversal to the others.

To complement this description, the SIG tools also identified major trends that should impact the future of scholarly communication, namely: preprint servers, artificial intelligence, data papers, and user-centric developments. In conclusion, the white paper provides a list of recommendations able to address the challenges identified and to provide building blocks for the envisioned open scholarly infrastructure. The recommendations suggest: to establish user-centric criteria for tools, a tools' observatory, a set of training materials, guidelines about publishing workflows, and collaborations with other community initiatives.

2. DECLARATION OF INTEREST

Some co-authors of this OPERAS SIG Tools are members of organizations either developers of tools or providers of services described in this white paper. The collaborative writing helped to ensure an equal description of all the listed tools. However, for the sake of transparency, it is brought to the attention of the readers that some co-authors of the paper belong to:

- OpenEdition, which is the developer of the open source software Lodel and Bilbo;

¹ OPERAS Tools Research and Development White Paper, July 2018: <https://zenodo.org/record/1324110>.

- Public Knowledge Project (PKP), which is the developer of the open source software OJS, OPS and OMP;
- Net7, which is the developer of the open source software Pundit and Muruca.

3. INTRODUCTION

The digital shift has impacted the research environment under many aspects, and this is particularly true under the technological aspect of the scholarly publishing tools. The changes can be direct, through the development of software allowing for digital authoring or publishing, or indirect, through the transition from article-based to data-centric research, which implies at least a redefinition of the scholarly publishing concept. The new potentialities opened by the digital shift appear clearly in the current landscape of scholarly publishing tools: there is a variety of tools for a variety of finalities available to a variety of users. Such richness fully exploits the flexibility of the virtual environment and the networking capabilities of a connected infrastructure. It appears as much clearly that this landscape requires guidance for the users, be they researchers, publishers, or research support actors. Considering more precisely the context of Social Sciences and Humanities (SSH), experienced with well rooted practices related to text publishing, in journals and in monographs, some specific support and clarification is required. Furthermore, if we can consider that the infrastructure in its narrow technological meaning is already achieved, through cables, data-centers, and satellites, the infrastructure in the broader meaning of a coherent social environment supported by technology has not yet naturally arisen from the multiple initiatives come to light. With these two aspects in mind—the guidance and the infrastructure to invent,—the OPERAS Research Infrastructure (RI) has set up in 2017 a working group dedicated to the tools for scholarly publishing in the SSH. The group released a first version of its outcomes in 2018; the present work represents a second version established during 2021 by the OPERAS Special Interest Group (SIG) on tools and the related R&D. Rather than an entirely new version of the former work, the group concentrated its efforts on the landscape analysis and the final recommendations, in order to add more context and provide more concrete leads for potential actions.

3.1. LANDSCAPE

After the [OPERAS Tools Research and Development White Paper](#) (2018), important publications and reports addressed the scholarly publishing current context, with slightly different focuses.

In 2019, a collaboration between the Educopia Institute and the Invest in Open Infrastructure Initiative (IOI) produced a report called [Mapping the Scholarly Communication Landscape](#), based on the survey of over 40 Scholarly Communication Resources (SRCs), including tools, services, and platforms. Stressing the need for clearer strategy policies and shared best practices, the report also recommends providing guidance to the SRCs about software development and maintenance. Significantly, it also recommends establishing a taxonomy for all the publishing functions characterizing the SRCs' services.

The same year, a project from the MIT press involving the Canadian Institute for Studies in Publishing led to the publication of [Mind the Gap](#), a report analyzing extensively the landscape of open source publishing tools. Together with a description and classification of open source tools, the

report questions the sustainability of the existing environment and examines paths towards a robust and consistent “open infrastructure” that would be an alternative to proprietary services.

At the European level, a European Commission’s expert report was published in 2019 about the [Future of Scholarly Publishing and Scholarly Communication](#). With attention, this time, not only to SRCs or developers, but to the whole research ecosystem, the report indicates ways to consolidate the open science potentialities. Alongside with major considerations around the rewarding system linked to publications, the report also acknowledges the contrast between a persistent wide use of traditional formats like PDF and the evolution of the “scholarly publishing” concept towards more innovative practices.

Based on a survey of European Open Science Infrastructures (OSIs), the 2020 report from SPARC Europe [Scoping the Open Science Infrastructure Landscape in Europe](#) also addressed widely the research ecosystem, analyzing more specifically how OSIs facilitate open science, and rely also themselves on openness. Highlighting the strength of these OSIs and their commitment to open science, the report also notes that they face challenges in fully adopting open standards and open source software. Recommending more support and coordination, the report also indicates that the European OSIs actually operate at an international level, which implies to consider the open science global infrastructure at a broader level.

With a narrower but complementary scope, a report was issued at the beginning of 2021 about the “diamond” journals, i.e. journals free for the user and the author: the [OA Diamond Journals Study. References Library](#). Commissioned by CoalitionS and written by a consortium of organizations including OPERAS, the report identified a wide range of often small and scattered publishers that would benefit from coordinated support. On the technical side, this support would seek to enhance the flexibility of open source publishing tools, and to provide guidance through the creation of a Capacity Center for diamond journals.

Through all these works, it appears that the current environment is characterized by sustainability issues, in terms of funding but also in terms of stability, an environment where competition often acts in place of coordination. Options for a better sustainability may vary from one report to the other, but they all confirm that coordination through clarification, monitoring, and guidance would help to have an actual and functional open infrastructure. Considering the titles and the topics, we also see that the reports address, often indirectly, but sometimes also directly, the shift from scholarly publishing to scholarly communication. Although the starting point for publishing tools is often still the traditional publishing workflow, the main trend for present developments is the integration of new functions and new objects in the scholarly publishing workflow, which precisely defines a broader scholarly communication. Finally, the pace of publications on the subject itself is indicative not only of a shared observation, but also of a need for dedicated actions. The work of the SIG Tools attempts to take these three aspects into account in formulating its recommendations (see [Recommendations](#) section).

3.2. SCOPE AND OBJECTIVES

The first outcome of the SIG was a white paper published in 2018. This second edition contains more synthetical additions, especially a list of recommendations aiming at translating the SIG's work into concrete actions and projects.

In order to provide a thorough analysis of tools based on the expertise of the SIG members, the scope of the white paper is focused on scholarly publishing tools. For the purpose of this work, scholarly communication is therefore considered here in close relation with writing and publishing.

The final aim of the SIG tool is to contribute on its specific topic to the building of an infrastructure for open scholarly communication. Openness in this paper is mainly focused on openness of the tools' source code, but it should be more broadly understood. To be fully open, an infrastructure should also rely on open standards, community-led governance, and inclusiveness, to cite only a few dimensions of openness. With a particular attention to the openness of the tools themselves, the SIG tools white paper intends to insert its analysis and recommendations in the broader context of these various dimensions of openness.

The objectives of the OPERAS SIG on tools for scholarly communication are more specifically to identify the main perspectives of development, implementation, or coordination useful to the OPERAS members, and more broadly to the open scholarly communication community.

The work of the SIG consists of:

- A technical watch on reports, developments, and trends.
- A list of relevant tools, detailing features and functionalities.
- A common approach and criteria for choosing tools.

The list of the tools described in this paper is reported in [Annex I](#). However, this white paper does not aspire to provide an exhaustive catalogue or a detailed benchmark of publishing tools, which should be in fact the objective of future projects. It provides instead an identification of scholarly publishing main functions, tools, and trends, as well as minimal guidance for the users and areas of potential developments.

In order to do so, the white paper is structured as follows: after a section defining the publishing **tools main characteristics**, tools are then analyzed according to the main functions of publishing.

Traditionally, it is acknowledged that scholarly publishing's main functions are "registration, certification, preservation, dissemination"². However, as authoring and the tools used for authoring have direct implications on the general process, it seemed more useful to structure the paper around these three functions: **authoring, certification, and publishing**. The section on publishing includes various aspects of registration, dissemination, and preservation.

An additional section focuses on the specific function of **communicating**, with commenting and annotating tools that can serve the purposes of the previous functions.

² J.-Cl. Guédon, "Scholarly Communication and Scholarly Publishing", Open Access Scholarly Publishing Association blog, April 21, 2021, [<https://oaspa.org/guest-post-by-jean-claude-guedon-scholarly-communication-and-scholarly-publishing/>].

Perspectives for the field are then presented in a section dedicated to the existing **major trends**. Finally, the last section gives a series of **recommendations** in the prospect of OPERAS future projects.

It is important to stress that the activities of OPERAS members are very diverse and related to different functions or stages of the scholarly communication process. Although such diversity can provide a good coverage of the overall process, it may have affected the balance between the various functions here described. However, the entire renewal of the SIG members between the two versions of the paper, as well as the review by external experts, have allowed to increase the level of completeness and to provide, so do we hope, a better-balanced version of the paper.

3.3. USERS AND USAGES

Working on this white paper, the SIG Tools discussed the importance of considering the usage when reporting on tools. It can concern how the tool is easy to use, what are the skills needed, how it addresses users' needs or perhaps creates new needs. In our perspective, users can be authors, readers, editors, scholarly publishers, and providers of publishing services. We can therefore schematically distinguish three types of users with different types of usage: researchers as end-users, editors as intermediate users, and providers of publishing services (publishers, libraries, archives, etc.) as advanced users. The white paper addresses all these types of users, keeping in mind that their needs are very different and that not all of them are concerned by all types of tools, that there is a diversity of contexts, and that they are nevertheless strongly linked to each other. Lastly, we identified in our survey this focus on users as a major trend in scholarly publishing. That appears, for instance, in the STM Association's [Top Tech Trends 2024](#) and a recent post by Roger Schonfeld on [The Scholarly Kitchen](#) (2021, with an interesting conversation in comments), calling for greater consideration of user needs and experience. Generally, the intent is to make things easier for researchers, for instance in handling multiple logins in the submission or review process from different journals or publishers, or managing multiple requirements for submissions, including different reference formats.

3.3.1. Researchers

As end-users, researchers can be authors, readers, and reviewers. Thus, we can suppose that they are concerned with authoring, peer reviewing, annotating, reference managing tools, and of course by a lot of other tools considering the complete scholarly communication process. Their practices take place in a professional context and their uses are therefore also social uses. For instance, as it is well known about authoring tools, LaTeX is largely used in some communities, especially in STM, while Microsoft Word remains the most used tool in SSH. But SSH are far from being a uniform entity and practices can be very different from one community to another, depending on research fields, language, etc. Moreover, the adoption and use of a tool can depend on several factors, like for instance, among others, the age of the user, or the institutional context (including research structure of the country, time dedicated to research, technical, financial and human resources, etc.). When considering potential barriers to tool use by end-users, these social factors are important to consider, especially in terms of reward and recognition. These questions are not well documented yet, even if

some studies exist³, and must be taken into account to support innovation. In other words, a tool is a success if it meets users' needs and if it is used by a large community.

In parallel, the adoption of new tools by researchers can be related to different types of criteria. These can be technical, aiming at improving the quality of integration into existing workflows, interoperability, etc., but they can also correspond to more overarching goals such as openness, transparency, efficiency, the robustness of the research process, including publication and dissemination. Within OPERAS project (task 4.1) a prototype tool has been designed to help researchers in selecting between the services (including several tools) provided by OPERAS partners, called [Pathfinder](#).

3.3.2. Editors

Editors operate the publishing process of journals and books, more and more including their connection with data. They can contribute to all parts of the process: managing peer review, copy-editing, typesetting, content structuring, dissemination, etc. They can be professionals (employees or subcontractors) or researchers (early carrier or senior), depending on the kind of structure they work in, the economic model of the publications, and the institutional context of the country, particularly concerning employment in this field.

As they are professionals, they are trained to use a variety of tools, from authoring (to support authors in writing) to publishing tools, including all types of tools that help to check quality. They can be considered intermediate users as they are supposed to have more technical skills for using specific tools than end users, without necessarily having the skills for installing, maintaining, or developing complex tools. In many cases, editors are also introducing new tools to the authors and are included in the development of the publishing tools.

3.3.3. Providers of publishing services

Working for platforms, libraries, publishing structures, and others, these users can be considered advanced or expert users. Their missions are part of the international scholarly publishing and communication ecosystem and address technical issues like interoperability, dissemination, conservation, etc. They can host, maintain and develop tools, and they also have to work with other providers of services. Their needs are therefore very different from other types of users, but they have to meet the needs of the community they serve and train or support the people who use their services. In mentioning the publishers in that category, we shouldn't forget the intellectual work they realize by building a catalog with collections, or defining an editorial line for journals, as well as by accompanying authors in writing, dissemination, or legal issues. The tools considered in this paper support this work carried out in close collaboration with editors and researchers.

This paper intends to provide useful information to these different types of users, and it is a major objective of the final recommendations to address their distinct needs more specifically.

³ For instance, starting from the predominance of Microsoft Word usage in all aspects of text production in SSH (writing and editing), Nicolas Sauret (2020) analyses that usage and discusses the concept of digital literacy. The forthcoming 'Future of scholarly writing in SSH' study conducted in the OPERAS-P program will give useful insights on that question. See also, about reading and other documentary practices of researchers, the recent Couperin report (Baligand, Colcanap, Harnais, Rousseau-Hans, and Weil-Miko 2020).

4. DEFINITIONS AND CRITERIA

4.1. Definitions

Given the above, we understand that tool types and characteristics can be very different, as a tool can be:

- A technical brick (software libraries or frameworks) which still needs to be adapted and/or integrated into a wider application to be utilised (e.g. Grobid)
- A software application which still needs to be configured, installed and maintained (e.g. OJS, Lodel)
- A ready-to-use software as a service fulfilling one or more functions of the scholarly communication activities (e.g. Scholastica, Publons)

Note that software as a service tool includes end-user applications and technical services (APIs) to be used by other software, via the network. Moreover, especially in the case of OPERAS members, some major characteristics are crucial in analyzing and/or selecting the publishing tools, as they can be: integrated or independent, open or close, more or less interoperable.

4.1.1. Integrated vs independent tools

As already mentioned, the functionalities may be more or less integrated or available in separate software products which may need more or less custom development or configuration, so as to interoperate and form a complete publishing chain. At the same time, integrated tools may imply acquiring an important set of skills to be managed with ease.

The tools considered in this paper address a number of functions that are part of the researcher workflow according to R. C. Schonfeld (2017): writing, collaborating, reviewing, and publishing. Nevertheless, as recent trends indicate (Schonfeld, 2018), the major commercial players in the academic publishing sector are integrating more and more functions and services to cover the workflow more completely, and this presents for the researcher community the serious risk of being locked in a particular suite of tools.

4.1.2. Open vs closed source

This lock-in risk also legitimates one of the key assumptions of the OPERAS RI: scholarly publishing is an essential part of research activity, and the SSH community (including the OPERAS partners) should have a certain control over tools and contribute to tool development. In other words, we believe that scholarly communication tools should be community driven. This is why we will have a particular focus on Open Source tools, as they can (at least potentially) be adapted and extended by the community. However, we will also mention closed source software that are widely used or have interesting features.

The CEO of Hindawi publishing, Paul Peters, stresses the risks of relying on proprietary scholarly communications infrastructure and promotes the move towards an open scholarly infrastructure, which will be challenging. In his views, "in order to prevent private companies from owning and

controlling this infrastructure a radically open approach to its development is required" (Peters, 2017). The proposition is to ensure simultaneously Open Source, Open Data, Open Integrations, and Open Contracts. In fact, not only the data should be open but also the infrastructure managing them and the implementation type of the services (Neylon, 2015). The publication software editor and OA journals publisher Scholastica (Scholastica, 2017 and 2018) stresses the importance for the academic community of having a consistent toolbox in order to take back control of the publishing process.

In fact, and more precisely, 'open source' is not necessarily a guarantee in the sense that the startup that produced the software may be bought by a larger company and the licence may evolve overnight towards a closed source license (Pooley, 2017). Although a community may still fork the initial (open source) code, in practice it means that the "true" openness criterion is that the tool should be managed by an open community.

4.1.3. Interoperability

Alongside the governance issues, the use of many publishing tools in many environments also implies Interoperability challenges. Such challenges are common to virtually all tools: how to enable a user moving (easily) data and documents from one tool, platform, environment to another. Interoperability is mostly addressed by another OPERAS SIG dedicated to standards. However, the question will also be considered here as it represents a specific aspect of the practical issues faced by many publishers, especially small ones.⁴

4.2. How to select the appropriate tool?

Following this first general level of analysis, we propose below a simple method for selecting tools according to a set of clear criteria.

4.2.1. Method

1) The first step is to have **a clear idea of the requirements**. This is often not easy, and, in the OPERAS context, the SIG Best Practices can help to clarify the requirements. In the case of OPERAS, the question may be not only to cover the particular needs of one partner, but to find an open source tool that can be reused and adapted to cover the needs of several organisations, at least the needs of the publishers. The requirements can be summarised as a list of criteria, which can be grouped into technical, functional, usage and governance.

2) It is also necessary **to be knowledgeable of the "tool landscape"** (or market) to be able to select candidate tools to examine more closely. This is where a list of tools grouped by function can be useful. The business need is often complex and not limited to a single well-defined functionality. The available software or services may not cover the need completely, or, to the contrary, may cover much more than what is needed. However, it should also be noted that the tools do not actually offer

⁴ J. Bosman and B. Kramer, from the 101innovations project, have discerned and discussed a few other criteria: non-profit, open licensed data, free to use, stakeholder governed. See: <https://docs.google.com/spreadsheets/d/1h0Aq6NYleVnLDw33vx1SGnv1jbE2B7widbHhU7tpiUw/edit#gid=2141288902>.

themselves like they would in a supermarket: some contextual parameters limit the options (users' capabilities, authors and publishers division of work, institutional support for specific tools).

3) It is then possible to **compare candidate tools** to assess which is best suited to specific needs and to evaluate what further development still needs to be done to meet requirements.

4.2.2. Criteria

Some criteria are common to all tools; some are, of course, specific (features). Also, and more importantly, the border between tools is not always clear: as said above, many services or platforms include several tools, so an authoring tool, or a publishing tool might be associated to one platform (or worse, cannot be used outside this platform); or several tools may be part of a software suite and are designed to work together and cannot be used separately (without a large adaptation effort, when they are open source).

As mentioned before, in the perspective of this paper, openness of the tools is one key aspect for the analysis, and therefore also an important concern for the selection's criteria.

Technical criteria

The technical analysis helps to narrow down the choices and can address these questions:

- What type of tool is it? A technical brick, an application software, a running service?
- Is the tool mature, regularly updated, or in a development stage? Does it rely on sound and sustainable technologies (programming language version, coding best practices, etc.)?
- Is the tool based on open standards? E.g. which structured document formats are supported? Does the tool follow NISO standards⁵?
- Does it support persistent identification of books, journals, authors, institutions, funders, etc.?
- On which technology (e.g. language, framework) is it based?
- Is the tool part of an integrated tool suite (risk of vendor lock-in)?
- How does the tool perform⁶ (e.g. response time...)?

Usage criteria

The questions about usage should help to define the service provided by the tool and to assess its quality:

- Is the tool easy to use?
- Is it easy for a newcomer to understand what the tool really does (this is far from always the case, when first visiting a tool's website)?
- How large is the user community?

⁵ See for example <http://www.niso.org/standards-committees/ebmd> or <http://www.niso.org/standards-committees/odi>

⁶ This criterion is often not very critical nowadays, but can be important depending on the use case.

- What is the scope of the tool: e.g. for which kind of publication is it intended: journals, books, both, other kinds of documents?
- Is it well documented? Does the tool have tutorials, FAQ, forum, etc. for the questions by the users? Is the community local, national or international?
- Is the software available in different languages?
- Does the software license allow further building, integration, dissemination?
- If it is an application/component, is it open source? If it is a service, is there a transparent use policy?
- Is the tool accessible via an existing platform with a good quality of service? Or does it need to be installed and operated by the user's organisation?
- Does it offer guarantees for privacy, no selling and destroying of user data?

Governance criteria

Governance criteria are key for assessing perennity of the tool:

- What is the software license?
- Who owns the software? Is the tool owned by a private company, an institution, or is it governed by a community?
- Is the tool free of charge or does it have a transparent pricing policy?
- Are the governance rules defined somewhere?
- Is the tool non-profit or for-profit? Is there a membership model, license fee, donation based?
- Does the tool have a roadmap? Is the development active?
- Is the software editor a member of an industry coalition (such as AAK for annotation, etc.)?
- Does the tool provide a contingency plan? Is the tool's sustainability demonstrated?

Functional criteria (features)

Of course, features are very dependent on the kind of tool (peer review, authoring, publication). A feature list can be established based upon the feature list of existing software and services on a tool's website.

In the 2018 version of this paper were drafted sets of criteria and features to analyze the various tools. This preliminary work led to establishment of a more comprehensive list reported in [Annex II](#). A more detailed table gives an example of online annotation tools analysis in [Annex III](#).

5. AUTHORING

5.1. Towards online and collaborative authoring tools

The web publication domain is very active: the W3C has a dedicated [Publishing Working Group](#) and open source software are flourishing. In fact, within recent years, a large number of native web authoring tools have been developed, often within the academic environment.

This seems to be a promising and important trend, as it may greatly facilitate the authors' work and transform the editing process. A key feature that goes along with online authoring is access to collaborative features (synchro, version control, etc.), as, in principle, any authorized user can edit a document concurrently with another user.

In a broader prospect, online editing capabilities can impact the whole publishing workflow, considering two key aspects: collaboration and interoperability.

In the SSH context, the authoring software is usually still Microsoft Word, and the peer review process is done on the Word document and managed by a workflow to produce a PDF publication. Online collaborative tools can greatly modify this process by enabling online writing or typesetting, and especially collaborative peer review. Publishing functions (see Publishing section) also may be impacted and become more seamless when linked to an online authoring tool. In that sense, online and easy-to-use tools could be a critical opportunity to move "away from PDFs" and the traditional publishing process (Scholastica et al., 2017). It should be mentioned also, however, that collaborative authoring, in research and beyond, is currently highly dependent on the proprietary suite from Google. The connection between advanced collaborative features, such as on the Google drive, and the communication and discovery Google services further increase the risk of lock-in.⁷ In fact, from Google, to Microsoft and then to Adobe, it is possible to cover an almost entire publishing workflow only with proprietary tools, interoperable mainly with one another.

However, online tools based on structured formats can operate through a specific workflow, especially as far as formatting or typesetting are concerned. In a traditional workflow, when the article or the book is ready for publication, it is usually converted to an exchange format. The exchange format often uses a [markup language](#) such as XML/TEI for SSH, JATS for medicine and biology (and for SSH, in the case of Scielo), LaTeX for maths and physics. In the case of online tools, as they are natively based on these exchange formats, the conversion challenge is solved in great part. However, it is still needed to ensure interoperability between different exchange formats, especially when tweaked versions are used.

With these kinds of technologies, it is then possible to achieve or envisage interoperability use cases such as:

- The conversion between structured formats Markdown, LaTeX, XML (TEI, JATS...), Word/Office styles

⁷ The risk of lock-in can actually take a variety of other forms (see for instance J. Bosman, <https://twitter.com/jeroenbosman/status/1194618057181794306>). In particular, the capturing of users' data and information by a tool provider represents both a cause and an effect of an increased lock-in.

- Exchange with Peer Review tools and with publication tools (e.g. FidusWriter and OJS, Lodel and OJS etc.)

In fact, interoperability between various tools is of major importance for the community as it is able to build a continuous environment of production. The example of Fiduswriter providing a formatted content used in the OJS Peer Review process could, in turn, inspire integration of the same Peer Review process with Lodel's XML-TEI file generation.

An important example of such open source tools for conversion between mark-up languages is offered by [Pandoc](#). The software allows conversion through command lines from and to LaTeX, Markdown, Epub, PDF among others, but also from and to word processors' files, MS Word, OpenOffice/LibreOffice.

Tools which exemplify the new trends mentioned in this introduction, in particular open source software are listed and briefly described below.

5.2. Structured formats in open source tools

5.2.1. Markdown & LaTeX

General purpose in-browser Markdown editors are getting mature, e.g. [Dillinger](#), [StackEdit](#) or [Pandao/Editor.md](#), all three open source software.

In-browser editors for scholarly publishing are not so mature, but development is active:

- [Manifold](#) is developed and used by the University of Minnesota. It allows for book editing.
- [ProseMirror](#) is an open source toolkit for building collaborative text editors, used in two projects in the scholarly publication community:
 - [FidusWriter](#) is funded as a German research project and has interesting plugin features which include an OJS plugin and the ProseMirror editor.
 - MIT's [PubPub](#) editor is both an open source editor and a publishing community.
 - [Sciflow](#) is free for single users but not for organizations. It proposes advanced editing and collaborative tools. It is also based on the ProseMirror toolkit and uses HTMLBook format as its pivot format (<https://www.sciflow.net/en/faq>). It uses open source bricks at the moment and should be fully open source in the future.

All three rely on Markdown and support LaTeX.

- [ManuscriptsApp](#) is an authoring tool for Mac users, which is open source (the code is available here: <https://gitlab.com/mpapp-public>). The tool is part of the Connect suite by Atypen, which includes Scitrus (discovering) and Authorea (publishing).
- [Stylo](#) is developed by the Canada research Chair in Digital Textualities. Based on Pandoc, it integrates a metadata editor, a version manager, a bibliography manager, different export formats (PDF, XML, epub, HTML 5, etc.), and an annotation tool via Hypothesis. It is both an authoring and an editing tool and it has been integrated into the French RI Huma-Num services in 2020.

5.2.2. XML

eLife Sciences is developing a suite of tools, under the name [Libero](#), to assist in the full journal publication process. The [Libero Editor](#) tool provides a user-friendly, “what you see is what you get” (WYSIWYG) interface for editing high-quality JATS XML. This builds from the work of the Substance Consortium to develop the Texture JATS XML editor⁸. Former Substance Consortium members PKP, Érudit and SciELO are actively participating with eLife in the development of Libero Editor. PKP is currently examining how to best support Libero Editor within its applications Open Journal Systems (OJS) and Open Preprint Systems (OPS).

Some of the developments from the eLife team are based on the framework [PubSweet](#), created by the California-based Collaborative Knowledge Foundation (CoKo). CoKo aims at building an open source editing and publishing framework, however now mostly focused on HTML production. CoKo developed a book production platform ([Editoria](#)) and a micro-publication platform ([micropublication.org](#)). Other organizations developed other platforms thanks to PubSweet, for instance: a journal submission and peer review platform by the Hindawi team, a manuscript submission and peer review platform built by the eLife team.

5.3. Proprietary software as a service

Closed source tools may be somewhat more mature, e.g. the proprietary platform [LeanPub](#). Specializing in book authoring, the Leanpub web editor allows direct export to PDF, EPUB and Mobi. The service also comes with a selling storefront.

Other proprietary tools specialized in article editing and publishing, like [Authorea](#) and [Overleaf](#), provide a full set of services.

Here is for instance the official Authorea [feature list](#) :

- Service: Hosted installation, 24x7 support;
- Data management: Host Data for tables and figures, Mint a DOI, Version control (Git);
- Authoring: History view, Templates for leading conferences, institutions, and journals, Collaborate and manage co-authors, Comments, Equations editor, Interactive figures;
- Publishing: Multiple markup languages (add blocks of Markdown and LaTeX to your document as needed), Advanced export and journal styles, Direct submissions to a growing number of journals.

Overleaf offers similar services and is based on LaTeX and Rich Text. For its range of services, [Authorcafé](#) seems to be more a platform than a specific tool but doesn't make its technical environment very specific.

⁸ As reported in our 2018 paper, the same organizations were before involved in the development of Texture through the Substance Consortium.

5.4. Companion tools for authoring

Commercial tools assist authors in finding the best journal for publishing their research and/or adaptation of their article to the submission rules of the journal, such as:

- APA Style Central: <http://apastylecentral.apa.org> (no OA/Free version)
- Manuscript Matcher: <http://endnote.com/product-details/manuscript-matcher> (free version only as trial)
- Open Journal Matcher: <https://ojm.ocert.at/> (free);
- JournalGuide (American Journal Expert): <https://www.journalguide.com/> (free);
- Open Access Journal Finder (Enago academy):
<https://www.enago.com/academy/journal-finder/> (free).

Reference management tools are a key aspect of authoring and detailed information can be found on a dedicated Wikipedia page.⁹ Among the open source tools, some are already well known, like for instance [Zotero](#) or [BibSonomy](#). The already mentioned [FidusWriter](#) also includes as one of its main features the reference management. There are currently innovative tools such as [recite](#) (beta version), that allows to check the consistency of the references against the text's content.

Related to another aspect of researchers' authoring activity, MECA¹⁰ is a proposed mechanism (ZIP folder with JATS-like XML files) to simplify transfer of manuscripts across publishers. Participating organizations (and systems) include Clarivate Analytics (ScholarOne), Aries Systems (Editorial Manager), eJournal Press (GEMS), HighWire (BenchPress). Although the use case for MECA is in STM and Biology, it may be also of interest in the SSH context.

Tools that help researchers to improve written quality of their works (translation, grammar check, synonyms suggestions, etc.) can be considered as companion tools for authoring, and can also be used by editors for copy-editing. We decided not to include them in the scope of the White Paper for two reasons. First, they are strongly related to language and identifying a list of these tools for many languages reflecting the OPERAS community is a big challenge. Secondly, most of them are not specifically dedicated to academic writing. We can however mention two IA based tools with a free version: [DeepL](#) for translation that seems to be more and more used in the academic field, and [Grammarly](#), a writing assistant that is also widely used by researchers, for English only.

6. CERTIFICATION

Structured and uniform certification practices are a prerequisite for creating a standard for scholarly material that works across platforms, academic sub-disciplines, publishers and geographic regions. The aim to increase the accountability of research within humanities and social sciences is dependent on publishers and libraries continuing to develop services for authors and readers and making them digitally accessible and searchable to a greater extent, especially if they want to catch up with publishing within sciences, technology and medicine and with the quickly developing journal

⁹ https://en.wikipedia.org/wiki/Comparison_of_reference_management_software

¹⁰ See <https://scholarlykitchen.sspnet.org/2017/08/17/meca-new-manuscript-exchange-initiative/> See also [this presentation](#) for more details.

platforms. This concerns in a first step the detection of potential plagiarism, an important area of technological developments in recent years described in our first subsection. The following subsections dedicated to peer-reviewing are meant to outline available tools for managing peer review and spot the gaps or challenges with available services.

6.1. Anti-plagiarism tools

Plagiarism is a continuing concern for scholarly publishers, especially as the pace of publishing online only seems to increase and the academic job market seems to become even more aggressive. At the same time, organizations such as DOAJ endorse the implementation of a plagiarism policy and/or the usage of plagiarism tools to improve journal quality (see eg. <https://doaj.org/apply/guide/>). While plagiarism attempts or inadvertent mistakes may ideally be caught as a submission makes its way through a journal's initial screening, review and acceptance processes, some publishers may also wish to include an automated plagiarism detection service to provide additional safeguards or to help with particularly heavy workloads. Automated plagiarism tools may help.

Most automated plagiarism tools compare manuscripts against an internally developed manuscript and website corpus to check for similarity in the manuscript text. (These data corpuses are built by harvesting publicly available content from the web, including web pages and published article full-text where available; and usually also include licensed content from publishers and other aggregators). They then provide a similarity score, which might indicate the level of potential plagiarism identified. Most tools also provide some other sort of evaluation of the manuscript, for example a report that includes the relevant matching paragraphs. Some tools have also been developed with authors in mind, to allow them to do a self-check of their content before it is sent for review (Grammarly, for example).

It is important to remember that the automated tools listed below are not perfect, and that any warnings be very carefully reviewed in order to identify false positives, including against the author's own work. In Canada, for example, there has been some anecdotal evidence to suggest that some of these service providers, who have harvested university institutional repositories as part of broader university agreement, have then falsely flagged researcher manuscripts based on pre-prints as plagiarized, simply because the author's manuscript has matched their own preprint. Likewise, when it comes to supported languages or disciplines, the plagiarism tools are only as effective as the data corpus they have. Therefore, some of these services are better in some regions, disciplines and languages than others, simply because of the contents of their data corpus.

Finally, it is crucially important to clearly and publicly state the publisher's plagiarism policy and procedures, and to obtain consent from the authors if the publisher will be working with a third party service provider. This is particularly the case in particular with plagiarism services like Turnitin/iThenticate, who may add the submitted manuscripts to their data corpus for re-use. Such activities must be clearly agreed upon by the authors.

A recent survey and evaluation of the most commonly-used anti-plagiarism tools was published in the [International Journal of Educational Technology in Higher Education](#). As mentioned before, there is a distinction to make between applications and services, between open and proprietary tools:

most of the anti-plagiarism tools are proprietary third-party services, often with little transparency about their software code or data policy. Some of these tools provide a free version that may imply limitations of the service.

Among the widely used anti-plagiarism tools, we can mention [iThenticate/Turnitin](#). iThenticate is an anti-plagiarism tool developed by Turnitin LLC (and sometimes the names are used interchangeably). This is probably the most widely used anti-plagiarism tool available. iThenticate already has integrations with a wide variety of manuscript management tools, and a REST API that can be used for future integrations (see <https://www.ithenticate.com/products/fags#partners>). Crossref's Similarity Check service (<https://www.crossref.org/services/similarity-check/>) is also an iThenticate integration; using Similarity Check allows you to use the full iThenticate service. [Ouriginal](#) is a combination of the former Urkund and PlagScan services, Ouriginal is geared primarily towards university learning management systems (LMS's) to evaluate student work. The PlagScan software (<https://www.plagscan.com/en/>) is still available for single users and other uses, however. [Grammarly](#) is a writing assistant meant to be used during the authoring process, and also includes a plagiarism checker (<https://www.grammarly.com/plagiarism-checker>).

6.2. Peer-reviewing overview

First and foremost, there are standards describing the outline of the peer review process, such as guidelines for editors, reviewers and authors on an international level provided by the [Committee of Publication Ethics](#) (COPE) or national initiatives like the Belgian [GPRC](#) mark (Guaranteed Peer Reviewed Content). The DOAB [toolkit for Open Access books](#), recently launched, provides specific guidance regarding peer reviewing. In this report, we focus on the more technical aspects of peer review to facilitate the standard, such as systems for peer review (including open peer review), peer review tracking, and tools for standardising paper submission workflow to different publishers (to ensure a smooth review process).

The process of peer review can be anonymised, partly anonymised or completely open, depending on the academic subject area and the scholarly community available within that realm. The different types of peer review and the application within academic disciplines have been well described in the article 'A multi-disciplinary perspective on emergent and future innovations in peer review' (Tennant, J et al., 2017). Most proprietary and open source publishing platforms, such as [OJS](#), for management of academic journals include a module for peer-review as a part of their core services. These systems allow editors and management to maintain a structured process and to create an archive for editorial processing to enable transparency. Most commonly used systems for peer review management, such as [Editorial Manager](#) or [ScholarOne](#) include sophisticated modules for reporting on user activity, automating process and measuring quality of submitted reviews. However, the need for more powerful reporting about editorial activity in open source software seems to be on the wishlist for many editors.¹¹

The peer review process for books is handled slightly differently from journals, as this is an evaluation process that in the past was managed under the discretion of the academic publishers.

¹¹ As of OJS 3.2, there are now editorial activity metrics and an emailed report included directly within OJS.

With the growing movement among university presses, where a lot of emphasis has been on creating spaces for Open Access monographs, there has also been a push for developing tools for peer review of such publications. Examples of systems supporting the evaluation process are the Public Knowledge Project platform [Open Monograph Press \(OMP\)](#), and the [Rua platform](#) provided by Ubiquity Press. These systems both provide management platforms for the entire editorial process related to monographs and edited volumes, including a module for conducting structured peer review, but also the production and distribution of electronic books. Both OMP and Rua are open source systems available for free download and adaptation by users. Publishers are already experimenting with annotation tools like [Hypothes.is](#) to provide a more transparent editing processes and open peer review for books.

To get further acquainted with ideas on how to develop better peer review tools, please consult documentation from the [Peer Review Transparency Workshop](#). This group of scholarly publishers, academic librarians, and IT experts is working to establish peer review standards and possible peer review labels comparable to the Open Science framework badges for open practices.¹²

6.3. Open peer review

Funders of research and academic institutions are currently aiming towards a higher level of transparency within the scholarly communications arena, to take back some of the control over the current quality assurance process for articles from publishers who have been criticised for not doing a proper job. This process has also been the purview of commercial publishers for a long while, as they have had the resources to develop tools for improving procedures. Following the development of more open practices within scholarly communication, such as open access to publications and research data, as well as the increased use of preprint servers to release early stage works for critique, it seems natural to also open up the peer review process to scrutiny. This corresponds to open peer review (OPR) or post-publication review. Most academic publishers already have systems in place to manage peer review, but few have yet opened up the peer review process for readers to access the information from the process.

Open peer review means that the item is published online first, and reviewers are invited to publish their comments online. Usually, this procedure also includes versioning of the item to allow the author to submit subsequent revisions based on the reviewer comments. Other parameters can also be taken into consideration to ensure OPR: open identities (degree of blindness during the peer-review and/or names published), open peer-review reports, etc. Using open peer review could potentially address several perceived problems with the current practice of scholarly quality control, such as unreliability and inconsistency, as well as a lack of incentives for peer reviewers (Ross-Hellauer, 2017).

¹² <https://www.cos.io/initiatives/badges>.

6.4. Tools for open peer review

Some current platforms offering open peer review, for example [F1000 Research](#)¹³ or [ScienceOpen](#), consider the open review procedure as a fundamental part of the publishing service; filtering is not done by the editors but by the open peer reviewers. Entire open peer review networks are emerging, for example [Peer Community In](#), where the creators aim to develop an open community for researchers interested in OPR, to develop best practices, and to provide a list of potential experts who can be invited. Now counting 11 communities, the platform also contains subject-specific networks, such as the first community in evolutionary biology (<https://evolbiol.peercommunityin.org/>), where authors can upload their preprints and get comments from peers before they submit to journals.¹⁴ The [Open Review Toolkit](#) enables anyone to convert a book manuscript into a website that can be used for Open Review using the Markdown format. Developed at Princeton (relying on [Pandoc](#) and [hypothes.is](#)), this software takes a book manuscript (currently formatted only in Markdown, which is quite limiting), converts it to HTML, and enables an Open Peer Review for that document. Other examples of innovations or platforms for developing and opening up the peer review process are ‘[Peerage of Science](#)’, ‘[Publons](#)’ and ‘F1000 Research’. These three services are described in the article ‘What’s next for peer review?’ (Research Information, 2016). Within Peerage of Science, the reviewers are not selected by the editors, but by the authors, and the reviewers are evaluated by the platform’s users. Publons offers a database of reviewers with the record of their previous reviews. In the field of Life Science, publishers launched another innovative initiative, ReviewCommons, aiming at providing journal-independent pre-submission peer-review.¹⁵

6.5. Anonymised peer review

Anonymised peer review is a challenge for many editors, as this is something that has to be done at an individual article level where the author has uploaded a document, and thus most of the work is done manually in the software used for writing. A tool for anonymisation would need to include checking references for self-citing and reviewing the linkage data in the actual text, as well as the user settings in each document for information that would reveal the author’s identity. A truly anonymised work is in practice extremely hard to achieve, especially in small academic fields where many researchers already know each other from meeting at conferences or other networking. There seems to be a need, however, to develop such a tool, so this could be something to consider, like for example building a plug-in to OJS to check for the author name being mentioned in the submitted material. Another useful tool to preserve author/reviewer integrity would be to use automated checks for conflicts of interest between authors and reviewers (answering questions like: have they collaborated on the same project or worked together in the same department?).

¹³ F1000 Research is owned by Taylor & Francis and implemented in the open publishing platforms of three major science funders: Gates Open Research, Wellcome Open Research, Open Research Europe.

¹⁴ Recently, 15 journals agreed to outsource peer-review through PCI:

<https://www.sciencemag.org/news/2021/04/fifteen-journals-outsource-peer-review-decisions>.

¹⁵ <https://www.embo.org/press-releases/review-commons-a-pre-journal-portable-review-platform/>.

6.6. Peer review tracking

Peer review is a critical mechanism for the scholarly communications landscape to function. The added value of peers who donate their time to evaluate potential publications for consistency and accuracy is enormous. Most of this work is done by researchers without any guarantee of recognition or reward, as the work is considered to be intrinsic in what it means to be an academic. In recent years there has been an ongoing discussion within academia that questions the added value for those who spend considerable time commenting on the work of others. Digital practices in publishing allow more opportunities, however, to do something about this lack of information on the number of completed reviews per year per researcher. We have, therefore, seen an emerging trend of tools being developed to better track and ease peer review activity (Tattersall, 2014). This would, however, demand that systems for peer review be aligned with tools to recognise users with unique identifiers, such as [ORCID](#). The integration of data about peer review activity is already being used by OJS (via an ORCID integration plugin), F1000, American Geophysical Union (AGU) and Publons. In 2017, these services added information to 9,800 ORCID records, to add to users' personal pages.¹⁶ The challenge with the tracking of peer-review on a wider basis is that it requires a digital workflow standard¹⁷ that not all systems deliver at the moment.

Many of the tools we found in this category seem to be proprietary in one way or another, apart from the platforms for managing the editorial process for books. There seems to be an open market for tools to enhance the editing process, where the paid-for services appear to be most used for the time being. [Publons](#) is, for example, free to use for researchers, but publishers have to pay for the service to be integrated in their systems as well as for extracting data, which many smaller organisations may not be able to afford. The partners would ideally like to have a similar more open tool but to ensure that data can be collected for all research output. Actually, regarding the search for potential reviewers, the main challenge is to collect information in a reusable database; if the data were available, the software itself could be developed with an open source licence. Such a large database would, however, need to take into consideration the integrity of its users in relation to legislation connected to [GDPR](#) and exhaustion of reviewers who may have to turn down too many invitations to review. Both ethical and practical guidelines should therefore be developed to meet the requirements of the GDPR on how such data should be used and processed.

7. PUBLISHING

The publishing here considered concerns mainly platforms registering, hosting and disseminating the published contents. It should however be noted that the concept of “publishing platform” can include a wider variety of types, notably the aforementioned example of F1000research. Such platforms offer specific functionalities with respect to hosting platforms: preprints, post-publication peer review and peer-reviewed articles. This variety relates to the evolution of publishing platforms

¹⁶ <https://info.orcid.org/peer-review-at-orcid-an-update/>

¹⁷ <https://members.orcid.org/api/workflow/peer-review>

and their combination with features typical of data repositories. This evolution is described a bit further in the [“Main trends”](#) section below.

7.1. Features of a publishing platform

Once a manuscript has been reviewed and typesetting is complete, it is quite straightforward to make the publication available on the web; any CMS would do that (e.g. [Wordpress](#)), and the content could then be found on search engines or from the web sites known to the researchers of the particular discipline. This was the state of the practice maybe 10 or 15 years ago, notably in SSH. However, the standard services expected from a scholarly publishing platform are now much more demanding and tend to increase each year. A good example of a publishing tool widely used for this purpose by smaller journals is OJS, which can either be self-hosted or be hosted by PKP. OJS includes more features dedicated to academic publishing than a generic CMS like Wordpress, for example an end-to-end submission to publishing workflow, tools to easily organise content into issues before publication, automated export of articles for DOI registration, and OAI-PMH end-points.

The publication functions of a hosting platform include Content Management (version control, status), quality checks, metadata annotation (authors, affiliations, keywords), bibliographic reference management, linking citations to standards such as ORCID, Funder Registry, DOI...), format production (PDF, EPUB, print, HTML, XML...), metrics and altmetrics, fee processing (may be relevant even for open access e.g. for APC or for a freemium model where HTML is free and PDF is not) and so on.

An important feature of the publishing process is making the content discoverable beyond the publishing website — this is often through active distribution — pushing to indexes like Crossref, PubMed, or passive methods like OAI-PMH, and presenting metadata for Google Scholar. There are some indexes (Web of Science, Scopus) which will harvest content manually, so a site just needs logically structured pages. As seen above, the functionalities may be more or less integrated or available in separate software products or services which may need more or less custom development or configuration to interoperate and form a complete publication chain.

7.2. Software overview

Some academic tools or software lists exist, either with a broad scope such as Utrecht University Library’s work,¹⁸ or focused on Open Access publication, as for instance, the list established by [Radical OA](#) in the UK or the one published in 2018 by the Scholarly Kitchen (A. Michael, 2018).

In the first version of this paper, we mentioned that, in 2018, the life science publisher eLife made a call to the community of open source publishing tools.¹⁹ During the same year, in the North-American area, two conferences addressed the open source tools 2018: the Library Publishing

¹⁸ <https://101innovations.wordpress.com/>. The list has not been updated for about four years but the works continues, and does link to a dozen other tools list.

¹⁹ <https://elifesciences.org/labs/f66b5b23/open-source-in-publishing-community-call-february-22>.

Coalition Pre-Conference on open tools in Minneapolis²⁰ and the open source Bazaar pre-conference at the Society for Scholarly Publishing Annual Meeting in Chicago.²¹

Since that time, as reported in the introduction, the topic has been further enriched with various studies and initiatives from Educopia, IOI, the MIT press, or the European Commission, not to mention the OPERAS works on innovations in scholarly communication. More specifically, we should outline two important actions in what concerns open source tools support and inventory. First, the Joint Roadmap for Open Science Tools ([JROST](#)) was launched in 2018 and organized a second conference in December 2020, for building, supporting, and advocating for open tools in research.²² In the process, JROST was integrated with IOI's development. Second, the already mentioned report "Mind the gap" recently led to the creation of a catalogue of "scholarly communication open technologies", [SComCat](#).

Such studies and initiatives certainly indicate directions for OPERAS future actions, and potential ways for broader coordination. Without aiming at exhaustivity in this paper, we report below the analysis of some publishing tools, which can be the starting point for a more detailed and standardized inventory.

7.2.1. List of software

Here we present a list of notable open source solutions for scholarly publishing:

- PKP's [OJS](#) (for journals) and [OMP](#) (for books) are certainly the best known scholarly publishing open source softwares.
- OpenEdition's publishing software, [Lodel](#).
- MIT's [PubPub](#) is a new collaborative edition and publication software designed for academic communities.
- [Hyrax](#) is a web front-end for the [Samvera](#) open source digital repository framework (formerly known as fedora/hydra); the samvera community seems quite active in the USA. The platform is developed in Ruby on Rails. It also includes a discovery tool called [Blacklight](#), which is a web front-end for the SolR search engine. The majority of use cases lie in academic library and repository applications; however, Samvera has been recently adapted by the University of Michigan for setting up [Fulcrum](#), a publishing platform; [Heliotrope](#) is the name of the software adaptation of Hyrax to meet publishing needs. It is interesting to note that the [LeverPress](#) project of a peer-reviewed, open access, scholarly, digitally native Press is based also on Fulcrum.
- Birkbeck Center for Technology and Publishing's [Janeway](#) journal platform.
- The SciELO platform (<https://github.com/scieloorg>), OS, well documented but large and complex.
- [elife](#), a UK-based non-profit biomedical publisher, has developed interesting open source software (<https://github.com/elifesciences/>):
 - [Lens](#) an online (JATS XML) document reading tool

²⁰ <https://librarypublishing.org/owned-by-the-academy-preconference/>

²¹

<https://customer.sspnet.org/404.aspx?aspxerrorpath=/ssp/2018-Meeting/Event-Home/ssp/AM18/Home.aspx>.

²² <https://investinopen.org/community/jrost-2020-conference/>.

- [Libero](#) (new name of Continuum) for journal publication
- [Pressbooks](#) is an open source platform for writing, editing and publishing e-books based on Wordpress. The files are in html format. Pressbooks can produce pdf (with Prince) and ePub files, with all the classic functions related to the production of books, manuals or collective works.

7.3. Integration with third party services

As previously stated, the publishing process is part of an ecosystem of interdependent services and platforms, which are not all part of the core CMS software but are provided by external service providers. Therefore, the publishing platform has to provide “hooks” enabling those services to be available (or third party software to be installed and to interoperate with the platform). It has to be noted that not all these third party services are free and open.

Given the wide range and the sometimes high specificity of these types of tools, it would be outside the scope of this paper to propose an exhaustive list. However, here are some examples of the functionalities and the challenges related to publications’ integration:

- Dissemination through data identifier:
 - DOI registration agencies such as Crossref (for scholarly publishers) or Datacite (working with the repository community)²³ allow for the published objects (books, chapters, articles) and also parts of the content (supplementary material, figures, tables) to be identified. After DOI registration, Crossref supports publishers adding them to reference lists either via the Crossref API, via authoring tools in production (e.g. eXtyle) or just by pasting the reference list into Crossref’s [Simple Text Query](#). The provision of metadata to Crossref upon registration also enables richer linking and discovery.
- Dissemination through author identifiers or author’s profiles:
 - [Kudos](#) (can be pushed by publisher, or by author)
 - [ORCID](#) (can be pushed by publisher directly²⁴, by publisher via Crossref²⁵, or by author)
 - [Academia.edu](#) (can only be pushed by author)
 - [Impactstory](#) (pushed by author)
 - [Social Science Research Network](#) (normally pushed by author, mostly intended for sharing)
- Dissemination through funder and/or research organizations identifier:
 - Publisher deposits Funder IDs associated with the DOI of the published object. Crossref make this available to funders via their [API](#) or [searchable database](#).
- Dissemination through aggregators (indexes are the most important discovery route, according to [Gardner and Inger](#)).

²³ <https://support.datacite.org/docs/datacite-or-crossref>.

²⁴ <https://members.orcid.org/api>

²⁵ <https://www.crossref.org/blog/crossref-to-auto-update-orcid-records/>

Establishing a typology of scholarly outputs' aggregators is challenging, given the diversity of processes, services and scopes. The list of aggregators below is approximately ordered from the broader to the narrower scope:

- [Google Scholar](#): pulls books and articles from the HTML metadata ([DublinCore](#)) on the publisher's site
- [Scopus](#) (journals and books): publisher either pushes PDFs to Scopus FTP or (for journals) Scopus pulls content from publisher's article browse pages (no technical requirements, other than logical browse pages)
- [Web of Science/Web of Knowledge](#) (journals and books): WoS/WoK pulls content from article browse pages, books and book series are submitted for evaluation
- [Dimensions](#) (journals, books and preprints) pulls openly available data from Crossref and others, but also gets information from publishers. It connects publications with other datas: grants, datasets, online mentions, policy documents, clinical trials, and patents.
- [Lens](#) (journals and books) pulls openly available data (Crossref, PubMed, Microsoft Academic, etc.).
- [Semantic Scholar](#) pulls data from repositories, publishers, and data partners like Microsoft and Unpaywall.
- [Core](#) aggregates repositories, OA journals, and full text of hybrid OA journals from all over the world, and hosts several millions of full texts.
- [Base](#) aggregates repositories and OA journals.
- [JSTOR Open](#) (OA publications and others): publisher pushes metadata and PDFs to JSTOR FTP.
- [DOAJ](#) (OA journals only): publisher pushes DOAJ XML metadata to DOAJ either manually via upload form or [via API](#)
- [DOAB](#) (OA books only): publisher pushes metadata via online form or file upload
- Subject-specific repositories — there are more subject-specific repositories as there are subjects, some notable examples are given here:
 - [PubMed](#) / [PubMed Central](#) (Biomedical journals only): Publisher pushes JATS-based article package to PubMed Central FTP (OJS supports PubMed export for journals which do not have JATS)
 - [PsycINFO](#) (Psychology): publisher emails PDFs on completion of issue
 - [HeinOnline](#) (Law): publisher emails PDFs on completion of issue
 - [Isidore](#)²⁶ is a search engine developed by Huma-Num providing access to digital data from the Social Sciences and Humanities (SSH), aggregating repositories, OA publications and more, enriching and highlighting digital data and documents.
- Country/language-specific repositories or indexes include:
 - [Latindex](#) (journals from Latin America, the Caribbean, Spain, Portugal only): publisher pushes metadata via online form.

²⁶ An enhanced version of Isidore gathering data from European repositories is currently being built by OPERAS: the [GOTRIPLE](#) platform.

- [CNKI](#) (journals from everywhere, which are then localised for a Chinese audience): publisher pushes PubMed XML metadata to FTP.
 - [Oasisbr](#) (portal aggregating metadata from all Brazilian institutional archives).
- Institutional repository: often this is done manually by the author or the repository manager — any more automated solution (eg via [SWORD](#)) depends on the repository
- [Portico](#) (journals and books): publisher pushes metadata and PDFs to Portico FTP to ensure long-term access to the publications.
- [CLOCKSS/LOCKSS](#) (journals and books): intended for long-term archiving, the mechanism pulls metadata and content from LOCKSS manifest pages on the publisher site.
- Metrics:²⁷
 - [Google Analytics](#) — supported by most publishing systems; reports on views, downloads, interactions, and user flow
 - [Altmetric](#) — enabled through Javascript insert or API; reports on alternative metrics (mainstream media coverage, social media coverage, citations etc), and provides weighted article ‘score’
 - [Plum Analytics](#) — enabled through Javascript insert or API; reports on alternative metrics (media coverage, social media mentions, citations etc), can include views/downloads, and provides weighted article ‘score’
 - [Crossref Event Data](#) — currently in Beta; enabled through [API](#); reports on social media coverage, citations, annotations via Hypothes.is, etc.
 - [OPERAS metrics service](#) — still in beta version, reports on social media coverage, citations, annotations via Hypothes.is, views and downloads²⁸.

8. COMMUNICATING

Online commenting and annotating publications, more than a function distinct from the previous functions of authoring, reviewing, and publishing, represent a transversal function, which is open to a variety of applications. Online annotations are a legitimate form of authoring, especially when made public and a part of a broader scientific conversation. From this point of view, online annotations are also a publishing tool, sometimes receiving the same persistent identifiers as scholarly publication. Moreover, the annotation tools can be integrated with a publishing tool, they can also, as mentioned above, be used as an instrumental piece of an online peer reviewing workflow, whether it is before or post-publication (see section “Tools for open peer review”).

The SIG focused its analysis on online annotation tools. In [Annex III](#) will be found an analysis of:

²⁷ Although WoS and Scopus primarily offer dissemination services, it has to be noted that the metrics they provide are critical for many publishers and authors and therefore highly influence their dissemination strategies.

²⁸ Since 2021, the OPERAS metrics service is referenced on the EOSC portal:
<https://marketplace.eosc-portal.eu/services/operas-metrics-service>

- [B2NOTE](#),
- [GetLiner](#),
- [Hypothes.is](#),
- [Remark Lite](#),
- [Pundit](#),
- [Notesalong](#).
- [Weava](#).

9. SCHOLARLY PUBLISHING MAIN TRENDS

The SIG Tools' survey identified and selected a few important trends in scholarly publishing: preprints, development of Artificial Intelligence (AI), and data papers. It is indeed necessary to be aware of what is identified by the major players as the future of scholarly publishing and to keep track of the envisioned innovations regarding tools and technologies.

9.1. Preprints

Preprints are not a new trend but the covid 19 pandemic highlights the need to better support their development and their role in scholarly communication. Born in communities of researchers and still the subject of developments, preprints are also a target for big publishers which are developing strategies to take control of it.

The preprints servers acquisitions and developments confirm a trend that was identified in our previous 2018 report around the notion of "Next Generation Repositories" promoted by [COAR](#) (Confederation of Open Access Repositories). The increasing number of disciplinary²⁹ and regional³⁰ preprint servers allows for the overlay publishing model, where peer-review and publishing workflows build upon a distributed, interoperable, and sustainable network of digital archives, repositories, and preprint servers.³¹ In a similar way, COAR recently published an updated paper on Pubfair, a "distributed framework for open publishing services"³². As reported in 2018, COAR studied various user stories showing the similarities between repositories and publishing platforms³³:

- Discovering:
 - Discovering metadata that describes a scholarly resource
 - Discovering the identifier of a scholarly resource
 - Discovering usage rights
- Usage:
 - Commenting, annotating, and peer-review
 - Automated recommender systems for repositories

²⁹ E.g. [SocArXiv](#) and [PsyArXiv](#) in 2016, [EdArXiv](#) in 2019.

³⁰ [INArxiv](#) in 2017, [AfricArxiv](#) and [ArabiXiv](#) in 2018, [IndiaRxiv](#) and [Preprints.ru](#) in 2019.

³¹ Cf. the experimental workflow tested on the platform Episciences.org (Berthaud et al. 2014).

³² <https://www.coar-repositories.org/news-updates/pubfair-version-2-now-available/>.

³³

<https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories/>

- Providing a social notification feed
- Data mining
- Supporting researchers' workflows
- System Management:
 - Recognizing the user
 - Resource syncing and notification
 - Comparing usage
 - Preservation

The shift towards publishing platforms enabling an integrated and modular workflow (from submission, through peer review, to publishing) is endorsed by the EU initiative to establish Open Research Europe. This no-fee publishing platform intended for all Horizon 2020 and Horizon Europe has been launched in 2021 ([‘Open Research Europe: Open Access Publishing Platform. Beyond a Research Journal’ 2020](#)).³⁴

Roger Schonfeld (2020) proposed an overview of strategies and recent investments of big publishers such as Elsevier, Springer Nature, Wiley, or Taylor and Francis in a field that was originally born in communities of researchers and built on repositories. The recent investments can of course be viewed as a strategy to take back control of these growing practices.

In terms of tools, publishers first started to build their platform and services to connect preprints with their article submission workflow tools, for instance, Springer Nature’s [“In review”](#) (with Research Square platform and Editorial Manager) or Wiley’s [“Under Review”](#) (powered by Authorea). But this investment in preprints is also visible in the acquisition of Social Science Research Network (SSRN) by Elsevier in 2016 and F1000 Research by Taylor & Francis in January 2020. The first one allows the company to add a preprint community value and the second provides a model of workflow based on post-publication open peer review.

9.2. Artificial Intelligence

Artificial intelligence and machine learning are emerging as a major trend in all publications that attempt to forecast scholarly publishing (eg. [Scholastica 2020](#), [Ann Michael 2019 in a Scholarly Kitchen post](#), STM Association posters). AI needs data, and its development is linked to the need for interoperable metadata and machine-readable (i.e. structured) content, as promoted by organizations, from the European Commission to cOAlitionS.

Scholastica’s post “5 Scholarly Publishing Trends to Watch in 2020” highlights curation and interpretation as two potential areas for AI applications. Peer review is another area where AI could be and is already being used, for example, to help find relevant reviewers by analyzing the content of a manuscript, or to assist human decision making by automatically detecting, for instance, "potentially low-quality or controversial studies" (Checco et al. 2021), to save time and human resources in the peer review process. Several tools exist that enable automatic detection, for instance, statistical reporting elements, presence or absence of other required elements. While not

³⁴ The contract for the tender has been awarded to the company [F1000](#): <https://ted.europa.eu/udl?uri=TED:NOTICE:134703-2020:TEXT:EN:HTML&tabId=1>

in AI in themselves, they can contribute to AI facilitated peer review. AI can therefore help to assess quality in complement with peer reviewer’s assessment. Of course, AI can also be used in other parts of the publishing workflow, such as typesetting, content translation, automatic annotation, or structuring (e.g. [Bilbo](#) or [Grobid](#)). Nevertheless, artificial intelligence has biases, risks and presents ethical problems, particularly considering its usage in peer review, when assessing quality objectively can be difficult. For instance, Checco et al. (2021) point that as “machine-learning techniques are inherently conservative, as they are trained with data from the past,” they could contribute to the persistence of inequalities for under-represented countries, groups, or individuals.

9.3. Data papers

Another area of development, somehow intersecting with the two previously identified trends, concerns the data papers. Datasets, but also software code can be considered as elements of scientific publication and communication, as they are important, for instance, for the administration of proof and the reproducibility of research. Data citation is, therefore, an important aspect of this trend and the development of data papers represents one specific use case. A publication that describes or more simply reports on datasets, the data paper could benefit from the increased connection between data repositories and publishing platforms, as well as from the developments of AI. Existing tools, still more widely used in the STM (Science, Technology, and Medicine) context, can extract metadata from a dataset and automatically generate a data paper. An open source example is [Data2papers](#), which provides two services: the production of a data paper after analysis of the metadata of the dataset; suggestions of journals that can host the article. Data2Papers is now integrated with the OpenAIRE Scholix implementation.³⁵ [Scholix](#) is a framework for data citation and the current implementations connect datasets and publications through the Datacite and Crossref DOIs. An initiative from the STM publishers association in 2020, led to an improvement of data citation through the use of Scholix.³⁶ Commercial tools for data paper creation include the [Arpha Writing Tool](#), from the STM publisher [Pensoft](#).

10. RECOMMENDATIONS

10.1. Information: User-centric Criteria

Establishing a clear description of tools according to transparent criteria

A first recommendation is to provide the basis for a shared knowledge about the tools functions and usability. A list of transparent criteria to describe the tool should provide not only technical information, but also more detailed descriptions of what the tool can be used for and the skills needed to use it. As mentioned above, there are different types of users with different needs and skills and that list will need to reflect this. This work could dwell upon the examples of 101innovation’s classification for research tools and of the “Mind the gap” report’s classification for open source publishing tools. It could also integrate the suggestion from the Educopia’s report to

³⁵ <https://www.openaire.eu/data2paper-scholix-integration>.

³⁶ <https://www.stm-researchdata.org/>.

create a taxonomy for publishing functions. The OPERAS SIG on Tools has already sketched a list of criteria for publishing tools that could be improved with reference to the cited reports and through a validation by the community.³⁷

10.2. Sustainability: A Tools' Observatory

Organizing the community effort to keep the information accessible, up-to-date, and archived

As a complement to the first recommendation, the OPERAS SIG Tools recommends to ensure sustainability through an accurate and collective monitoring of scholarly publishing tools. Continuing the efforts of 101innovation and the “Mind the gap” report, the Tool Observatory will provide long-lasting and updated information on tools, including archives on suspended projects. The Tool Observatory could be maintained by a dedicated team including, but not limited to, members of the OPERAS SIG Tools and the OPERAS Lab, and could work with other initiatives dedicated to tools for scholarly publishing (eg. [SComCaT](#)) or to SSH research tools (eg. [SSH Open Market place](#)). The team will be able to use the Tool Observatory as an incubator by suggesting features' developments of specific tools, interoperability enhancements between tools, or hackathons on new functionalities.

10.3. Training: A Training Material Catalogue

Offering training materials for users based on tools' testing, demonstrators, and summaries

As emerged during the discussions of the OPERAS SIG Tools and outlined by the CoalitionS report on Diamond journals, guidance should be provided to the tools' users, taking into account their various needs and levels of expertise. Various training materials can indeed be designed according to specific needs: technical summaries extracted from the Tools Observatory, general information on the tools main functions, step-by-step usage guidelines, testing reports about the tools functionalities. Entering also into the scope of the Catalogue is the referencing of existing training materials for specific tools. The SIG Tools will provide a first example of such training material with a draft of general guidelines for choosing publishing tools targeting junior researchers.³⁸

10.4. Method: Workflows Guidelines

Providing guidance to authors, libraries, and publishers with main publishing possible routes

As often mentioned by the various publications and repeated in this very paper, a lack of global coordination strongly hinders the building of a healthy and robust open infrastructure for scholarly communication. One possible way to address this challenge is by providing a clear representation of the main possible types of publishing workflows, addressing the transitional nature of the scholarly publishing environment, especially in the SSH. Schematically, the representation can distinguish between the autonomous self-built workflow of the expert user, the workflow supported by an institutional publisher, and the workflow supported by a publisher for a fee. Such a representation of

³⁷ See [Annex II](#) for a provisional list of criteria.

³⁸ See [Annex IV](#) for a draft of guidelines.

possible workflows defines different kinds of users and needs which would help to design the training material and the tools' observatory.

10.5. Community infrastructure: Bringing the pieces together

Bringing together the components of a flexible and open community infrastructure

The final recommendation is more a general goal and is provided by the sum of the previous ones put in a more global perspective. The recommendations aim at providing the components of an open infrastructure which remains flexible by paying attention to the different types of users and usage. In order to do so, the recommendations also try to take into account the various inputs coming from the cited reports, adapting them to the scholarly communication practices in the SSH. Between the quickly evolving context and the actual state of affairs, the general objective of these recommendations is to accompany the transition towards an open, coordinated, and sustainable digital infrastructure led by the community.

11. CONCLUSION

By considering the entire research lifecycle from discovery to dissemination, addressing the data citation issue, and putting the scientific conversation at the center of research activity, scholarly communication goes beyond the traditional scholarly publishing model. Digital technologies and tools allow to transform well established practices inherited from a long history and related to specific objects (books, articles) and workflows. The development of preprints and post-publication Open Peer Review are examples of an actual profound transformation of the publishing model through innovative services.

Nevertheless, traditional objects, workflows, and actors still have an important role to play. Journals and books offer to the research communities a place where to discuss specific objects or methodologies. In the SSH context, the publishers also often offer the crucial opportunity to exchange and publish in other languages than in English, while repositories are generally managed in English and by English-speakers.

The role of communities is central in producing research, and it may require both technical innovations and social changes to be fully empowered. The growing success of Peer Community In is one example of such an approach, giving back to the communities the control over the process of certification. Another example is the success of tools like [Jupyter](#) Notebook, which was developed by the communities of researchers and, therefore, met their needs (Whitehouse, 2019). In fact, more generally, communities often appear and grow also around a specific tool or a technology, and this increases even more the necessity to have—and keep—community-led tools.

As we stressed in this report, and in reference to other works (The OA Diamond Journals Study, Future of Scholarly Communication for instance), it is important for all stakeholders (funders, infrastructures and providers of publishing services, etc.) involved in open scholarly communication to support communities and to prevent the risk of lock-in by building community-driven tools. It is also about understanding needs, motivations or barriers of users, supporting practices changes by

training to make innovation efficient. In this prospect, research communities are not only composed of researchers but include the other actors (or users from the tools point of view) that we identified in this report, all working together to finally make knowledge available to all citizens.

ANNEXES

Annex I. List of tools mentioned in this paper

The table focuses on the tools related to contents' production. Most third-party services used for dissemination are not listed in the table.

| NAME AND WEBSITE | OPEN SOURCE | AUTHORING | PEER REVIEWING | PUBLISHING | COMMUNICATING |
|---------------------------------|-------------|-----------|--------------------|------------|---------------|
| AJE | No | authoring | | | |
| authorcafé | No | authoring | | | |
| authorea | No | authoring | | publishing | |
| B2NOTE | Yes | | | | annotations |
| BibSonomy | No | | | | |
| DeepI | No | authoring | | | |
| EditorialManger | No | | | publishing | |
| F1000 research | No | authoring | peer review (open) | publishing | |
| FidusWriter | Yes | authoring | | | |
| Fulcrum | Yes | | | publishing | |
| Getliner | No | | | | annotations |
| Google Docs | No | authoring | | | |
| Hypothes.is | Yes | authoring | | | annotations |
| Hyrax | Yes | | | publishing | |
| Janeway | Yes | | | publishing | |
| Jupyter | Yes | authoring | | publishing | |
| Knora | Yes | | | | annotations |
| LeanPub | No | authoring | | publishing | |

| | | | | | |
|--|-----|-----------|------------------------|------------|---------------|
| Libero (elife Continuum) | Yes | | | publishing | |
| LodeI | Yes | | | publishing | |
| Manifold | Yes | authoring | | | |
| Manuscripts App | Yes | authoring | | | |
| Muruca | Yes | authoring | | publishing | communicating |
| Notesalong | No | | | | annotations |
| OpenReview wToolkit | Yes | | peer review (open) | | |
| overleaf | No | authoring | | | |
| Pandoc | Yes | authoring | | | |
| Paperhive | No | | | | annotations |
| Peer Community in | No | | peer review (open) | | |
| PeerageOfScience | No | | peer review (tracking) | | |
| PKP (OJS, OMP) | Yes | | | publishing | |
| PressBooks | Yes | authoring | | publishing | annotations |
| ProseMirror | Yes | authoring | | | |
| pubfactory | No | | | publishing | |
| Publons | No | | peer review (tracking) | | |
| PubPub | Yes | authoring | | publishing | |
| PubSweet (Editoria) | Yes | authoring | | publishing | |
| Pundit | Yes | | | | annotations |

| | | | | | |
|-----------------------------------|-----|-----------|--------------------|------------|-------------|
| recite | No | authoring | | | |
| Remarg | No | | | | annotations |
| Rua | Yes | | | publishing | |
| ScholarOne | No | | | publishing | |
| scholastica | No | | | publishing | |
| Scielo | Yes | | | publishing | |
| Science _Open | No | | peer review (post) | | |
| sciflow | No | authoring | | | |
| Stylo | Yes | authoring | | | |
| Weava | No | | | | annotations |

Annex II. List of criteria and features for analyzing tools

Basic criteria for all

1. General
 - a. name
 - b. homepage
 - c. description
 - d. ownership
 - e. Business model
2. Scope
 - a. books
 - b. journals
 - c. blogs
 - d. (structured research) data
 - e. multimedia
 - f. others
3. Openness
 - a. Code
 - b. Data
 - c. Standards
 - d. License
4. Type
 - a. service
 - b. component
 - c. application
5. Subtype
 - a. RTU = A **ready-to-use** software as a service fulfilling one or more functions of the scholarly communication activities (eg. Texture).
 - b. TB = A **technical brick** (software libraries or frameworks) which still needs to be adapted and / or integrated into a wider application to be utilised
 - c. CONF = A software application which still **needs to be configured**, installed and maintained.
6. Pricing
 - a. Free (no limitations)
 - b. Free with limitations (describe)
 - c. Fee (describe)
7. Technical requirements
 - a. Operating system
 - b. Programming language
 - c. Required additional software
 - d. Pivotal format
 - e. Other import/export formats
 - f. Technical support (manuals, tutorials, forums etc.)
8. Publication

- a. First stable version (version + date)
 - b. Last stable version (version + date)
 - c. Number of published versions in between
 - d. How many downloads all together
 - e. Who are the biggest users (community)
9. ...

Authoring tools additional criteria

1. Does the tool allow off-line synchronisation?
2. Can you work with:
 - a. pictures
 - b. mathematical formulas
 - c. tables
 - d. graphs
 - e. multimedia content
 - f. non-latin scripts
3. Does the tool have additional editing functions (sorting, indexing, TOC etc.)
4. ...

Publishing tools additional criteria

1. Webpage
 - a. Search
 - b. Browse (by type, by date, by collection, by most popular, by similar items etc)
 - c. Entry Display (including support for different file formats)
 - d. Social Networking and Collaboration Tools
 - e. Other Plug-in Tools
 - f. Personalization and Custom Publishing
 - g. Distribution
 - h. Metadata (DOI, ORCID, keywords, resources, licencing and usage rights etc.)
2. Workflow
 - a. Automated emails and reminders
 - b. Consents / Agreements with the authors and reviewers
 - c. GDPR Consent
3. Administration
 - a. E-commerce integration
 - b. Revenue Model Support
 - c. Access Control
 - d. CMS
 - e. Content Ingestion / Publication Management
 - f. Library Features
 - g. Reporting
 - h. Digital preservation
4. Dissemination: indexes and other channels
5. Output formats

- a. ePub
- b. PDF
- c. XHTML
- d. JATS
- e. TEI

Annex III. Analytical table of annotation tools

Goal: providing an assessment of (web) annotation tools that can be used to apply digital marginalia on content published on one of the publishing tools mentioned in the SIG document. This excludes “closed” tools like Evernote for example, where the content must be imported in the tool before applying annotations on them.

A recap of the work that has been done:

- tools that were analysed:
 - All tools in the original version of the OPERAS tools white paper
 - Hypothesis
 - Colwiz
 - Paperhive
 - Remarq
 - Pundit
 - Bibsonomy
 - Other tools that were considered:
 - B2NOTE
 - Knora.org
 - Weava: <https://www.weavatools.com/>
 - GetLiner: <https://getliner.com/>
 - Notesalong: <https://notesalong.com/>
- Not assessed/discarded
 - Tools specialised to tag data for machine learning tasks
 - Prodigy: <https://prodi.gy/>
 - Label Studio: <https://labelstud.io/>
 - Not proper generic web annotation tools
 - Brat: <http://brat.nlplab.org/>

- This is not a generic web annotation tool, but a desktop application to apply linguistic annotations on text. Also the web version doesn't work, possibly because the development seems to have been stopped long ago.
- WebAnno: <https://webanno.github.io/webanno/>
 - A web based annotation software for linguistic annotations. Not a real web tool: You should either install the system on your PC or on a server. Also, specific document formats are necessary to apply (linguistic) annotations. See https://webanno.github.io/webanno/releases/3.4.5/docs/user-guide.html#sect_formats
- Recogito: <https://recogito.pelagios.org/>
 - Albeit Recogito is a beautiful service (it received several recognitions, it supports semantic annotations, etc) it cannot be considered a general purpose web annotation tool. Content to be annotated (only text files and images) must be first imported into the platform.

Results were divided in two tables, the Part 1 with the tools closer to the scope of the paper, and Part 2 with those which seem to be out of focus of this study. In this paper, therefore, we present only the Part 1.

| | HYPOTHES.IS | PUNDIT | B2NOTE | WEAVA | GETLINER | NOTESALONG | REMARQ LITE |
|---------|---|---|---|---|---|---|---|
| Website | https://web.hypothes.is/ | https://thepund.it | https://b2note.eudat.eu/ | https://www.weavertools.com/ | https://getliner.com/ | https://notesalong.com/ | https://remarkable.com/web/index.html |
| Type | web annotation tool | web annotation tool | Not a generic web annotation tool, since it is integrated with the B2Share service. It has been included here because theoretically the | web annotation tool | web annotation tool | web annotation tool | Web annotation tool plus platform for sharing and discovery users and annotated content. |

| | | | | | | | |
|----------------------------------|---|---|---|---|-----|--------------------------------|-----------------------------------|
| | | | service can also be used via a widget on other web sites. It is also a EOSC-HUB service and it is listed in the EOSC marketplace. | | | | |
| Annotator Functionalities | | | | | | | |
| Highlighting | Yes | Yes | No | Yes | Yes | Yes | Yes |
| Different highlight colors | No <i>(still planned?)</i> | No in the current version Yes in the new one in the making | No | Yes | Yes | Yes | No |
| Commenting | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Tagging | Yes. Free text tags can be associated to an annotation | No in the current version Yes in the new one in the making | Yes An annotation can be a free text tag. | Yes, tags can be associated to a highlight color and organized in folders.. | Yes | Yes. also support for emoticon | Yes (private for the single user) |
| Annotating a text fragment | Yes | Yes | No | Yes | Yes | Yes | Yes |
| Annotating a whole web | Yes (they are called "Page | Yes | Yes, but only if published in | Yes | Yes | No | No |

| | | | | | | | |
|---|---------|----------------------------|---|----------|---|---|-------------|
| Page/resource | notes") | | B2Share (resource description page plus attachments) or if the content provider integrates B2Note in the web site | | | | |
| Semantic annotations (RDF triples in annotations) | No | Yes | Yes but limited: only link to vocabularies (not known the predicate used). | No | No | No | No |
| Social annotation | Yes | Yes | Yes | Not sure | Yes somehow: annotations can be exported to FB, Twitter, etc. | Yes somehow: annotations can be shared through a link to the inweb.notesalong.com proxy service | Yes |
| Replies | Yes | Yes | No | No | no (?) | no | Yes |
| Personal annotation | Yes | Yes (via Private notebook) | Yes | Yes | yes | yes | Yes |
| Groups | Yes | Yes (Notebooks) | No | Yes | no | no | No but it's |

| | | | | | | | |
|--|---|--|----|-----|-----------------------------------|-----|---|
| (Annotation Container) | | | | | | | possible to publish them in public groups |
| Public discussion | Yes | Yes | No | No | no | no | Yes for public groups. Remarq Lite does not support public comments. This is possible only for those journals and publications that integrate the full version of Remarq. |
| Share an annotation | Yes | No in the current version. Yes in the new one in the making | No | Yes | yes | yes | No |
| Sharing target page with annotations activated | Yes, very effective (via central proxy web service that also redirects to the destination URL if it detects that the Hypothes.is extension is | Yes (via central proxy web service - Feedthepundit) | No | No | yes (couldn't check but possible) | no | No |

| | | | | | | | |
|----------------|---|---|--|--|---|---|--|
| | already activated) | | | | | | |
| Versioning | No | No | No | No | no | no | No |
| Direct links | Yes | No in the current version Yes in the new one in the making | No | No | Yes | Yes, through the inweb.notesalongo.com proxy service | No |
| HTML support | Yes | Yes | Limited. Only B2Share resources page | Yes | Yes | yes | Yes |
| PDF support | Yes | No in the current version Yes in the new one in the making | No | Yes | Yes | yes | Yes |
| EPUB support | Yes | No | No | No | no | no | No |
| Mobile support | Limited: possible through the proxy service via.hypothes.is. Decent on a tablet, quite impossible to use on a mobile phone. | No | Yes when embedded in the B2Share service, since the latter supports responsive design. | Yes but only through a dedicated iOS App | Yes but only through a specific App and a dedicated web browser | Limited: annotations can be viewed on a mobile device through the inweb.notesalongo.com proxy service | Apparently not this version (Remarq Lite) but the commercial edition, that can be embedded on publishers' sites, supports responsive |

| | | | | | | | |
|---------------------------------|---|-----|---|----|----|----|--|
| | | | | | | | design. |
| Annotate over publisher content | Yes | Yes | Yes but only if publisher includes it in the content page | No | No | No | Yes |
| Publisher Moderation | Yes | No | No | No | no | no | |
| HTML<>PDF cross format | Yes | No | No | No | no | no | Possibly with the full version and not the Lite one (see https://remarkable.com/web/faq.html) |
| DOI support (in the HTML pages) | Yes | No | No | No | no | no | |
| Markdown/ Rich text format | Yes | No | No | No | no | no | Yes |
| Math support | Yes (via LaTeX) | No | No | No | no | no | Yes |
| Rich media | Yes (It is possible to insert an image in the annotation; | No | No | No | no | no | Yes, images and video |

| | | | | | | | |
|------------------|---|---|---|---|--------------------------------|--|--|
| | <i>apparently also audio files can be added in an annotation)</i> | | | | | | |
| Follow | No <i>(still planned?)</i> | | | | | | Yes (groups) |
| Social Login | No <i>(still planned?)</i> | Yes Facebook, Google. Also EGI AAI Check-In | Yes, through B2Access and OpenAire | Yes (Google) | Yes (Google) | no | Yes (Apple, Google and Orcid) |
| Image Annotation | No <i>(still planned?)</i> | No | No | No | no | no | No |
| Search | Yes | Yes. "Filter annotations" option | Yes | Yes | yes (tag and color filters) | yes yes (tag, color filters and folders) | Yes (people, groups, annotations, articles) |
| Advanced search | No? | Yes. Annotations can be filtered by several attributes, including author, notebook, and date | Yes. Annotations can be filtered by those associated to all files of the present resource, type, public/only owned by user. | Yes, color and associated tag filter. | idem | idem | Not a real advanced search but a discovery feature is available. |
| Indexed | | | | No | no | no | |

| | | | | | | | |
|--|--|---|--|---------------------------|------------------------------------|------------------------------|---|
| (Crossref Event Data) | | | | | | | |
| Activity Feed/ Page <i>(lists of user's activities, e.g. annotations done)</i> | | | | No | no | no | Yes: it is possible to follow users and groups. |
| Service characteristics | | | | | | | |
| Centralized dashboard/App | Yes <a href="https://hypothes.is/users/<user_name>">(https://hypothes.is/users/<user_name> . Quite basic apparently) | Yes (https://thepund.it/app . View notebooks and annotations, search, filtering - faceted, annotation export via file and API). The new version is accessible at https://app.thepund.it | No. Apparently annotations can be seen only in the annotated B2Share page | Yes | yes (user private dashboard) | yes (user private dashboard) | Yes |
| Exporting the annotations | No apparently (only via coding/APIs) | Yes those of a Notebook via file (XSLX, DOCX, ODT, JSON-LD/W3C Web Annotation | Yes, those of a B2Share resource, via file (JSON-LD, RDF/XML, RDF/Turtle) | Yes (word, txt, csv, xls) | yes (word, onenote, evernote, txt) | no | No apparently |

| | | | | | | | |
|---|---|--|---|-------------------------|-------------------------|-----------------------|--|
| | | Data Model) or ready to use API endpoint but only of entire notebooks (no filtering) | | | | | |
| Works everywhere | Yes | Yes | No, only web documents published through B2Share | Yes (browser extension) | yes (browser extension) | no (chrome extension) | Only for personal notes |
| API | Yes | Yes | Yes. See https://b2note.docs.apiary.io/ | No ? | no (?) | no | No |
| Supported Hosting Type | Centralized cloud based for the Public Service; Self hosted for the open source version. | Centralized cloud based for the Public Service; Self hosted for the open source version but current version is very hard to install. The new one should solve this issue. | Centralized cloud based for the Public Service; Self hosted for the open source version. | not specified | ? | ? | Centralized cloud based service |
| Customization to fit publisher platform | Yes | Yes. The new version has also a Wordpress | Yes | No | no | no | Yes, see https://remarque.com/web/fa |

| | | | | | | | |
|------------------------------------|---|--|---|---------------|---------------|---------------|---|
| | | plug-in (seen on the OPERAS web site) | | | | | q.html |
| Ecosystem & maintainer | | | | | | | |
| W3C standard - data model (W3C WA) | ? (they declared "Yes" in the previous version of the SIG Tools document. Actually the public API doesn't seem compliant to W3C WA. See: for example https://hypothes.is/api/search and https://www.w3.org/TR/annotation-model/) | Partial support in the current version Full support for the Web Annotation Data Model in the new one for some public APIs | Yes, when exporting the annotations of a resource | not specified | not specified | not specified | In the old OPERAS document it was stated as "Claimed" but no evidence can be found. |
| W3C standard - protocol | No. Declared in progress in the previous version of the document but no evidence about that in their web site | No. | No | not specified | not specified | not specified | In the old OPERAS document it was stated as "No" but no evidence can be found. |

| | | | | | | | |
|-----------------------------|--|---------------------------------|--|-------------------------------|-------------------------------|---|---|
| Open source | Yes | Yes | Yes https://github.com/EUDAT-B2NOTE/b2note | No | No | No | No |
| <i>Documentation</i> | Very active blog for users, APIs documentation | Only (neglected) blog for users | Help on line | poor (FAQ and tutorials) | poor (a FAQ and a forum) | no | No documentation available |
| Non-profit (Maintainer) | Yes Hypothesis is a 501 non-profit organization. | No (<i>Net7 is a SME</i>) | Yes. It is a service maintained by the Eudat infrastructure | no (free version and premium) | no | not specified (extension is free) | No |
| Annotation License (Public) | Yes CC0 | No | No | No | No | No | |
| Member of AAK coalition | Yes | Yes | No | No | No | No | Yes (Redlink - the previous version) |
| Type of support provided | Mailing-list, Google Forum, Public chat Slack channel, direct mail to developers | Poor: direct mail to developers | Limited: Through an on line form (Eudat.eu support service) | Online form. | Online forum, not very active | assistance through the google extensions platform | Apparently no support in the Lite (free) version. |

Annex IV. Training material draft

The SIG tool member, CNRS Médiçi network, worked on a draft for first-step information material on scholarly publishing tools. Below are examples of infoboxes which could be integrated in a synthetic leaflet.



AUTHOREA

- Name: Authorea
- URL: <https://www.authorea.com/>
- Developed by: Authorea, New Jersey
- Category: Proprietary
- Type: service
- Purpose: authoring, publishing



- Name: Editoria
- URL: <https://editoria.pub/>
- Developed by: Coko Foundation, San Francisco
- Category: Open
- Type: service
- Purpose: book publishing

✱



- Name: Manifold Scholarship
- URL: <https://manifoldapp.org/>
- Developed by: University of Minnesota Press and the CUNY Graduate Center
- Category: Open
- Type: Service
- Purpose: authoring



- Name: Manuscripts
- URL: <https://www.manuscriptsapp.com/>
- Developed by: Atypon
- Category: Proprietary
- Type: Application
- Purpose: authoring

✱

REFERENCES

- '5 Scholarly Publishing Trends to Watch in 2020', Scholastica Blog, 2020
 <<https://blog.scholasticahq.com/post/scholarly-publishing-trends-to-watch/>>
- Baker, Stewart, 'Assessing Open Source Journal Management Software', The Journal of Electronic Publishing, 23.1 (2020) <<https://doi.org/10.3998/3336451.0023.101>>
- Baligand, Marie-Pascale, Colcanap, Grégory, Harnais, Vincent, Rousseau-Hans, Françoise, Weil-Miko, Christine, 'Les pratiques de recherche documentaire des chercheurs français en 2020 : étude du consortium Couperin' [Technical Report], Rapport Couperin N°2, 2021, <[hal-03148285](https://hal.archives-ouvertes.fr/hal-03148285)>
- Bilder, Geoffrey, Jennifer Lin, and Cameron Neylon, 'Principles for Open Scholarly Infrastructures-V1', 2015, 35186 Bytes <<https://doi.org/10.6084/M9.FIGSHARE.1314859>>
- Bosman, Jeroen, and Bianca Kramer, '101 Innovations in Scholarly Communication: How Researchers Are Getting to Grip with the Myriad New Tools', LSE Impact Blog, 2015
 <<https://blogs.lse.ac.uk/impactofsocialsciences/2015/11/11/101-innovations-in-scholarly-communication/>>
- Butchard, Dorothy, Simon Rowberry, Claire Squires, and Gill Tasker, 'Peer Review in Practice', in Academic Book of the Future: BOOC (UCL Press, 2017)
 <<https://doi.org/10.14324/111.9781911307679.15>>
- Checco, Alessandro, Loreti, Pierpaolo, and Pinfield, Stephen, 'AI-Assisted Peer Review', Humanities & Social Sciences Communications; London, 8.1 (2021)
 <<https://doi.org/10.1057/s41599-020-00703-8>>
- Eriksson, Jörgen, Christer Lagvik, and Emma Nolin, 'Moving towards Open Science? Conference Report: The 9th Conference on Open Access Scholarly Publishing, Lisbon, September 20–21, 2017', Nordic Perspectives on Open Science, 1 (2018) <<https://doi.org/10.7557/11.4307>>
- European Commission. Directorate General for Research and Innovation., Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission. (LU: Publications Office, 2019) <<https://data.europa.eu/doi/10.2777/836532>>
- Fecher, Benedikt, and Tony Ross-Hellauer, 'Journal Flipping or a Public Open Access Infrastructure? What Kind of Open Access Future Do We Want?', LSE Impact Blog, 2017
 <<https://blogs.lse.ac.uk/impactofsocialsciences/2017/10/26/journal-flipping-or-a-public-open-access-infrastructure-what-kind-of-open-access-future-do-we-want/>>
- Foltýnek, T., Dlabolová, D., Anohina-Naumeca, A. et al. 'Testing of support tools for plagiarism detection', Int J Educ Technol High Educ 17, 46, 2020,
 <<https://doi.org/10.1186/s41239-020-00192-4>>
- Langlais, Pierre-Carl, Critical Study of the New Ways of "Editorialising" Open Access Scientific Journals (Bibliothèque Scientifique Numérique, November 2016)
 <<https://hal.archives-ouvertes.fr/hal-01399286>>

- Lawrence, Rebecca, 'What's next for Peer Review?', Research Information, 2016
<<https://www.researchinformation.info/feature/whats-next-peer-review>>
- Lewis, David W., A Bibliographic Scan of Digital Scholarly Communication Infrastructure | Educopia Institute (Atlanta, Georgia: Educopia Institute, 2020)
<https://educopia.org/wp-content/uploads/2020/05/Lewis-Bibliographic-Scan_FULL.pdf>
- Maxwell, John W., Erik Hanson, Leena Desai, Carmen Tiampo, Kim O'Donnell, Avvai Ketheeswaran, and others, Mind the Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms, 1st edn (PubPub, 2019) <<https://doi.org/10.21428/6bc8b38c.2e2f6c3f>>
- Michael, Ann, 'Open Access Technology Options', Scholarly Kitchen, 2018
<<https://scholarlykitchen.sspnet.org/2018/02/22/open-access-technology-options/>>
- Peters, Paul, 'An Open Approach to Developing Infrastructure for Open Science', Hindawi Blog, 2017
<<https://www.hindawi.com/post/a-radically-open-approach-to-developing-infrastructure-for-open-science/>>
- Pooley, Jefferson, 'Scholarly Communications Shouldn't Just Be Open, but Non-Profit Too', LSE Impact Blog, 2017
<<http://blogs.lse.ac.uk/impactofsocialsciences/2017/08/15/scholarly-communications-shouldnt-just-be-open-but-non-profit-too/>>
- Ross-Hellauer, Tony, 'What Is Open Peer Review? A Systematic Review', F1000Research, 6 (2017), 588
<<https://doi.org/10.12688/f1000research.11369.2>>
- Sauret, Nicolas, 'De la revue au collectif : la conversation comme dispositif d'éditorialisation des communautés savantes en lettres et sciences humaines', PHD thesis, Université Paris Nanterre, Université de Montréal, 2020, <<https://these.nicolassauret.net>>
- Scholastica et al., 'Democratizing Academic Journals: Technology, Services, and Open Access', Copyright, Fair Use, Scholarly Communication, Etc., 42, 2017
<<https://s3.amazonaws.com/marketing.scholasticahq.com/Democratizing-Journal-Pub-WP.pdf>>
- , 'Why Academic-Led Journal Publishing? Liberating Research Through Tools and Services', Scholastica Blog, 2018
<<https://blog.scholasticahq.com/post/academic-led-journal-publishing-liberating-research-tools-services/>>
- Schonfeld, Roger C., 'Publishers Invest in Preprints', The Scholarly Kitchen, 2020
<<https://scholarlykitchen.sspnet.org/2020/05/27/publishers-invest-in-preprints/>>
- , 'Publishers Still Don't Prioritize Researchers', The Scholarly Kitchen, 2021
<<https://scholarlykitchen.sspnet.org/2021/01/26/publishers-fail/>>
- , 'Workflow Lock-in: A Taxonomy', The Scholarly Kitchen, 2018
<<https://scholarlykitchen.sspnet.org/2018/01/02/workflow-lock-taxonomy/>>

- Skinner, Katherine, Mapping the Scholarly Communication Landscape – 2019 Census (Atlanta, Georgia: Educopia Institute, 2019) <<https://educopia.org/2019-census/>>
- Tattersall, Andy, 'Comment, Discuss, Review: An Essential Guide to Post-Publication Review Sites.', 2014
<<https://blogs.lse.ac.uk/impactofsocialsciences/2014/11/08/comment-discuss-review-an-essential-guide/>>
- Tech Trends 2024, 2020
<<https://www.stm-assoc.org/standards-technology/stm-tech-trends-2024-focus-on-the-user-connect-the-dots/>>
- Tennant, Jonathan P., Jonathan M. Dugan, Daniel Graziotin, Damien C. Jacques, François Waldner, Daniel Mietchen, and others, 'A Multi-Disciplinary Perspective on Emergent and Future Innovations in Peer Review', F1000Research, 6 (2017), 1151
<<https://doi.org/10.12688/f1000research.12037.3>>
- Tulley, Christine, 'Guest Post — Emerging Trends in the Academic Publishing Lifecycle', The Scholarly Kitchen, 2019
<<https://scholarlykitchen.sspnet.org/2019/03/27/guest-post-emerging-trends-in-the-academic-publishing-lifecycle/>>
- Whitehouse, Tyler, 'Guest Post — A Look at the User-Centric Future of Academic Research Software — And Why It Matters, Part 1: Trends', The Scholarly Kitchen, 2019
<<https://scholarlykitchen.sspnet.org/2019/10/07/guest-post-a-look-at-the-user-centric-future-of-academic-research-software-and-why-it-matters-part-1-trends/>>; 'Part 2: Implications'
<<https://scholarlykitchen.sspnet.org/2019/10/08/guest-post-a-look-at-the-user-centric-future-of-academic-research-software-and-why-it-matters-part-2-implications/>>