

D3.2

Report on challenges and help measures faced by OA journals and platforms

Submission Date 2023.12.22	Version 2.0 – Approved by EC		
Due Date 2023.12.31	PU	Public	X
	SEN	Sensitive	
	R-UE/UE-R	EU classified	
Deliverable Title	Report on challenges and help measures faced by OA journals and platforms		
Deliverable No.	D3.2		
Lead beneficiary	TSV		
Contributing WP	WP3		
Type	Report		
	HORIZON-INFRA-2022-EOSC-01 Grant Agreement: 101094397		

Project Full Title	Creating a Robust Accessible Federated Technology for Open Access
Project Acronym	CRAFT-OA
Project No.	101094397
Start Date	2023.01.01
End Date	2025.12.31
Duration	36 Months
Project Website	https://craft-oa.eu
Authors	Mikael Laakso (https://orcid.org/0000-0003-3951-7990), Xenia van Edig (https://orcid.org/0000-0003-2510-0529), Jorina Fenner (https://orcid.org/0009-0001-0469-3110), Leonidas Pispiringas (https://orcid.org/0000-0002-7111-094X), Arnaud Gingold (https://orcid.org/0009-0000-0466-5542), Barbara Svetina (https://orcid.org/0000-0002-9636-4102), Clara Armengou (https://orcid.org/0000-0003-2124-5165)
Abstract	Initiatives like CRAFT-OA and DIAMAS aim to enhance Open Access Diamond Journals' (OADJs) capacities through digital solutions, yet the technological landscape remains heterogeneous. This deliverable conducts a gap analysis, assessing the alignment of OADJs with recommended technical best practices outlined in the Report on standards for best publishing practices and basic technical requirements in the light of FAIR principles (Armengou et al., 2023). The analysis employs a versatile approach, combining quantitative analysis of various datasets and desk research on studies and reports. Case studies of European platforms further enrich the evaluation. The assessment identifies technical gaps and successes and lacks a comprehensive OADJs registry that the CRAFT-OA project will address through solutions like the Diamond Discovery Hub.

Version and Revision History

Version	Date	Author/Reviewer/Contributors	Comments
0.1	2023.06.26	Authors: Mikael Laakso (TSV), Xenia van Edig (TIB), Jorina Fenner (TIB)	First draft
0.2	2023.11.30	Authors: Mikael Laakso (TSV), Clara Armengou (DOAJ), Xenia van Edig (TIB), Jorina Fenner (TIB), Arnaud Gingold (AMU), Leonidas Pispiringas (OpenAIRE AMKE), Barbara Svetina (ZRC SAZU) Contributors: Andy Byers (Janeway), Ljiljana Jertec Musap (SRCE), Tabea Klaus (UGOE), Ignacio Lamata Martinez (EGI), Antti-Jussi Nygård (TSV), Janne Pölönen (TSV), Cezary Rosiński (IBL PAN), Nikodem Wołczuk (IBL PAN)	Version sent to reviewers
0.3	2023.12.10	Reviewers: Sy Holsinger (OPERAS), Aysa Ekanger (UiT)	Review
0.4	2023.12.15	Authors: Mikael Laakso (TSV), Xenia van Edig (TIB), Jorina Fenner (TIB), Leonidas Pispiringas (OpenAIRE AMKE), Arnaud Gingold (AMU), Barbara Svetina (ZRC SAZU), Clara Armengou (DOAJ)	Final text, handed over for final formal check
0.5	2023.12.21	Contributors: Tabea Klaus (UGOE), Theresa Waldmann (UGOE)	Final formal check
1.0	2023.12.22		Submitted version
1.1	2025.01.09	Contributors: Lisa Müller (UGOE), Theresa Waldmann (UGOE)	Correction of formal mistakes, e.g., spelling mistakes
2.0	2025.01.09		Approved by EC



CRAFT-OA is funded by the European Union under Grant Agreement no. 101094397. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.



This deliverable is licensed under a Creative Commons Attribution 4.0 International License.

List of Acronyms

AAI	Authentication & Authorisation Infrastructure
AAM	Author Accepted Manuscript
AHCI	Arts & Humanities Citation Index
AMU	Aix-Marseille Université
APC	Article Processing Charge
API	Application Programming Interface
ARIS	Slovenian Research and Innovation Agency
ARK	Archival Resource Key
AVOTT	National Open Science Coordination
CC	Creative Commons
CC0	Creative Commons Zero
CC BY	Creative Commons Attribution
CC BY-SA	Creative Commons Attribution Share-Alike
CC BY-NC-ND	Creative Commons Attribution NonCommercial NoDerivates
CERIF	Common European Research Information Format
CINES	Centre Informatique National de l'Enseignement Supérieur

CLOCKSS	Controlled LOCKSS (Lots of Copies Keep Stuff Safe)
CNRI	Corporation for National Research Initiatives
COAR	Confederation of Open Access Repositories
CoARA	Coalition for Advancing Research Assessment
COUNTER	Counting Online Usage of NeTworked Electronic Resources
CROASC	Croatian Association for Scholarly Communication
CSV	Comma-Separated Values
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
DDH	Diamond Discovery Hub
DIAMAS	Developing Institutional Open Access Publishing Models to Advance Scholarly Communication
dLib	Digital Library of Slovenia
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identifier
eISSN	electronic International Standard Serial Number
EOSC	European Open Science Cloud

ePub	electronic Publication
EQSIP	Extensible Quality Standard for Institutional Publishing
ESCI	Emerging Sources Citation Index
FAIR	Findable, Accessible, Interoperable, and Reusable
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
GWGDG	Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen
HTTP	Hypertext Transfer Protocol
HTML	Hypertext Markup Language
IBL PAN	Instytut Badan Literackich Polskiej Akademii Nauk
IPSP	Institutional Publishing Service Provider
IPTP	Institutional Publishing Technology Providers
ISSN	International Standard Serial Number
JASPER	Journals are Preserved forever
JATS XML	Journal Article Tag Suite Extensible Markup Language
JSON	JavaScript Object Notation
KBART	Knowledge Bases and Related Tools

LOCKSS	Lots of Copies Keep Stuff Safe
MARC	Machine-Readable Cataloging
MODS	Metadata Object Description Schema
MPA	Multi Primary Administrator
NISO	National Information Standards Organisation
OA	Open Access
OADJ	Open Access Diamond Journal
OADJ Study	Open Access Diamond Journals Study
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OASPA	Open Access Scholarly Publishers Association
OE	OpenEdition
OEJ	OpenEdition Journal
OJS	Open Journal Systems
OMP	Open Monograph Press
ONIX	Online Information Exchange
OpenAIRE	Open Access Infrastructure for Research in Europe
OPS	Open Preprint Systems

ORCID	Open Researcher and Contributor Identifier
PID	Persistent Identifier
PKP	Public Knowledge Project
PDF	Portable Document Format
ROR	Research Organization Registry
RSS	Rich Site Summary
SAML	Security Assertion Markup Language
SCIE	Science Citation Index Expanded
SEO	Search Engine Optimization
SciELO	Scientific Electronic Library Online
SRCE	University of Zagreb, University Computing Centre
SSH	Social Sciences and Humanities
T	Task
TDM	Text and Data Mining
TEI	Text Encoding Initiative
TIB	Leibniz Information Centre for Science and Technology and University Library
TSV	The Federation of Finnish Learned Societies

UGOE	Georg-August-Universität Göttingen
UiT	The Arctic University of Norway
WoS	Web of Science
WP	Work Package
XML	Extensible Markup Language
ZDB	Zeitschriftendatenbank
ZRC SAZU	Scientific Research Centre of the Slovenian Academy of Sciences and Arts

Table of Content

1	Executive Summary	14
2	Introduction	16
2.1	Aims and scope	16
2.2	Background	16
3	Data sources.....	20
3.1	DOAJ data	20
3.1.1	Rejected DOAJ applications	20
3.1.2	Public journal-level data of OADJs currently included in the DOAJ	22
3.1.3	Article-level metadata of OADJs making such data available.....	22
3.2	Diamond OA Study	23
3.3	Data from the DIAMAS project.....	24
3.3.1	DIAMAS Web survey and associated landscape analysis.....	24
3.3.2	DIAMAS Gap analysis	24
3.4	Technical support for standards among open source publishing content management systems.....	25
3.5	Data for future work.....	25
4	Gap analysis	26
4.1	Identifiers	27
4.1.1	ISSN/EISSN	28
4.1.2	ORCID identifier	30
4.1.3	ROR identifier.....	32
4.1.4	ARK identifier	34
4.1.5	Handle identifier	36
4.1.6	DOI identifier	38
4.1.7	Funder DOIs/PIDs	41
4.2	Metadata	44
4.2.1	Metadata exchange for harvesting (e.g. Dublin Core, OpenAIRE, OAI-PMH s etc.) 46	
4.2.2	OpenAIRE Guidelines	49

4.2.3	Mass metadata export (as CSV files, ONIX XML feeds or in any other established format)	51
4.2.4	KBART	52
4.2.5	Metadata about OA status	55
4.2.6	Registering of a self-archiving policy	56
4.2.7	Direct deposition in an OA repository	58
4.2.8	Metadata under CC0	59
4.2.9	Open Citations" standard compliance	61
4.3	Content	64
4.3.1	Human- and machine-readable information about the open access status, copyright holder and licensing in each publication in a standard non-proprietary format	64
4.3.2	Full text in machine-readable format (JATS XML or equivalent (e.g. TEI))	66
4.3.3	Text and Data Mining is technically supported	68
4.3.4	Deposited in a digital preservation service	71
4.4	Website features	74
4.4.1	SEO	75
4.4.2	Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review (where relevant)	76
4.4.3	Unique URLs for landing pages	79
4.4.4	URLs linking to related research objects	80
4.4.5	COUNTER	81
5	Case studies	84
5.1	Journal.fi	84
5.1.1	Introduction	84
5.1.2	Background	84
5.1.3	Journal.fi service	85
5.1.4	Compliance with technical standards and their adoption among journals	88
5.1.5	Conclusion	89
5.2	HRČAK – Portal of Croatian scientific and professional journals	93
5.2.1	Introduction	93
5.2.2	Scope and mode of operation	93

5.2.3	Compliance with technical standards and their adoption among journals.....	94
5.2.4	Conclusion	96
5.3	ZRC SAZU journals	99
5.3.1	Introduction	99
5.3.2	Compliance with technical standards and their adoption	101
5.3.3	Conclusion	102
5.4	OpenEdition	104
5.4.1	Introduction	104
5.4.2	Context	105
5.4.3	OEJ journals on DOAJ: a tentative assessment.....	105
6	Discussion	110
7	Conclusions and outlook.....	112
8	List of References	114
9	List of Websites	117
10	List of Figures	120
11	List of Tables	121
12	Annex.....	122

1 EXECUTIVE SUMMARY

Open Access Diamond Journals (OADJs) play an essential and strategic role in the Open Science landscape, supporting the free circulation of knowledge in contemporary societies. Numerous current initiatives aim to strengthen OADJs' capacities, including the CRAFT-OA and DIAMAS¹ projects, which contribute to the creation and maintenance of robust and sustainable research outlets based on modern digital solutions. Currently, the landscape of OADJs is technologically heterogeneous and distributed, with some OADJs facing challenges in meeting the recommended practices in terms of technological standards and services. This deliverable assesses potential gaps between recommended technological best practices and the OADJ's current practices. This gap analysis is performed by assessing the degree of uptake of recognised standards and practices among OADJs based on data from a variety of sources that includes previous surveys on the OADJ landscape and data stemming from the Directory of Open Access Journals (DOAJ).

The standards and practices chosen for closer inspection in the gap analysis were previously identified and described in the "Report on standards for best publishing practices and basic technical requirements in the light of FAIR principles"² and are briefly summarised in this deliverable as well. The methodology to gauge this uptake is twofold: 1) data-based quantitative analysis, and 2) desk research of written studies, papers and reports. Quantitative analysis has been performed mostly on the DOAJ public dataset, anonymised data on rejected DOAJ applications, the data of the Open Access Diamond Journal Study (OADJ Study) and various data generated by the DIAMAS project. This versatile approach of collecting evidence supporting the gap analysis from multiple sources enabled broad coverage for all the included standards and practices. In addition to analysing individual technologies and their gaps in uptake for individual journals, the deliverable also includes four holistic case studies of European journal portals (OpenEdition, Journal.fi, HRČAK, ZRC SAZU) that cater for technologies to many OADJs through the services that they provide. These portal case studies provide a unique perspective on support and adoption of these standards, showcasing how the gaps manifest in different national and technological contexts.

The gap analysis reveals that, in most cases, the uptake for the various recommended standards has substantial room to grow. By observing the various gaps, we could also perceive indications that the gaps are often different: some stem from technological challenges, others from resource/expertise limitations, and, quite often, a lack of awareness. Identifying the

¹ The acronym DIAMAS stands for Developing Institutional Open Access Publishing Models to Advance Scholarly Communication. <https://diamasproject.eu/>

² Armengou, C., Edig, X. van, Laakso, M., & Umerle, T. (2023). *CRAFT-OA Deliverable 3.1 Report on Standards for Best Publishing Practices and Basic Technical Requirements in the Light of FAIR Principles* (under EC review), Zenodo. <https://doi.org/10.5281/zenodo.8112662>

main type of challenge has implications for what kind of training materials and other support would be needed to help close the various gaps. At the same time, it must be emphasised that the lack of a comprehensive registry of OADJs impacts the analysis. This is a gap that the CRAFT-OA project aims to address by delivering solutions such as the Diamond Discovery Hub (DDH).

As part of Task (T) 3.3 of the CRAFT-OA project, we aim to address the identified gaps with a suite of training resources and events specifically tailored to the OADJ audience. These events and training resources will include workshops, a self-assessment checklist, online learning materials and a FAIRification toolkit. The materials will also be available after the project. They will enable journals, platforms, and infrastructures to assess their own “compliance level” with the various standards presented and get help in overcoming these challenges.

In addition to the materials to be produced by T3.3, other Work Packages (WP) also deal with training activities. As all of these activities have the common goal of enhancing the operations of OADJ, they are strongly interconnected. The conclusions of this deliverable will inform both the user documentation and the DDH service.

2 INTRODUCTION

CRAFT-OA deliverable 3.2 conducts a gap analysis, assessing the alignment of Open Access Diamond Journals (OADJs) with recommended technical best practices outlined in the Report on standards for best publishing practices and basic technical requirements in the light of FAIR principles³. The analysis employs a versatile approach, combining quantitative analysis of various datasets and desk research on studies and reports. Case studies of European platforms further enrich the evaluation.

2.1 Aims and scope

In this challenge and gap analysis, we strive to identify obstacles for OADJs and their publishing platforms. We focus on compliance with recommended technical standards, publishing best practices and requirements set by funders (e.g. cOAlition S), indexing services and aggregators (e.g. Directory of Open Access Journals (DOAJ), Open Access Infrastructure for Research in Europe (OpenAIRE)), and other players in the field of scholarly communication⁴. The results of this analysis will feed into identifying suitable training topics (T3.3 “Training and education to enable the adoption and implementation of technical specifications”) to develop materials in order to enhance the operational performance of these journals.

2.2 Background

Open Access (OA) journals, and especially those OA journals operating without article processing charges (APCs), also referred to as OADJs, are very diverse. They cover different languages, disciplines and scholarly communities.⁵ The Open Access Diamond Journal Study (OADJ Study)⁶ estimated that the ecosystem of Diamond OA Journals comprises 17,000 to 29,000 journals, and thus contributes substantially to scholarly communication. OADJs are commonly smaller in terms of article outputs⁷ compared to commercial (both APC-based OA and subscription based) journals.

As outlined in the Report on standards for best publishing practices and technical requirements in Light of the FAIR principles⁸, journals are confronted with various publishing

³ Armengou et al., 2023.

⁴ Cf. Armengou et al., 2023.

⁵ Cf. Ancion, Z., Borrell-Damián, L., Mounier, P., Rooryck, J., & Saenen, B. (2022). *Action Plan for Diamond Open Access*. Zenodo. <https://doi.org/10.5281/zenodo.6282403>

⁶ Bosman, J., Frantsvåg, J. E., & Kramer, B., Langlais, P.-C., & Proudman, V. (2021). *OA Diamond Journals Study. Part 1: Findings*. Zenodo. <https://doi.org/10.5281/zenodo.4558704>

⁷ Frantsvåg, J. E., & Strømme, T. E. (2019). Few Open Access Journals Are Compliant with Plan S. *Publications*, 7(2): 26. <http://dx.doi.org/10.3390/publications7020026>

⁸ Armengou et al., 2023.

best practices in which they need to comply. Keeping up with these recommended technological practices requires planning and continuous effort, which often necessitates some degree of funding. Laakso & Multas⁹ conducted a study covering all individual European countries concerning journals and publishers active in each country, assessing journal counts, publisher sizes and relative shares of OA. They found that European countries are currently publishing 16,387 journals from small- and mid-sized publishers, with 36% of them already adopting an OA model. The allocation and distribution of public funds to journals was found to vary significantly among countries, ranging from inclusive subsidies to competitive grant funding, and for several countries like the United Kingdom, the Netherlands, Germany and Greece - no funding at all. Furthermore, the authors observed that because subscription income dwindles and might eventually cease when switching over to OA publishing models, journals must explore alternative funding avenues to sustain their operations. One such alternative approach involves publishing agreements with international commercial publishers to secure financial stability and predictability. However, a common challenge in such arrangements is that multilingualism is frequently sacrificed in favour of English, an aspect that was found particularly strong among the OADJs from small- and mid-sized publishers. This is the exact opposite direction from the one that some recent developments such as the Helsinki Initiative¹⁰ and the Coalition for Advancing Research Assessment (CoARA) working group¹¹ on Multilingualism and language biases in research assessment they advocate for. Since OADJs often operate with modest funding, it is important to have low-cost and easy to operate technological solutions that lessen the pressure on OADJs to consider signing agreements that enable an easy way of technological compliance at the potential costs of diversity or no-fee OA model.

In a study conducted by Frantsvåg & Strømme¹², the compliance of OA journals with Plan S¹³ requirements were assessed. Their analysis was based on DOAJ metadata (which, however, does not address all aspects requested by Plan S). The following technical aspects were analysed: licensing, copyright retention, DOI usage, digital archiving, machine-readable full-text format and embedded licensing information. Overall, the authors found that only 8.8% (or 1,085 of 12,350 analysed journals) met all Plan S criteria they could observe with the DOAJ data. Thus, a majority of the surveyed OA journals were not Plan S compliant at this time. The authors emphasised that especially small OADJs from the Social Sciences and Humanities (SSH) faced challenges in complying with the requirements set out by Plan S.

⁹ Laakso, M., & Multas, A.-M. (2023) European scholarly journals from small and mid-size publishers: mapping journals and public funding mechanisms, *Science and Public Policy*, 50(3), 445-456.

<https://doi.org/10.1093/scipol/scac081>

¹⁰ <https://www.helsinki-initiative.org/>

¹¹ <https://coara.eu/coalition/working-groups/>

¹² Frantsvåg & Strømme (2019).

¹³ <https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/>

Frantsvåg presented further research on this topic within a chapter of a report tied to the OADJ Study¹⁴. In addition to analysis based on DOAJ metadata, survey data from OADJs (553 journals not indexed in DOAJ and 1,087 indexed DOAJ) was analysed. Many survey respondents had answered “Unknown” and/or “No” to the more technical questions of the OADJ questionnaire. In general, the use of publishing platforms such as the Open Journal Systems¹⁵ (OJS) has a positive impact on compliance, while a lack of human and financial resources seems to be the biggest obstacles for compliance.

Overall, Bosman et al. concluded that:

“OA diamond journals are not yet fully compliant with the standards specified in the Plan S technical requirements. Of the six criteria surveyed, a mere 4.3% of OA diamond journals comply with all criteria, and only 37% comply with more than half of all criteria. Regarding the use of open licences, 37% of the journals use a CC-BY licence. Only 49% of the journals embed machine-readable licences in their metadata as required by Plan S, and around 55% use a DOI. The size of the journal correlates with their ability to attribute such identifiers to their articles. A majority (68%) of OA diamond journals have no preservation policies. Of those who do have a preservation policy, 60% use a standard archiving system that may comply with Plan S requirements. In terms of content structuration and formatting, 75% of journals are unable to format their content either in XML or HTML, providing only PDF in most cases.”¹⁶

A particularly serious and concerning finding threatening the longevity of published content from the study was the high share of journals without any preservation policy. Recently, Laakso et al.¹⁷ explored what some outcomes of such circumstances might lead to and found that at least 174 OA journals had vanished from the web between 2000 and 2019 without being comprehensively archived. The authors stated that:

“Especially small-scale and APC-free journals might have limited financial resources and, as a way to keep operating costs low, might opt for lightweight technical solutions, such as university websites and servers or content management systems (...)” [Those journals] “do not protect against technical instabilities, and if the journals cannot afford to enrol in preservation schemes, long-term access to their websites cannot be ensured (...)” “(...) 4 in 10 journals indexed in the DOAJ reporting enrollment in at least one preservation or archiving scheme”¹⁸.

¹⁴ Bosman et al., 2021.

¹⁵ <https://pkp.sfu.ca/ojs/>

¹⁶ Bosman et al., 2021, p. 8.

¹⁷ Laakso, M., Matthias, L., & Jahn, N. (2021). Open is not forever. A study of vanished open access journals. J Assoc Inf Sci Technol. 72, 1099-1112. <https://doi.org/10.1002/asi.24460>

¹⁸ Laakso et al., 2021.

The lack of compliance with technical standards is not only a problem in meeting funding criteria, but also impacts the visibility of OA journals: Khanna et al. (2022)¹⁹, e.g., used OJS installation data to assess the number of OA journals that are outside of known indexing services. They analysed 25,671 journals that are largely absent from common journal counts, as well as Web of Science and Scopus. Out of them, 84% follow the OA diamond model and only 1% are indexed in the Web of Science and 6% in Scopus.

As the results from the different studies mentioned within this section indicate there is currently room for improvement when it comes to how OADJs implement various recommended technical standards and practices; however, there is a lack of comprehensive knowledge about exactly which standards this currently concerns and with what severity. With this report, we aim to conduct a comprehensive review of all technical standards and practices identified in CRAFT-OA deliverable D3.1²⁰ and systematically pull together evidence and observations from various datasets and published literature in order to provide a comprehensive snapshot of the current situation. Through this, we will be able to provide solid direction and foundation for the development of relevant training materials and resources in CRAFT-OA T3.3.

¹⁹ Khanna, S., Ball, J., Alperin, J.P., & Willinsky, J. (2022). Recalibrating the scope of scholarly publishing: A modest step in a vast decolonization process. *Quantitative Science Studies*, 3(4): 912-930.

https://doi.org/10.1162/qss_a_00228

²⁰ Armengou et al., 2023.

3 DATA SOURCES

We used various data sources to inform our gap analysis (section 3). These sources are described in the following subsections. Moreover, we also included information from literature and other data sources that supported our gap analysis or that contributed additional information where appropriate.

3.1 DOAJ data

DOAJ is a central infrastructure serving the complete OA journal landscape. For this deliverable, we have elected to include as much data as possible stemming from DOAJ to inform our gap analysis of adoption for the various standards. As DOAJ has comprehensive metadata for all its indexed journals, it is easy to filter any analysis to focus only on the subset of non-APC journals of the index, i.e. OADJs. **Figure 1** provides a summary of the three different categories of DOAJ data that were used for supporting the analysis. In the following three subsections, we describe what each category of data is and how it was collected, and in some cases processed, in order to feed into the analysis process.

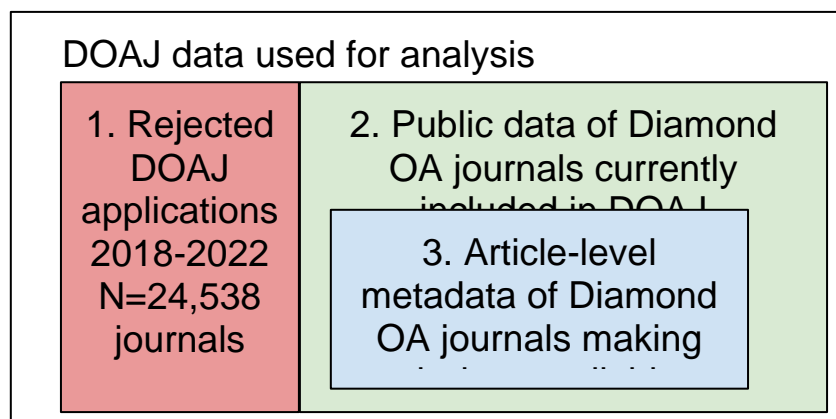


Figure 1: The three types of DOAJ data used for the analysis

3.1.1 Rejected DOAJ applications

The first category of DOAJ data concerned reviewing the application data from journals that had applied to DOAJ during the timeframe of 2018-2022, a total of 40,904 applications. By looking only at journals already included in the DOAJ one is only able to perceive a self-selected group of journals that have already overcome many central technical challenges, but the larger dataset of applications provides a way to consider a broader and more diverse set of journals. This unique dataset extracted from the DOAJ admin system is not publicly available and had not been examined systematically prior to this task. The data contains both structured form data submitted by the applicant journals as well as open text fields primarily relating to the DOAJ review and motivation for non-acceptance. Analysing this dataset, we

were interested in finding out what the most common reasons for DOAJ rejection were, and figuring out to what degree such reasons relate to failure of implementing the basic technical standards that DOAJ requires and that are also included in the scope of this deliverable. **Table 1** provides a breakdown of the received and rejected application counts per year.

Year	Received Applications	Rejected Applications	Successful Applications	% of Successful Applications
2018	5,969	3,749	2,220	37.2
2019	6,760	3,897	2,863	42.3
2020	8,416	5,082	3,334	39.6
2021	10,034	5,867	4,167	41.5
2022	9,725	5,943	3,782	38.9
TOTAL	40,904	24,538	16,366	40.0

Table 1: Received and rejected journal application counts for the DOAJ 2018-2022

Out of the 24,538 rejected applications, we were able to analyse and categorise the main reason for rejection of 13,149 applications. We were not able to analyse all of the rejected applications because DOAJ did not have a standardised notes system in the past, so the process of categorisation for this analysis had to be done manually. **Table 2** presents the categories that emerged by reviewing the 13,149 rejected journal applications, which include technical and non-technical reasons.

Reason for application rejection	Number of applications
The journal's licensing policy is not available or unclear.	3,358
The journal or publisher has been previously rejected or removed from DOAJ.	1,963
The journal's copyright policy is not available or unclear.	1,589
The ISSN is incorrect, provisional or not registered with ISSN.	1,449
The application has incorrect answers or the URLs given do not provide the required information.	1,281
The journal has not published enough research content to qualify for DOAJ inclusion.	597
The information in the journal implies that it does not employ a fair & robust peer review.	552
This application is a duplicate.	434
The journal does not employ good publishing practices.	319

The journal title in the application and/or website does not match the title at ISSN.org.	310
The URL(s) provided in the application do not work.	284
It has been less than 6 months since last application was rejected.	170
Already in DOAJ.	145
The information about the journal is in different languages.	141
The journal is not Open Access.	126
False claim of being in DOAJ.	78
The full-text articles are not available article by article with individual links.	84
Endogeny	18
Other issues	251
TOTAL	13,149

Table 2: Categories of reasons for journal application rejection to the DOAJ

3.1.2 Public journal-level data of OADJs currently included in the DOAJ

The second category of DOAJ data is straightforward: it is the publicly available journal-level data that is available for download²¹, which contains data that describes features of all journals currently included in the DOAJ. From the data, it is, for example, possible to discern which journals fulfil the basic DOAJ requirements and which ones also fulfil all the recommended requirements thus obtaining the DOAJ Seal. Among others, the DOAJ's CSV²² file of journal metadata publishes information about article processing charge (APC), which for this study was crucial, as all journals with the value of "No" in the APC column were considered Diamond Journals (13,503 journals as of 30 October 30 2023).

3.1.3 Article-level metadata of OADJs making such data available

The third category of DOAJ data relates to the journals included in the DOAJ that provide article-level metadata to DOAJ. Quantitative analysis of the article metadata of objects in the DOAJ repository is helpful in identifying gaps in relation to standards and requirements and measuring the size of the problem. The use of this metadata, which is available through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), can provide insight into where there are significant gaps in metadata coverage for this subset of journals. The presence of identifiers (Open Researcher and Contributor Identifier (ORCID), Digital Object Identifier (DOI), Research Organization Registry (ROR), Archival Resource Key (ARK) and other

²¹ <https://doaj.org/csv>

²² The acronym CSV stands for Comma-Separated Values.

Persistent Identifiers (PIDs)) and the number of other typical metadata such as title, creator, publisher, subject, etc. can be measured to see how DOAJ objects relate to standards such as the OpenAIRE Guidelines²³ or others. Of the entire available database, which consists of 9 million records, only OADJs were in our field of interest. As a guideline for how to distinguish them, we used the publicly available CSV journal-level file presented as the second category of DOAJ data. We extracted both ISSN (International Standard Serial Number) and eISSN (Electronic International Standard Serial Number) for all the journals without APC in order to connect and filter the DOAJ database of article-level records to only those belonging to OADJs.

In harvesting the DOAJ database using the OAI-PMH protocol, we focused on the `oai_doaj` format which organises information about persistent identifiers in a more structured manner than the `oai_dc` format. The Diamond Journals list of ISSNs and eISSNs was a determining factor for the analysis. Each DOAJ bibliographic record was checked to see if the identifier was present. We extracted data from 2018 up to the date of harvesting (13 November 2023). Records collected this way were analysed as a whole set and one journal after another. We obtained 1,626,481 bibliographic records from 9,890 OADJs. The provided dataset contains descriptive statistics for 9,890 Diamond Journals from the DOAJ database, focusing on various attributes such as abstract, author, DOI, eISSN, Full-text Uniform Resource Locator (URL), ISSN, keywords, language, ORCID and Additional Publisher IDs. The publication date is not included, as it was used as a filtering criterion when OAI-PMH harvesting. The detailed overview of obtained bibliographic records and each metadata are covered in the individual sections of the gap analysis.

3.2 Diamond OA Study

This survey is the largest web survey conducted on the global population of OADJs yet. The widely promoted survey contained 93 questions and was open for responses from 22 July to 11 September 2020. Many of the questions also included technical aspects of the journals that are included in the scope of this deliverable, giving a rare glimpse at what the status and opinions of a global set of OADJs were at that time. In addition to written outputs²⁴, the project also made its data publicly available, including the survey responses for the 1,619 OADJs that responded to the survey²⁵.

²³ <https://guidelines.openaire.eu/en/latest/>

²⁴ Bosman et al., 2021.

²⁵ Bosman, J., Frantsvåg, J. E., & Kramer, B. (2021a). *OA Diamond Journals Study. Dataset* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4553103>

3.3 Data from the DIAMAS project

The DIAMAS project²⁶ (Developing Institutional Open Access Publishing Models to Advance Scholarly Communication) is a 3-year EU-funded project that started in September 2022. Its goals and involved organisations are closely related to CRAFT-OA. Within DIAMAS, some data has been produced that is of direct relevance as input for this particular task: 1) a widely distributed web survey to institutional publishers in European countries and 2) a gap analysis that used survey data, website observations and focus groups to determine where some of the major challenges (including technical) currently reside among institutional publishers in European countries.

3.3.1 DIAMAS Web survey and associated landscape analysis

Between 21 March 2023 and 10 May 2023, WP2 “Mapping the European landscape of IPSP” of the DIAMAS project distributed a large survey to institutional publishers in European countries which included questions on a broad spectrum of aspects related to the activities of the institutional publishers. It included 60 questions and received responses from 685 publishers, of which 577 published academic journals. Since some of the question topics included standards, we include an analysis of responses to those questions in relevant parts of this deliverable to better understand the potential gap in technical standard implementation and use. The landscape report materials will be published by the DIAMAS project in early 2024.²⁷

3.3.2 DIAMAS Gap analysis

WP3 “Setting standards and assessing quality gaps for IPSP” of the DIAMAS project had earlier in 2023 conducted a gap analysis where they compared DIAMAS survey answers against the goals set up in the Extensible Quality Standard for Institutional Publishing (EQSIP) 1.0. To complement this, a manual web scraping/observing task has been carried out where all Institutional Publishing Service Providers (IPSP) web pages have been visited to record factors that are present in the EQSIP but were not part of the survey questions. Additionally, DIAMAS WP3 has conducted regional focus groups/group interviews with a handful of publishers at a time. The results of this gap analysis, focusing on the IPSP level of policies and features rather than individual publications, are available in a public report²⁸, and relevant content from that publication is reflected upon and built upon in this deliverable.

²⁶ <https://diamasproject.eu/>

²⁷ Due to the parallel involvement of several CRAFT-OA partners in the DIAMAS project, the results of the DIAMAS survey were already available for the gap analysis in CRAFT-OA D3.2.

²⁸ Brun, V., Torny, D., & Pontille, D. (2023). *DIAMAS Deliverable 3.3 Report on the gap analysis results (1.0)* (under EC review). Zenodo. <https://doi.org/10.5281/zenodo.10083615>

3.4 Technical support for standards among open source publishing content management systems

In November 2023 we reached out to experts on the technical support among publishing software for the technical standards identified in Armengou et al.²⁹ with regards to identifiers, metadata, content and website features. The three experts were each asked to briefly describe if and how support for the technical standard is available through the publishing platform of their expertise. In this way, we were able to gain technical insights into OJS, Janeway and Lodel (from the perspective of OpenEdition (OE) which is the primary user of Lodel). The results are presented along with the gaps and challenges of each technical standard. The software features introduced in these sections represent the state of development as of November 2023.

3.5 Data for future work

The data sources described in sections 2.1-2.4 were the ones we deemed most appropriate for reaching the goals set out for this deliverable, however, we could mention a few important additional ones that we ended up not using for this purpose but could be useful for future related work. Khanna et al.³⁰ provide a large dataset of describing the PKP's (Public Knowledge Project's) dataset of known journals running OJS. While this list is likely the most comprehensive in what it sets out to do, it only contains limited metadata concerning technical features used by the journals, and as a result, it was not useful for the particular type of detail we wanted to include in this deliverable without substantial additional data collection.

Another data source is Crossref, which in particular through their Labs interface³¹ provides interesting analytics concerning the metadata deposited by publishers. However, for our specific purpose of focusing on OADJs and not all scholarly journals, the interface provides no filtering options that would enable isolation to only OADJs in some way.

A final service we would like to mention in this context is the OpenAIRE Research Graph³². While it contains metadata for a lot of journals and articles within them, it lacks a filtering option to only OADJs similar to Crossref, so obtaining any specific information about the specific subset of journals we are interested in for this deliverable is not possible.

²⁹ Armengou et al., 2023.

³⁰ Khanna, S. Raoni, J., Smecherr, Al, Alperin, J., Ball, J., & Willinsky, J. (2022a). *Details of publications using software by the Public Knowledge Project* (Harvard Dataverse; Version 3) [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/OCZNVY>

³¹ <https://prep.labs.crossref.org/>

³² <https://explore.openaire.eu/>

4 GAP ANALYSIS

This section is divided into four subsections based on the categorisation of standards as provided in D3.1: identifiers, metadata, content, and website features. Within each of the four sections, we provide an individual gap analysis for each standard. However, before zooming in on individual standards, there are some key contextual observations about the software environments that OADJs are running on today, which help frame the interpretations of the gap analysis for the individual standards. In the DIAMAS survey carried out in 2023, there was a question that queried what publishing system the institutional publishers were using. The results of how the 577 IPSPs involved in journal publishing responded to this question are presented in **Table 3**. Please note that some IPSPs responded by saying that they use multiple systems, so the total number and total % exceeds the total number of responses/100%. While most (62%) were using OJS, the rest of the software list is quite long, testifying to the heterogeneity of the technical landscape.

Open Journal Systems (OJS)	355	61.5%
Customisation or own development	67	11.6%
WordPress	56	9.7%
Open Monograph Press (OMP)	43	7.5%
Don't know	43	7.5%
Other commercial software	42	7.3%
Other open source software	35	6.1%
Lodel	34	5.9%
DSpace	31	5.4%
Drupal	20	3.5%
No answer	18	3.1%
ScholarOne	14	2.4%
Editorial manager	11	1.9%
Janeway	9	1.6%
Manifold	2	0.3%
Pressbooks	1	0.2%

Table 3: DIAMAS survey responses for the question what software the institutional publisher is using

Another relevant question in the DIAMAS survey queried how the institutional publishers have organised the maintenance and update of their technical infrastructure. The results for how the 577 IPSPs involved in journal publishing responded to this question are presented in

Table 4. With over half of responses relating to some in-house variant of organisation, it provides important information for understanding in what type of contexts many OADJs are operating.

In-house by an IT department personnel	238	41.2%
Partially outsourced	97	16.8%
In-house by a dedicated publishing department	91	15.8%
Mainly outsourced	69	12%
In-house across different departments	66	11.4%
Fully outsourced	61	10.6%
Don't know	24	4.2%
No answer	18	3.1%
Other	6	1%
No provision	5	0.9%

Table 4: DIAMAS survey responses for questions concerning how the institutional publishers services and / or technical infrastructure are maintained and updated?

These are key observations about the software environments of OADJs. In the light of these observations, the following four sections will zoom in on the gap analysis of the individual standards across the four main standards categories.

4.1 Identifiers

Following the FAIR³³ principles, the use of identifiers promotes findability and accessibility of resources from both a human and a computational perspective as shown in Armengou et al.³⁴. However, many journals are challenged by complying with the wide range of standards concerning identifiers for their content. This may be exemplified by Martínez-Galindo et al.³⁵ who observed, referring to Spanish OA journals, that “DOAJ shows 680 Spanish journals as being published OA, though half do not offer any identifier (DOI, Handle, etc.)....”³⁶. The following section addresses each of the persistent identifiers (PIDs) presented in Armengou et al.³⁷ with a description of the standard, where it is required, and presents challenges and

³³ The acronym FAIR stands for Findable, Accessible, Interoperable, and Reusable.

³⁴ Armengou et al., 2023.

³⁵ Martínez-Galindo, F.J., Rubio, F., Hernández-San-Miguel, J. & Fernández Burguete, S. (2019). Plan S: Challenges and Opportunities in Spain. *Insights*, 32(1). <https://doi.org/10.1629/uksg.467>

³⁶ Martínez-Galindo et al., 2019, p. 5.

³⁷ Armengou, C., Redhead, C., & Rooryck, J. (2023a). *DIAMAS Deliverable 3.5 Extensible Quality Standard in Institutional Publishing (EQSIP)_V1.0* (under EC review). Zenodo. p. 25-26. <https://doi.org/10.5281/zenodo.7923916>

gaps regarding the standard or requirement along with an overview of technical support among selected publishing software (OJS, Janeway, Lodel) and preliminary conclusions. Journals and platforms should also be aware of the potential risks involved when selecting specific persistent identifiers for implementation. The organisation that issues and maintains the technical infrastructure needs to be reliable since persistent identifiers often rely on their operations to direct users to the correct content or record in their database. A recent report contains case studies of PIDs failing to be maintained by the issuing organisation, and situations where PID implementations have been observed to be unreliable in their operation without proper plans for contingency³⁸.

4.1.1 ISSN/EISSN

Description of the standard “ISSN/eISSN”

The International Standard Serial Number (ISSN) is an international identifier for continuously publishing resources such as journals, newspapers or magazines. It is available for both print and online publications and follows the structure of “ISSN-[four digits] [four digits]”. While the digits carry no internal meaning, the ISSN number allows for a clear identification of publications even if their titles or names are similar or the same. ISSNs/eISSNs are jointly managed by the ISSN International Centre based in Paris and has national centres in 93 countries. The different national centres assign ISSN identifiers for continuing publications if they are first published in the respective countries. The ISSN remains the same throughout the publication of e.g. a journal, unless the title is changed, it is merged with another publication, or split into two or more new journals.

For journals, this means that they will often only apply for an ISSN once, since it is assigned to a specific unit of recurring publication, and not, for example, on the article level. However, journals and publishers have to adhere to specific administrative procedures, identify the appropriate ISSN centre and supply the necessary supporting documents as part of this process.

Using the ISSN identifier facilitates the clear identification of the respective journal even if there are several journals with a similar name currently in existence. Additionally, all ISSNs may be found through the ISSN Portal and other large databases. For instance, in Germany, this means the catalogue of the national journal library, the national bibliography and the journal database³⁹.

³⁸ de Castro, P., Herb, U. Rothfritz, L., & Schöpfel, J. (2023). *Failed PIDs and unreliable PID implementations*. Knowledge Exchange. [\(hal-04245144\)](#)

³⁹ The journal database in Germany is the Zeitschriftendatenbank (ZDB) <https://zeitschriftendatenbank.de/startseite>

Where is an “ISSN/eISSN” required?

ISSN/eISSN identifiers are a prerequisite by DOAJ in order to be eligible for inclusion in the directory. However, ISSNs are also essential for being eligible for indexing in major indexing services such as Scopus or Web of Science.

Challenges and gaps in relation to “ISSN/eISSN”

The OADJ Study from 2021⁴⁰ calculated that between 17,000 and 29,000 OADJs operate globally using an ISSN. It is hard to get an exact estimate for what share of OADJs are actively publishing without an ISSN, but we have listed the evidence we have been able to gather in this section.

The 2023 DIAMAS survey distributed to institutional publishers, primarily based in Europe, found that of the 577 responding publishers that published academic journals, only 332 (58%) did so using an ISSN identifier.

Investigating the DOAJ application data, 1,449 of the 13,449 applications analysed were submitted (10.7%) where the ISSN was incorrect, provisional or not registered with issn.org.

When considering the dataset of known public installations of OJS, Open Monograph Press⁴¹ (OMP) and Open Preprint Systems (OPS) by Khanna et al.⁴², almost half of the journals using OJS do not have an ISSN entered in the data that is pinged back to the Public Knowledge Project. It is worth noting that this observation comes with some caveats, as many of these installations could have been created for test purposes or for purposes other than academic publishing, been abandoned a long time ago or it may even be that someone has simply failed to enter the ISSN data into the correct field in the control panel of the installation.

Technical support for “ISSN/eISSNs” among software

Identifier	OJS	Janeway	Lodel
ISSN/eISSN identifier	Supported in OJS core	Supported in Janeway core	Both are supported

Technical support within publishing systems for ISSN/eISSN is given in OJS, Janeway and Lodel.

Conclusions “ISSN/eISSN”

Based on this overview, it would seem that even though ISSN identifiers are not technically complex, their presence and use among journals is not something that automatically happens, and particularly, journals starting out may not have applied for such an identifier yet. Fulfilling

⁴⁰ Bosman et al., 2021.

⁴¹ <https://pkp.sfu.ca/software/omp/>

⁴² Khanna et al., 2022a.

the requirement of having an ISSN requires a one-time application process rather than ongoing maintenance or extra work, which makes it very achievable for all serious journals in terms of the effort needed. In terms of additional support, awareness-raising about the requirements of the application process could be something to consider.

4.1.2 ORCID identifier

Description of the standard “ORCID identifier”

ORCID, which stands for Open Researcher and Contributor ID, is an international standard for persistent unique identifiers for researchers and scholars. It provides a persistent digital identifier that distinguishes individual researchers and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between researchers and their professional activities, ensuring that their work is recognized. The ORCID system is operated by a non-profit organisation ORCID INC, which is registered in the United States.

These identifiers help eliminate name ambiguity by ensuring that researchers are correctly attributed for their work, regardless of any changes in their name, affiliation or research focus. ORCIDs are widely used in academic and research communities, allowing for better tracking of research outputs, collaborations and impact. Researchers, institutions, funding agencies and publishers can integrate ORCID identifiers into their systems to streamline workflows and enhance the accuracy of research databases. Overall, ORCID plays a crucial role in improving the transparency and efficiency of the research ecosystem.

Integrating ORCID identifiers into journal workflows enhances efficiency, reduces errors and ensures proper recognition of authors' contributions. It also contributes to the overall improvement of the scholarly publishing ecosystem by promoting transparency, data integrity and collaboration among researchers, journals and institutions.

Where is the “ORCID identifier” required?

ORCID is mentioned in the EQSIP 1.0⁴³, and having author PIDs is also a strongly recommended criterion for publishing venues by cOAlition S/Plan S, which mentions ORCID as an example (at the moment, it is not a mandatory feature for compliance).

Challenges and gaps in relation with the “ORCID identifier”

In the OADJ Study⁴⁴ from 2021, 525 out of 1,619 responding journals (32%) declared their use of ORCID identifiers.

⁴³ Armengou et al., 2023a.

⁴⁴ Bosman et al., 2021.

The DIAMAS survey from 2023 was not really tailored to inquire into ORCID use specifically, as this was not included as a multiple-choice option for PIDs used by institutional publishers. 4 out of the 577 respondents that published journals declared the usage of ORCID identifiers when asked via a free-text question about which PIDs they are using outside of the pre-defined options available.

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, of the 13,503 OADJs, 3,404 (25%) have reported “Yes” concerning “Article metadata includes ORCIDs”, 2,997 report “No” (22%) and 7,102 (53%) lack information about this aspect in the DOAJ journal-level metadata file. The respective numbers for the 6,562 APC-based journals in DOAJ were 3,004 (46%) “Yes”, 1,242 (18.9%) “No”, 2,316 (35.3%) with lack of information. This information was based on DOAJ data from the 31st of October 2023.

Regarding the article-level metadata for the whole DOAJ database for Diamond Journals, the largest gap concerns the ORCID identifier, where only 9% of the articles allow confirmation of authorship beyond doubt using an authority database. For the data grouped by particular journals, the average presence of ORCIDs is low, at 10.44%, with a high standard deviation of 23.99%, which indicates that ORCID inclusion is not yet a common practice across Diamond Journals.

In 2022, The Scholastica “State of Journal Production and Access”⁴⁵ survey included 82 respondents from academic organisations that publish independently without external publishers. More than half of the respondents stated that they were using ORCIDs in their metadata.

Technical support for “ORCID identifier” among software

Identifier	OJS	Janeway	Lodel
ORCID identifier	Supported with a plug-in, maintained by PKP, planned to be moved into OJS core https://github.com/pkp/orcidProfile	Supported in Janeway core	Not supported

The ORCID identifier is supported by OJS with a plug-in that is maintained by PKP and will prospectively be moved into the core functionalities. It is also possible to add ORCID identifiers manually to the author metadata in OJS although this increases the likelihood of

⁴⁵ Scholastica (2022). *The State of Journal Production and Access 2022: Report on survey of independent academic publishers. Scholastica Report.* <https://lp.scholasticahq.com/journal-production-access-survey-2022/>

false or unauthenticated IDs and a higher workload. Janeway includes support for the ORCID identifier as a core functionality. Lodel does not support the ORCID identifier.

Conclusions “ORCID identifier”

Despite the advantages of ORCID identifiers in unambiguously linking researchers and their work independently from names or affiliations, the gap analysis shows a relatively low uptake, also in comparison to APC-based journals in the DOAJ. Only in the Scholastica survey with a much smaller scope, more than half of the respondents stated that they were using ORCID identifiers. However, the full status remains somewhat unclear, considering that the DIAMAS survey did not address this issue directly and 53% in the DOAJ data about OADJs do not offer information on ORCID usage. From a software point of view, OJS allows for the integration of ORCID identifiers with a plug-in, and Janeway has ORCID IDs as a core functionality. Seeing that journals could include the ORCID identifier in their workflow when collecting author information in a one-step process, more information and awareness for the identifier could lead to a higher proportion of usage.

4.1.3 ROR identifier

Description of the standard “ROR identifier”

The Research Organization Registry (ROR)⁴⁶ identifier makes it possible to uniquely identify and connect research organisations to researchers and research outputs, which helps both human readers but in particular for machine-reading purposes. If organisation names are only provided as free-text in various metadata, there is a substantial challenge in disambiguating research organisation names when aggregating data, as the names can be outdated, misspelled, wrong or in a different language. This has implications for how easy and reliable it is to create systems and search tools based on such data. ROR data provides similar benefits to organisations that ORCIDs provide for individuals.

ROR registration and curation is a bit different in comparison to how other identifiers are created and maintained. It is a community-driven effort based on public records where anyone can suggest additions and changes to any of the organisational identifiers.

Journals need to have a metadata field for requesting authors to fill in any applicable RORs in conjunction with all other manuscript information at time of submission, and also make this data publicly available once the content is published. In addition to journals checking that the submitted ROR is linked to the correct and desired location indicated by the authors, the journal can also as extra steps: 1) suggest entering a suitable ROR if the authors have not done it themselves, 2) add an official organisation name in a different language to an existing ROR

⁴⁶ <https://ror.org/>

entry, and 3) suggest that a new ROR be created and added for organisations that are not yet part of the registry.

This way, individual journals contribute to a more cohesive, interlinked and transparent data environment for organisational data. The journals also benefit themselves e.g. submission systems can start suggesting to authors suitable affiliation records from the large public ROR dataset, which would remove the necessity for editors to check affiliation metadata. Making use of the ROR identifier also potentially increases the visibility of journals among research funding organisations who are tracking the publication outlets of researchers.

Where is the “ROR identifier” required?

Currently, the implementation of ROR among publication metadata is mentioned as a basic requirement in EQSIP V1.0⁴⁷.

Challenges and gaps in relation with the “ROR identifier”

In the OADJ Study from 2021⁴⁸, there was not a dedicated question regarding ROR use among journals. However, for the question of “Which persistent identifiers does the journal attribute or use attached to articles, authors, research grants?” In addition to predefined responses, there was also the option to select “Other”, but only 135 out of the 1,619 responding journals (8%) did so, indicating quite a low uptake since that response also includes identifiers other than only ROR.

The DIAMAS survey from 2023 was similarly not a focused inquiry into ROR use specifically, as it was not included as a multiple-choice option for PIDs used by the institutional publishers. One out of the 577 respondents that published journals declared usage of ROR identifier when asked through a free-text question about which PIDs they are using outside of the pre-defined options available.

Technical support for “ROR identifier” among software

Identifier	OJS	Janeway	Lodel
ROR identifier	Supported with a plug-in, maintained by PKP development partner https://github.com/withanage/ror	Planned development (https://github.com/BirkbeckCTP/janeway/issues/3168)	Not supported

⁴⁷ Armengou et al., 2023a.

⁴⁸ Bosman et al., 2021.

The ROR identifier is supported in OJS with a plug-in which is maintained by a PKP development partner. Adding the ROR identifier to Janeway is a planned development, and the ROR identifier is not supported in Lodel.

Conclusions “ROR identifier”

Overall, ROR identifier use still seems relatively low among OADJs, but ROR has been on a steady growth path, and with the recent work to integrate the data with that of the Crossref Open Funder Registry⁴⁹ to also include funding organisations in ROR, there seems to be substantial momentum behind this identifier. Therefore, it is likely that there will be more organisations putting pressure on its adoption. For journals and portals, the basic implementation of allowing for ROR identifier metadata at submission is not technically very challenging and OJS even has a plug-in for this purpose. Some additional manual work for each manuscript is introduced by publishers verifying the accuracy of submitted ROR identifiers and potentially contacting authors for corrections. Similarly to ORCIDs, one could see the benefit of training materials for this identifier in raising awareness and uptake of the standard among OADJs.

4.1.4 ARK identifier

Description of the standard “ARK identifier”

Archival Resource Key (ARK) identifiers are URLs that support long-term access to information⁵⁰ and may therefore act as a persistent identifier. They follow a standardised syntax and may be queried by appending different strings, such as ‘?’, to the URL to access the previously included data, such as article metadata.

While it was initiated by the California Digital Library and the DuraSpace/LYRASIS, it is currently supported by numerous institutions as well as the U.S. American NDSA (National Digital Stewardship Alliance). Assigning or using ARKs is free of charge and may be self-hosted. Assignment is handled by organisations holding a NAAN (Name Assigning Authority Number). Currently, over 1,000 organisations mainly from the GLAM (galleries, libraries, archives and museums) sector are registered as NAANs. ARKs serve as identifiers for a number of different categories such as digitised documents, archival objects, fine art but also for historical persons or organisations. However, in the context of the current analysis, it should be mentioned that ARK identifiers may be used on different levels in the field of scholarly publishing, to identify a journal, an article or an issue.

⁴⁹ ROR (2023, October 12). How ROR and the Open Funder Registry Overlap: A Closer Look at the Data. *ROR Blog*. <https://ror.org/blog/2023-10-12-ror-funder-registry-overlap/>

⁵⁰ <https://arks.org/about/>

Where is the “ARK identifier” required?

Based on the survey by Armengou et al.⁵¹, ARK identifiers are recommended for the DOAJ Seal⁵² as part of the requirement of using persistent identifiers for content. The exact wording on the DOAJ Seal criteria webpage is the following as of November 2023 “Articles must use persistent article identifiers. DOI, ARK or Handle are the most commonly used”⁵³.

Challenges and gaps in relation with the “ARK identifier”

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, of the 13,503 OADJs, 37 (0.2%) journals indicated the use of ARK identifiers as part of their persistent article identifiers. This can be contrasted to 9,172 journals that indicated use of DOIs (67.9%) and 117 journals that used Handles (0.8%). 4,221 journals (31.3%) had no information about persistent identifiers in their DOAJ journal-level metadata.

In the 2021 OADJ Study⁵⁴, there was no dedicated question regarding ARK use among journals. However, for the question of “Which persistent identifiers does the journal attribute or use attached to articles, authors, research grants?” there was, in addition to predefined ones, also the option to select “Other DOIs”, with 400 out of 1,619 responding journals (25%) doing so. Note that these responses also include indiscernible DOI identifiers other than only ARK.

The DIAMAS survey distributed in 2023 to institutional publishers, mainly based in Europe, found that of the 577 responding publishers that published academic journals, only 2 (0.3%) indicated use of ARK identifiers.

Technical support for “ARK identifiers” among software

Identifier	OJS	Janeway	Lodel
ARK identifier	Supported with a plug-in, https://github.com/yasielpv/pkp-ark-pubid	Not supported	Not supported

The ARK identifier has been available in OJS via a plug-in since version 3.1. It is not supported in Janeway and Lodel.

⁵¹ Armengou et al., 2023.

⁵² <https://doaj.org/apply/seal/#criteria>

⁵³ <https://doaj.org/apply/seal/#criteria>

⁵⁴ Bosman et al., 2021.

Conclusions “ARK” identifier

The persistent identifier ARK appears to be very sparsely used among OADJs, although recommended as an alternative next to DOI and Handle for the DOAJ Seal and already supported with a plug-in in OJS. Depending on geographic location, for most OADJs, DOIs might be more well-known and accessible to obtain, but ARKs fulfil many of the same functions so it is a viable alternative to consider depending on the individual circumstances of the journal.

4.1.5 Handle identifier

Description of the standard “Handle identifier”

The Handle identifier⁵⁵ is a PID first developed at the Corporation for National Research Initiatives (CNRI) and currently administered at the DONA Foundation based in Geneva. “Multi Primary Administrators” (MPA) are in charge of assigning PIDs under their respective naming authority. In this way, the Handle identifier forms the basis for other established systems like DOI assigned by the DOI Foundation. In the syntax of an identifier within the Handle system, the MPA is encoded in the first digits, the following digits encode the identifier for a digital object. For example, in the case of the DOI Foundation this prefix is “10”, the German Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) assigns identifiers with the prefix “21” and the CNRI uses “20” as the prefix. The CNRI allots identifiers with the prefix “20.500” with four to five digits following this which may be resolved by adding the respective identifier to “<https://hdl.handle.net/>”.

If journals use one of the Handle-based identifiers, they have to establish a workflow with the respective naming-authority, an example of which is described below for the DOI identifier. It is of importance to maintain the identifiers to ensure their persistent resolving.

Where is the “Handle identifier” required?

Based on the survey by Armengou et al.⁵⁶ handle identifiers are recommended for the DOAJ Seal as an option for fulfilling the requirement of using persistent identifiers for content. The exact wording on the DOAJ Seal criteria webpage is the following as of November 2023 “Articles must use persistent article identifiers. DOI, ARK or Handle are the most commonly used”⁵⁷. Plan S also mentions the Handle identifier as an example for a PID under the mandatory conditions for all publication venues⁵⁸.

⁵⁵ <https://www.dona.net/handle-system>

⁵⁶ Armengou et al., 2023.

⁵⁷ <https://doaj.org/apply/seal/#criteria>

⁵⁸ https://www.coalition-s.org/technical-guidance_and_requirements/

Challenges and gaps in relation with the “Handle identifier”

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, of the 13,503 OADJs, 117 (0.8%) journals indicated the use of Handle identifiers as part of their persistent article identifiers. This can be contrasted to 9 172 journals that indicated use of DOIs (67.9%) and 37 journals that used ARK identifiers (0.2%). 4,221 journals (31.3%) had no information about persistent identifiers in their DOAJ journal-level metadata.

The 2021 OADJ Study⁵⁹ did not have a dedicated question regarding Handle use among journals. However, for the question of “Which persistent identifiers does the journal attribute or use attached to articles, authors, research grants?” there was, in addition to predefined ones, also the option to select “Other DOIs”, with 400 out of 1,619 responding journals (25%) doing so. Note that these responses also include indiscernible DOI identifiers other than only Handle.

The DIAMAS survey that was distributed in 2023 to institutional publishers who were largely from Europe, found that of the 577 responding publishers that published academic journals, 33 (5.7%) indicated the use of Handle identifiers.

Identifier	OJS	Janeway	Lodel
Handle identifier	Not supported	Not supported	Handle.net is the by default PID system for all documents

Technical support for “Handle identifiers” among software

The Handle identifier is not supported in OJS nor in Janeway. Lodel uses Handle.net as the default PID for all documents.

Conclusions “Handle identifier”

Although Lodel supports the Handle.net identifier of the handle identifier system as their default PID, its uptake overall among OADJs appears to be low. Similar to the situation concerning ARKs, the conclusions regarding Handles are: DOIs are likely the more well-known and accessible persistent identifier for journals looking into options for persistent identifiers, but other identifiers like Handle might make more sense in some individual cases depending on geographical location and organisational context of the publisher.

⁵⁹ Bosman et al., 2021.

4.1.6 DOI identifier

Description of the standard “DOI identifier”

The DOI system establishes a framework for the lasting and distinct identification of various types of objects in the digital realm. DOI, standing for "Digital Object Identifier", emphasises the digital identification of an object rather than an identification limited to digital entities. In this international standard, the term "digital object identifier" specifically pertains to the system outlined herein.

Functioning seamlessly on the Internet, the DOI system assigns a permanent DOI name to an object, ensuring a stable and resolvable link to the latest information about that object. This link directs users to the object or pertinent information about it on the internet. Despite potential changes in the object's information, the DOI name remains constant. Within the DOI system, a DOI name can be resolved to various types of data related to the identified object, including URLs, email addresses, other identifiers and descriptive metadata.

The DOI system not only supports the creation of automated services and transactions but also serves diverse applications. These applications encompass information and documentation management, metadata organisation, streamlining electronic transactions, ensuring unique identification for any data form and facilitating both commercial and non-commercial dealings.

Each object linked to a DOI name is meticulously described through DOI metadata. This description relies on a structured and adaptable data model that allows precise and detailed association of metadata with the object. This flexibility ensures support for various descriptions and services, promoting interoperability between different DOI applications.

The DOI system's scope is not determined by the format or nature of the content it refers to but instead by the functions it offers and the context of its usage. Within networks of DOI applications, the DOI system guarantees unique identification, persistence, resolution, metadata management and semantic interoperability⁶⁰.

For journals, the DOI system helps in the identification and citation of articles. It provides a permanent link to the digital location of the article, making it easier for researchers and readers to access and cite the work. The DOI remains constant even if the URL of the article changes, as long as the publisher updates the new location to which the DOI should resolve to at the DOI registrar, facilitating persistent access.

⁶⁰ <https://www.iso.org/obp/ui/#iso:std:iso:26324:ed-1:v1:en>

The process of adding DOIs involves initial setup, configuration of DOI plug-ins, automatic minting of DOIs for new articles and ensuring that metadata is accurately transferred to the relevant registration institution (e.g. Crossref). This process does not have to be repeated once set up for every new piece of content, but new DOIs need to be minted for new articles.

The DOI system is beneficial for journals as it ensures persistent linking, enhances the findability and accessibility of articles, promotes interoperability, aligns with best practices in research data management and provides a standardised format for identification⁶¹.

Where is the “DOI identifier” required?

Use of persistent identifiers (PIDs) is a mandatory technical condition in Plan S and DOI numbers are the preferred solution. DOI is also one of the suggested PID options for being eligible for the DOAJ Seal, where the use of PIDs is also mandatory.

Challenges and gaps in relation with the “DOI identifier”

While Gorraiz et al.⁶² do not exclusively focus on OADJs, they did explore the coverage of DOIs in the Web of Science Core Collection and Scopus from 2004 to 2015. The authors found huge disciplinary differences in the uptake of DOIs. While they generally observed an increasing DOI usage over time, they observed that DOI usage in the Arts and Humanities remained comparatively low: 90% of the citable documents in the Web of Science Core Collection originating from the domain of the Sciences and Social Sciences had a DOI in 2014, but less than 50% of the documents from the Arts and Humanities used a DOI.

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, of the 13,503 OADJs, 9,172 (67.9%) journals indicated the use of DOIs as part of their persistent article identifiers. This can be contrasted with 117 journals that indicated use of Handle identifiers (0.8%) and 37 journals that used ARK identifiers (0.2%). 4,221 journals (31.3%) had no information about persistent identifiers in their DOAJ journal-level metadata.

Regarding the article-level metadata for the entire DOAJ database for OADJs, DOI usage remains a challenge, since over 20% of articles do not have such identifiers. Grouping the data by journals, 79.1% of journals have at least one DOI in the metadata. The standard deviation of 34.2% and the 25th percentile at 75% show a significant number of journals without DOIs.

⁶¹ Liu, J. (2021). Digital Object Identifier (DOI) Under the Context of Research Data Librarianship. *Journal of eScience Librarianship*, 10(2): 4. <https://doi.org/10.7191/jeslib.2021.1180>

⁶² Gorraiz, J., Melero-Fuentes, D., Gumpenberger, C., & Valderrama-Zurián, J. C. (2016) Availability of digital object identifiers (DOIs) in Web of Science and Scopus. *Journals of Informetrics*, 10(1), 98-109. <https://doi.org/10.1016/j.joi.2015.11.008>

The 2021 OADJ Study⁶³ contained a question formulated as “Which persistent identifiers does the journal attribute or use attached to articles, authors, research grants? (Tick all that apply)”. This question was optional for the 1,619 responding journals to provide answers to. Three of the predefined response options concerned DOIs, including Crossref DOIs, Datacite DOIs and Other DOIs. In their responses, 961 journals (59%) ticked Crossref DOIs, 125 (8%) Datacite DOIs, 401 (25%) Other DOIs and 393 (24%) did not select any of the three DOIs. Please note that it was possible for journals to select multiple DOI varieties, this is why the total number and percentage exceeds the total number of absolute responses.

The 2023 DIAMAS survey found that of the 577 responding publishers that published academic journals, 353 (61.2%) used Crossref DOIs, 61 (10.6)% DataCite DOIs and 44 (7.6%) other DOIs.

The 2022 Scholastica The State of Journal Production and Access⁶⁴ survey included 82 respondents from academic organisations that publish independently without external publishers. More than half of the respondents stated use of DOIs in their metadata.

The European Open Science Cloud (EOSC) Interoperability Framework⁶⁵ reports as a challenge for the community that "Multiple service providers for different types of PIDs exist"⁶⁶. Moreover, it identifies "a need to have a common and well-understood PID policy across communities."⁶⁷

In a recent blog post⁶⁸, the Confederation of Open Access Repositories (COAR) points out some concerns for the adoption of PIDs, especially DOIs. What COAR mainly raises is a concern about the DOI-centric research information environment that is being built up, which risks leaving out scholars and organisations from countries where the cost of obtaining DOIs and being a member of DOI registration services is prohibitive. COAR further brings forward the issue of developing a lock-in situation where the services issuing DOIs obtain increasing leverage both when it comes to pricing and decision power. COAR thus advocates for a more diverse, but interconnected, PID environment where other alternatives besides DOI can also

⁶³ Bosman et al., 2021.

⁶⁴ Scholastica, 2022.

⁶⁵ Corcho, O., Eriksson, M., Kurowski, K. Ojsteršek, M., Choirat, C., Sanden, M. van de, & Coppens, F. (2021). *EOSC interoperability framework – Report from the EOSC Executive Board Working Groups FAIR and Architecture*. <https://data.europa.eu/doi/10.2777/620649> as well as <https://eosc-portal.eu/eosc-interoperability-framework>

⁶⁶ Corcho et al., 2021, p. 14.

⁶⁷ Corcho et al., 2021, p. 15.

⁶⁸ COAR (2023, September 28). Persistent Identifiers: Addressing the challenges of global adoption. *COAR Confederation of Open Access Repositories*. <https://www.coar-repositories.org/news-updates/persistent-identifiers-addressing-the-challenges-of-global-adoption/>

contribute and be equally visible, interconnected and measurable in the digital research landscape.

Identifier	OJS	Janeway	Lodel
DOI identifier	Supported in OJS core	Supported in Janeway core	DOIs are supported for most journals contents but not for all types (e.g., book reviews). However, all the contents receive another type of PID, with handle.net.

Technical support for “DOI identifier” among software

In the OJS and Janeway core, DOI identifiers are supported. Lodel also supports DOIs, with OE assigning them to most but not all publication types.

Conclusions “DOI identifier”

The uptake of DOIs is rather high compared to many other identifiers, with results suggesting that around $\frac{2}{3}$ of all OADJs provide DOIs for their published content. This holds true in comparison to other PIDs used for articles, but also in comparison with other PIDs in general. Nevertheless, DOI uptake is not at 100%, and the other identifiers (ARK and Handle) do not fill the remaining gap with their uptake either. Hence, this implies that many articles are still published without a persistent identifier, which has a negative impact on findability and visibility. Various efforts should be undertaken to raise awareness and offer practical low-cost solutions that would be targeted at specifically reaching and engaging with the $\frac{1}{3}$ of known OADJs that are operating without PIDs for their content.

4.1.7 Funder DOIs/PIDs

Description of the standard “Funder DOIs/PIDs”

Funder PIDs are identifiers for uniquely identifying research funders, facilitating monitoring of research produced by different funders and their funding instruments. This enables highlighting connections of research activities, the research focus areas of grant-giving organisations, or tracking of funding results for easier compliance checking by the research funder. For example, journals registering their DOIs with Crossref can deposit funder DOIs when transferring the article metadata to Crossref. Funders, and anyone else interested, can

then query the Crossref Application Programming Interface (API) to discover which articles and other content has originated from research funded by a particular research funder. Recently, Crossref announced⁶⁹ that they will use RORs for funder identification (see the earlier section in this deliverable for a gap analysis on RORs) instead of the Open Funder Registry (formerly FundRef) they have been using so far. Against the background of this development, it is important to point out that the ROR identifying an author's institutional affiliation is not necessarily the same ROR identifying who funded the research of the same author. Developers of publishing software will need to make the necessary adjustments in their plug-ins and associated metadata export, so that the ROR for an author's institution and the ROR for the funder (in those cases where these are not the same organisation) are associated with the correct metadata entries.

PIDs for funding information encourages journals and platforms to collect standardised information on grants that enabled the research presented. Such PIDs need to be collected for every piece of content (if applicable) and each piece of content can have more than one funder PID. Journals need to provide a respective metadata field to cover this information. Moreover, they need to enable this information to be transferred to aggregators such as Crossref. Either authors can provide this information during submission or, less optimally, publishers can attempt to identify and extract relevant information from the funding and or acknowledgment section of a paper and enter it. In general, linked information is beneficial for visibility. For authors, such data can help with reporting; for funders, it strengthens traceability of funded research outputs.

Where are “Funder DOIs/PIDs” required?

Plan S states the following among the mandatory requirements for all publication venues “Metadata must include complete and reliable information on funding provided by cOAlition S funders (including, as a minimum, the name of the funder and the grant number/identifier)”⁷⁰ and among the strongly recommended additional criteria for all publication venues “Support for PIDs for authors (e.g., ORCID), funders, funding programmes and grants, institutions and other relevant entities”⁷¹.

Challenges and gaps in relation with “Funder DOIs/PIDs”

In the OADJ Study⁷² from 2021, there was a question formulated as “Which persistent identifiers does the journal attribute or use attached to articles, authors, research grants?” that was optional for the 1,619 responding journals to provide answers for. One predefined

⁶⁹ French, A., Hendricks, G., Lammey, R., Michaud, F., & Gould, M. (2023, September 07). Open Funder Registry to transition into Research Organization Registry (ROR). *Crossref Blog*. <https://www.crossref.org/blog/open-funder-registry-to-transition-into-research-organisation-registry-ror/>

⁷⁰ <https://www.coalition-s.org/technical-guidance-and-requirements>

⁷¹ <https://www.coalition-s.org/technical-guidance-and-requirements/>

⁷² Bosman et al., 2021.

response was Grant ID (i.e. the project number/grant number given by the research funder to be attached to funded outputs), which 153 (9%) of respondents chose. The OADJ Study, indicates that the hosting situation is connected to PID usage⁷³: International platforms show the highest adoption rate (17%) in Grant IDs, followed by commercial platforms with 11%. According to the OADJ Study, these are lower shares than the average on commercial hosting platforms.⁷⁴ Bosman et al. suggest that especially international and national platforms will be of elevated importance for future adoption of identifier standards in OA journals.⁷⁵

Kramer & De Jonge⁷⁶ conducted a study on the funding information present in journal articles funded by the Dutch national research funder, the Dutch Research Council (NWO). The over 5000 articles included in the study were published over the years 2011-2022 and, from the study, it was apparent that funding information developed a lot over the years, going from essentially no funder information in Crossref until 2014 to almost 100% for articles published in 2021 (2022 was still incomplete at the time the data was collected). In total, 45% of the articles included a Funder ID, of which the vast majority had been submitted by the publisher but also some that Crossref had identified and entered.

The Scholastica “State of Journal Production and Access” survey⁷⁷ had 82 respondents from academic organisations that publish independently without external publishers. Funder IDs were included by 20% of respondents.

With regards to Crossref, “(...) about 25% of Crossref records contain some kind of funding information. Over the years, this figure has grown steadily.”⁷⁸ There is no breakdown of this number for type of journal or whether PIDs were used, but the number gives some indication of what the overall uptake is for inclusion of funding information in the metadata of content.

Technical support for “Funder DOIs/PIDs” among software

⁷³ Bosman et al., 2021, p. 97.

⁷⁴ Bosman et al., 2021, p. 97.

⁷⁵ Bosman et al, 2021.

⁷⁶ Kramer, B., & de Jonge, H. (2022). The availability and completeness of open funder metadata: Case study for publications funded by the Dutch Research Council. *Quantitative Science Studies*, 3(3), 583-599. https://doi.org/10.1162/qss_a_00210

⁷⁷ Scholastica, 2022.

⁷⁸ de Jonge, H., Kramer, B. Michaud, F., & Hendricks, G. (2023, September 06). Open funding metadata through Crossref; a workshop to discuss challenges and improving workflows. *Crossref Blog*. <https://www.crossref.org/blog/open-funding-metadata-community-workshop-report/>

Identifier	OJS	Janeway	Lodel
Funder DOIs/PIDs	Supported with a plug-in, planned to be moved into OJS core https://github.com/ajnyga/funding	Supported in Janeway core	Crossref funding PID supported for journals

Funder DOIs/PIDs are supported with a plug-in in OJS and are intended to be included in the OJS core. Janeway includes funder DOIs/PIDs as a core functionality. Lodel supports Crossref funding PID for journals.

Conclusions “Funder DOIs/PIDs”

One challenge in monitoring the comprehensiveness of using PIDs for funders is that not all research is funded research, so substantial parts of the content are not within the scope of having such information added. However, there should be readiness among journals and platforms to receive and store such information in the metadata of content as funders are requiring grantees to provide such information for all funded outputs. While funder PIDs have various benefits, it does not seem like the uptake is very high yet based on available evidence. On the one hand, this is a bit surprising as various publishing platforms provide support for funder PIDs. On the other hand, it is not clear how many OADJs make use of Crossref DOIs and hence have the opportunity of adding funding information as part of the same metadata submission process. The transition by Crossref from using the Open Funder Registry (formerly FundRef) to ROR is also something that needs to be followed by publishing platforms as the development was only announced in September 2023.

4.2 Metadata

High quality metadata plays an important role in findability and interoperability in OA publishing, benefitting publishers, authors, readers and aggregators alike. Technical standards for metadata range from making metadata publicly available via standardised interfaces and vocabularies to related policies, such as self-archiving, deposition in repositories and publishing metadata under Creative Commons Zero (CC0). The DIAMAS survey from 2023 explored the challenges for publishers in relation to metadata: Responses were gathered from 577 institutional publishers whose activities included publishing scholarly journals. One of the included questions relevant for the gap analysis on metadata was one querying what type of challenges (if any) these organisations were having with “Metadata, PIDs, supplying and enriching metadata, or making metadata available for use”. The results from the 577 publishers are summarised in **Table 5**. Please note that publishers could indicate more than one challenge, so the total number of responses and % is higher than the simple total of respondents.

Lack of human resources	192	33.3%
Lack of expertise	131	22.7%
No answer	122	21.1%
This is not a challenge	111	19.2%
Financial constraints	104	18.0%
Administrative constraints	45	7.8%
Other	8	1.4%

Table 5: “Metadata, PIDs, supplying and enriching metadata, or making metadata available for use”

Lack of human resources and lack of expertise were found to be the most prominent challenges that the responding publishers highlighted. A similar question was posed for indicating potential challenges regarding “Trying to achieve and maintain interoperability with other services” with the results found in **Table 6**. Here as well, a lack of human resources together with a lack of expertise were the most prominent challenges.

Lack of human resources	169	29.3%
No answer	136	23.6%
Lack of expertise	110	19.1%
Financial constraints	106	18.4%
This is not a challenge	101	17.5%
Administrative constraints	50	8.7%
Other	8	1.4%

Table 6: “Trying to achieve and maintain interoperability with other services”

Challenges within this domain are also dealt with in the Scholastica report⁷⁹, which concludes that: “However, most publishers still appear to be working to fulfil article-level metadata best practices, like producing JATS XML⁸⁰ metadata for archiving/indexing, suggesting that meeting technical standards is not without challenges.”⁸¹

In the following section, the metadata standards identified in the survey by Armengou et al.⁸² are described individually and placed within the publication workflow of OA journals, gaps and challenges in the context of metadata properties are discussed next to a brief summary

⁷⁹ Scholastica, 2022.

⁸⁰ JATS XML stands for Journal Article Tag Suite Extensible Markup Language.

⁸¹ Scholastica, 2022, p. 40.

⁸² Armengou et al., 2023.

of the possibilities from the publishing software perspective. Similarly, to the previous section on Identifiers, conclusions on the gaps and challenges related to each standard are presented.

4.2.1 Metadata exchange for harvesting (e.g. Dublin Core, OpenAIRE, OAI-PMH s etc.)

Description of the standard “Metadata exchange for harvesting”

Metadata exchange for harvesting allows for mass-processing and harvesting of metadata following established standards such as Dublin Core (DC), OpenAIRE or OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) interfaces. The harvested metadata is used by aggregators as well as researchers and increases the findability of a journal’s publications and the interoperability of metadata. OAI-PMH is an open protocol for mass-metadata exchange by the Open Archives Initiative⁸³. It provides metadata over Hypertext Transfer Protocol (HTTP) that is based on the DC standard in XML (extensible Markup Language) encoding and may include additional formats. In this way, OAI-PMH offers information from six basic services (verbs), such as “Identify” to show all general information on the repository⁸⁴. Dublin Core is a metadata standard by the Dublin Core Metadata Initiative (DCMI) that includes a series of elements for the categorization of a digital object e.g. in terms of identifiers, formats, content or rights.

Offering metadata exchange for harvesting improves both the findability and interoperability of a journal while increasing compliance with common standards, like Dublin Core. The adherence to metadata formats in a journal’s workflow is a recurring process, while the implementation of a suitable exchange protocol, like OAI-PMH is part of the overall technical infrastructure set-up.

Where is “metadata exchange for harvesting” required?

As the survey in Armengou et al.⁸⁵ found, metadata exchange for harvesting is required by EQSIP V1.0, the OpenAIRE Guidelines for Literature Repository Managers v4, Plan S, and the DOAJ.

Challenges and gaps in relation with “metadata exchange for harvesting”

The 2023 DIAMAS survey included the following question “Does the IPSP release its metadata openly with a standard metadata description schema (MARC, MODS, DC, ONIX, JATS, TEI)?”. The distribution of the answers from the 577 responding publishers that published academic journals is found in **Table 7**. From those results, it can be derived that around 47% do so, with 19% not doing so and 27% not being sure about if they provide such metadata openly.

⁸³ <http://www.openarchives.org/>

⁸⁴ https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai_node.html

⁸⁵ Armengou et al., 2023.

Yes, under CC-BY or another Creative Commons licence	214	37.1%
Don't know	157	27.2%
No	109	18.9%
Yes, under Creative Commons Public Domain Dedication (CC0)	58	10.1%
Other	24	4.2%

Table 7: "Does the IPSP release its metadata openly with a standard metadata description schema (MARC, MODS, DC, ONIX, JATS, TEI)?"

OAI-PMH is a widely used protocol in the realm of digital libraries and repositories for the exchange of metadata. However, like any technology, it is not without its limitations. One significant technological limitation of OAI-PMH is its reliance on XML as the standard for data interchange. While XML is widely adopted, it comes with certain drawbacks. XML files tend to be large and verbose, leading to increased bandwidth and storage requirements. This can pose challenges, especially in environments with limited resources or slow network connections. Another limitation is the lack of support for real-time updates. OAI-PMH operates on a scheduled harvesting model, where repositories expose metadata at specified intervals. This asynchronous approach means that changes to the repository may not be immediately reflected in harvesting activities. For applications requiring up-to-the-minute data, this delay can be a notable drawback. Furthermore, OAI-PMH does not inherently address issues related to authentication and authorization. While there are extensions and supplementary mechanisms to enhance security, the core protocol does not provide a standardised solution. This limitation can be a concern in scenarios where strict access control is crucial.

The metadata available on the DOAJ service via the OAI-PMH protocol is split into two endpoints: the journals endpoint⁸⁶ and the articles endpoint⁸⁷. This makes the retrieval of some information require harvesting of both resources. It is also difficult to harvest collections of specific journals, as the DOAJ protocol only contains subject sets. The number of metadata formats available is also limited (oai_dc for journals, oai_dc and oai_doaj for articles), each with a fixed number of tags that may not be flexible enough to accommodate a wide range of metadata. The peculiarities of the OAI-PMH protocol and the shortcomings identified above make the filtering of harvested content very limited.

⁸⁶ <https://doaj.org/oai?verb=Identify>

⁸⁷ <http://www.doaj.org/oai.article?verb=Identify>

Metadata standard	OJS	Janeway	Lodel
Metadata exchange for harvesting	OAI-PMH supported in OJS core. OAI-PMH supports Dublin Core, MARC and RFC 1817 by default and can be easily extended to support any other metadata standard like JATS.	OAI-PMH is supported in Janeway core. Dublin Core and JATS outputs are supported with more planned.	Metadata for journals available in various standards: <ul style="list-style-type: none"> - DC, DC terms through OAI-PMH - KBART - MARC

Technical support for “metadata exchange for harvesting” among software

Metadata exchange for harvesting via established standards such as Dublin Core, OpenAIRE, OAI-PMH are supported by OJS and Lodel. More precisely, OAI-PMH is part of the OJS core and enables metadata exchange in DC, MARC (Machine Readable Cataloguing) and RFC 1817, with the possibility to include other standards too. OAI-PMH is also supported under Janeway with DC and JATS XML outputs. Lodel, as used in OE, allows for metadata exchange through OAI-PMH with DC and DC terms. KBART (Knowledge Bases and Related Tools) and MARC are also supported.

Conclusions “Metadata exchange for harvesting”

Metadata exchange for harvesting based on open standards is very important for the discoverability and indexation of published content, something which is also reflected by this being a technical feature required or recommended by many key actors. While this feature comes built-in among the investigated software solutions, the main challenge is present for those publishers that are using general-purpose content management systems or self-built solutions that do not include backend APIs for other services to harvest metadata from. In order for improvement to happen, publishers should make use of content management systems that include this functionality, whether that means updating or migrating from their current software environment.

4.2.2 OpenAIRE Guidelines

Description of the standard “OpenAIRE Guidelines”

This standard is a more specific subset of the broad category presented in the previous section covering use of essentially any open data metadata harvesting protocol. OpenAIRE is committed to the principles of Open Science by actively promoting and supporting interoperability and FAIRness, it has developed a set of metadata policies, the OpenAIRE Interoperability Guidelines. The development of the OpenAIRE Guidelines has been significantly influenced by community feedback and alignment with new developments in the Open Science landscape.

The OpenAIRE Guidelines have been built upon widely adopted standards, protocols and controlled vocabularies for data and metadata exchange. These include among others:

- The OAI-PMH protocol
- The Dublin Core, DataCite and Common European Research Information Format (CERIF) Metadata schemas
- COAR Resource Types Vocabulary
- Persistent Identifiers (PIDs) such as:
 - ORCID IDs for authors
 - Funder IDs, ROR IDs et al. for Research funding and Research Performing organisations
 - DOI, handle, ARK et al. for publications

The OpenAIRE Guidelines specify mandatory, recommended and optional elements that data sources, including platforms hosting OA Journals, should include in their metadata records to ensure compatibility with OpenAIRE. Consequently, the OpenAIRE Guidelines have gained widespread adoption among repositories throughout Europe and beyond. To accomplish the objectives of interoperability, FAIRness, and compatibility with EOSC onboarding, compatibility with the latest versions (v3.0 and v4.0) of the OpenAIRE Guidelines for Institutional and Thematic Repositories is essential.

The OpenAIRE Guidelines assist data sources, including platforms hosting OA Journals, in exposing their metadata in a way compatible with the OpenAIRE and the EOSC infrastructure. By implementing these guidelines, data sources can ensure the harvesting of their metadata records by OpenAIRE, their integration into the OpenAIRE Graph, and their inclusion in the integrated platform of the EOSC Portal Catalogue and Marketplace. Thus, the OpenAIRE Guidelines play a vital role in addressing the sustainability, openness and interoperability challenges faced by Scholarly Infrastructures.

Where are the “OpenAIRE guidelines” required?

Adherence to the OpenAIRE Guidelines for Literature Repository Managers v4 (Mandatory metadata fields) is recommended in EQSIP V1.0. Plan S recommends publication outlets to be

compliant with the OpenAIRE Guidelines for Literature Repository Managers v4 (Recommended metadata fields).

Challenges and gaps in relation with “OpenAIRE guidelines”

The OADJ Study⁸⁸ from 2021 included a question for all 1,619 responding journals querying their metadata compatibility with OpenAIRE standards. The results can be found in **Table 8** below. The responses were mainly split between “Yes” (40%) and those that reported “Unknown” (46%).

Yes	647	40%
No	106	7%
Unknown	742	46%
No answer	124	8%

Table 8: Are the journal’s metadata compliant with the OpenAIRE standard?

Metadata standard	OJS	Janeway	Lodel
OpenAIRE guidelines	OpenAIRE compliance supported with a plug-in ⁸⁹ : https://github.com/ojsde/openAIRE	Currently in the process of full compliance with OpenAIRE guidelines: DC metadata has passed the OpenAIRE validation, though additional work is required to pass all tests.	DC metadata is compliant with OpenAIRE guidelines (OE collections are harvested) https://oai-openedition.readthedocs.io/en/latest/index.html

Technical support for compliance with “OpenAIRE guidelines” among software

OpenAIRE guidelines compliance may be achieved in OJS with a plug-in. For the compatibility of OA Journals with the OpenAIRE Guidelines and the plug-in for OJS, the Journal Article Tag Suite (JATS) has been chosen. Janeway is currently in the process of achieving full compliance with the OpenAIRE guidelines with the Dublin Core metadata already having passed validation. The Dublin Core metadata of OpenEdition who are using Lodel, is compliant with the OpenAIRE guidelines and their collections are harvested.

⁸⁸ Bosman et al., 2021.

⁸⁹ Schirrwagen, J. (2019, April 26). Open Journal Systems (OJS) sets new standards to achieve OpenAIRE compliance with JATS. *OpenAIRE Blog*. <https://www.openaire.eu/blogs/open-journal-systems-ojs-sets-new-standards-to-achive-openaire-compliance-with-jats>

Conclusions “OpenAIRE Guidelines”

Journals and portals that are already providing some degree of open metadata for harvesting should look into the different levels of OpenAIRE compliance, as it is something that increases indexability and standardises the output to a certain recognized level. There is not a lot of data available about the exact share of OADJs that are compliant with the OpenAIRE guidelines at the different levels, but as this is a more specific subset of having open harvestable metadata as was explored in section 3.2.1, it can be estimated that under half of journals are compliant at the mandatory metadata field level, and less so at the recommended level.

4.2.3 Mass metadata export (as CSV files, ONIX XML feeds or in any other established format)

Description of the standard “Mass metadata export”

Mass metadata export allows for the export of metadata in an established format, such as CSV-files or ONIX⁹⁰-XML feeds. This standard supports journals in their internal organisation for back-ups or in case of migration of data even to a different publishing platform. Adhering to an established, standardised format increases the interoperability across platforms. The support for mass metadata export is also beneficial for researchers who, for example, study bibliometrics at the article-level. The use of standardised tag-sets or metadata schemes, like DC, provides searchable information on the rights and permissions of an article. Supporting mass metadata export forms a part of the technical set-up of a journal and while it requires maintenance, it does not have to be repeated in the workflow of a journal.

Where is “Mass metadata export” required?

EQSIP V1.0 states that a publishing platform “supports massive metadata export (as CSV files, ONIX XML feeds or in any other established format)”⁹¹.

Challenges and gaps in relation with “Mass metadata export”

Our data sources did not provide information on whether this is provided by journals or not.

⁹⁰ ONIX stands for Online Information Exchange.

⁹¹ Armengou et al., 2023a, p. 12.

Metadata standard	OJS	Janeway	Lodel
Mass metadata export (as CSV files, ONIX XML feeds or in any other established format)	OJS has a native XML import/export standard and has a plug-in interface that can be used to create an import/export plug-in to support any available standard.	Mass export is available via CSV, JSON and JATS XML formats.	Mass export is supported in CSV files, ONIX is also possible in some specific cases

Technical support for “Mass metadata export” among software

Mass metadata export, in file formats like CSV, ONIX-XML feeds, are available under OJS, Janeway and Lodel.

Conclusions “Mass metadata export”

There is (to our knowledge) no comprehensive data available about the uptake of this standard. However, from the scan of software platform features it can quickly be concluded that this is a feature that is quite widely supported on some level, meaning that the problem of compliance concerns mainly non-publishing oriented content management systems or self-built software solutions where this feature might be missing.

4.2.4 KBART

Description of the standard “KBART”

The National Information Standard Knowledge Bases and Related Tools (KBART) provides structured data to libraries and other actors, such as knowledge base vendors, and allows for managing electronic resources, licences, usage data or analytics.⁹² Since KBART Phase II in 2014, it has been administered by the KBART Standing Committee at NISO (National Information Standards Organisation) based in Baltimore. Publishers may use the human readable KBART title lists as one authoritative file to which other stakeholders may be referred. The usage of KBART files therefore leads to a higher level of standardisation, more consistency, fewer errors and therefore higher exposure in e.g. library records.

⁹² van Ballegoie, M., Meares, S., & Wilson, K. (2017). Deep Dive into KBART. *The Serials Librarian*, 72(1-4), 15-25. <https://doi.org/10.1080/0361526X.2017.1309826>

The metadata exchange as well as the file structure follow a recommended standard: The files should be posted and updated regularly via a Web page or File Transfer Protocol (FTP). Within the file naming scheme, the provider name, the region/consortium, the package name and the date of the file creation should be included. KBART compliance also includes adhering to a standardised file structure which includes specified fields and labels for metadata, like the publication title which is found in the field “publication_title” and follows UTF-8 encoding. The OA licence is also indicated in the KBART file.

Within the workflow of a journal, creation and maintenance of KBART files is a recurring process. Journals that provide current and accurate KBART files may benefit from a higher visibility in aggregated records like library catalogues.

Where is “KBART” required?

With regards to interoperability standards, EQSIP V1.0 outlines that the publishing platform is developed and maintained to adhere to current standards, including KBART.

Challenges and gaps in relation with “KBART”

The NISO web page⁹³ provides information on the KBART recommendations and information for prospective applicants. Potential challenges for journals are implied by the Standing Committee’s encouragement to send sample data “for review and feedback to ensure that it will meet the needs of your customers and work optimally with knowledge base suppliers’ products”. Additionally, a mailing list addressing questions regarding the KBART standard is offered. Next to the standard compliant usage of the KBART files a possible challenge may lie in the recommendation to use “public-facing and open metadata using your website”, which both have to be made available in this way and periodically maintained.

Van Ballegoie et al.⁹⁴ additionally stress the importance of the quality of metadata provided by the publishers’ side stating that “accurate data increases exposure and usage of full-text content and leads to greater interoperability and access”⁹⁵. Legacy systems may pose another challenge for journals to structure all necessary metadata in the recommended format and may pose challenges in cleaning the metadata as well. According to van Ballegoie et al., “especially for smaller publishers, these activities may compete with other priorities on an already full technical roadmap”⁹⁶.

⁹³ <https://www.niso.org/standards-committees/kbart/kbart-content-providers> as well as <https://www.niso.org/standards-committees/kbart>

⁹⁴ van Ballegoie et al., 2017.

⁹⁵ van Ballegoie et al., 2017, p. 16.

⁹⁶ van Ballegoie et al., 2017, p. 16.

Rathemacher et al.⁹⁷ (2022) report on potential challenges for journals in KBART phase III, mentioning that the KBART metadata standard is only suitable for serials and monographs. While journals clearly fall into the category of serials, related research objects may not and pose difficulties in compliance with the standardised KBART files. Additionally, the properties of hybrid OA publications appear to pose challenges which are partly resolved with a new value, “M,” for mixed content.

Metadata standard	OJS	Janeway	Lodel
KBART	Supported with a plug-in https://github.com/UB-Heidelberg/kbartExpo in OJS 3.2; support for newer versions only in OMP.	Supported in Janeway core.	Supported https://www.openedition.org/26973

Technical support for “KBART” among software

The KBART standard is supported with a plug-in in OJS so far only until version 3.2. It is supported in the Janeway core and under Lodel as used by OE.

Conclusions “KBART”

This is a standard recommended only by EQSIP 1.0, so overall, the pressure for its implementation is lesser than for standards where multiple other parties also require use of a particular standard. There is (to our knowledge) no comprehensive data available about the uptake of this standard among OADJs. Based on Rathemacher et al.⁹⁸, the use of KBART in the context of journals is not unproblematic, suggesting that the standard’s strengths lie with the description of other types of content. However, from the scan of software platform features, it can quickly be concluded that this is a quite widely supported feature on some level.

⁹⁷ Rathemacher, A., Ragucci, M., & Doellinger, S. (2022). Don’t wait, Automate! Industry Perspectives on KBART Holdings Automation. *The Serials Librarian*, 82(1-4), 91-97. <https://doi.org/10.1080/0361526X.2022.2019545>

⁹⁸ Rathemacher et al., 2022.

4.2.5 Metadata about OA status

Description of the standard “Metadata about OA status”

Information about OA status in article metadata is crucial to ensure the correct attribution of rights-holders of articles. Metadata standards, such as DC, include a tag-set for this purpose. In the case of Dublin Core metadata about OA status, it is found under the term “rights”. Adding metadata about OA status becomes part of a journal’s publication workflow and only needs to be updated if the OA status changes. Clearly stating the OA status in the metadata allows OA journals to highlight this property and might increase their visibility in indexing services.

Where is “Metadata about OA status” required?

Under Plan S, it is mandatory for compliant publication outlets to include information about the OA status in the metadata. EQSIP V1.0 requires information about the OA status in machine- and human-readable format at every publication. It is also necessary for the DOAJ Seal.

Challenges and gaps in relation with “Metadata about OA status”

In the OADJ Study⁹⁹ from 2021, there was a question included that was formulated as “Does the journal embed or display licensing information in its articles?”, where the alternatives were “Yes in PDF”, “Yes in HTML”, “Yes on the landing page”, and “No”. Overall 793 journals (49%) out of the 1,619 indicated at least one of these alternatives. 117 journals (10.9%) had the information in Portable Document Format (PDF), 116 journals (7.2%) in HTML, 169 journals (10.4%) on the landing page. 683 journals (42.2%) answered an explicit “No”.

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file there is a data field for reporting if a journal has “Machine-readable CC licensing information embedded or displayed in articles”. While being slightly more specific than the scope of the standard described here, it is one of the few data points there are. Of the 13,503 OADJs 6,158 (45.6%) journals indicated the use of machine-readable CC-licence for their articles.

⁹⁹ Bosman et al., 2021.

Technical support for “Metadata about OA status” among software

Metadata standard	OJS	Janeway	Lodel
Metadata about OA status	Journal and article OA status are stored as metadata in the system. Can be fetched for various purposes like in the case of the OpenAIRE (see: https://github.com/ojsde/openAIRE)	Article OA status is determined by the licence which Janeway stores against each article.	OA status is included in the metadata following the OpenAIRE guidelines (see: https://oai-openedition.readthedocs.io/en/latest/qdc.html#dcterms-accessrights)

OJS, Janeway and Lodel include possibilities to provide metadata about the OA status, e.g. in order to share with OpenAIRE.

Conclusions “Metadata about OA status”

This standard is possible to implement in different ways so it provides some flexibility for journals and platforms to select the way that works best for them. What seems to make most sense is implementation through the “rights” field in the DC metadata if a journal is already using such a metadata protocol, as this protocol is also widely supported by the surveyed software solutions. As such, the challenge here is highly related to the challenge of increasing the uptake of open harvestable article-level metadata overall.

4.2.6 Registering of a self-archiving policy

Description of the standard “Registering of a self-archiving policy”

A journal’s self-archiving policy describes the set of rules and norms that a journal stipulates in terms of authors’ self-archiving rights, i.e. may an author make available a copy of (or a manuscript version of) their article in an OA repository. This policy can, for instance, include statements about how the journal handles author rights to the different versions of a text, like a preprint or the author accepted manuscript (AAM). A record of this policy is placed on the website of a journal to contribute to transparency for all concerned parties. Self-archiving policies can be registered into databases such as SHERPA/RoMEO¹⁰⁰, which is currently the most comprehensive database for such information. This may generally be regarded as a one step process for journals, unless the policies need to be updated.

¹⁰⁰ <https://www.sherpa.ac.uk/romeo/>

Where is the “Registering of self-archiving policy” required?

Plan S and the DOAJ Seal recommend the registering of a self-archiving policy.

The Federation of Finnish Learned Societies (TSV) distributes subsidies granted by the Ministry of Education and Culture to scientific societies for their publishing and international activities¹⁰¹. One of the conditions of the publication subsidy is that a “publisher's open access policy must be stored in the international SHERPA/RoMEO database”¹⁰².

Challenges and gaps in relation with “Registering of self-archiving policy”

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, there is a data field called “Deposit policy directory” indicating where the journal’s self-archiving policy has been deposited (if anywhere). Of the 13,503 OADJs 3,543 (26.2%) journals indicated some location for where their self-archiving policy has been deposited.

While there is no global data available for the coverage of SHERPA/RoMEO for OADJs, there is at least one local observation that can give some anecdotal evidence. In Finland, there have been campaigns and workshops raising awareness of the importance of this practice over the years, but currently only 63 out of the 139 journals publishing on Journal.fi have registered their policy in SHERPA/RoMEO.

Conclusions “Registering of a self-archiving policy”

While some might think that being an OA journal negates the need for a self-archiving policy, that is not the case - such a policy is needed both for instructing authors of their rights before and during manuscript processing, as well as for clearly communicating where the final copyedited version can be distributed if the licence of the content does not make this clear (and even if it does, it is good to have it codified in a common database of policies). Having journal self-archiving policies stored openly in a standardised way benefits individual authors in helping them take informed action at any stage of the manuscript handling or publication process. But equally important is having them stored in machine-readable format that provides interfaces for various services to fetch data from, e.g. repositories for checking author rights in conjunction with submission of content. SHERPA/RoMEO provides both of these, so having journals submit information about their policies takes care of both the needs of humans and machines. From the perspective of the journal, this is not a highly technical standard and is roughly equal in terms for everyone since registration happens in an external service outside of the journal’s own domain; however, the data available suggests that there is still a substantial share of OADJs that have not stored their policies in any service.

¹⁰¹ <https://www.tsv.fi/en/grants/publishing-and-international-activities-scientific-societies>

¹⁰² <https://www.tsv.fi/en/grants/publishing-and-international-activities-scientific-societies/application-instructions>

4.2.7 Direct deposition in an OA repository

Description of the standard “Direct deposition in an OA repository”

To comply with this standard, all publications of a journal should immediately upon publication also be deposited in an OA repository, such as domain specific repositories, local institutional repositories or Zenodo¹⁰³. The recommendation refers here to the green OA model and implies that all publications and their respective metadata are automatically copied to an archive or repository. In this way, journal publications as well as their metadata are continuously mirrored to a secondary location/locations, making metadata and content more widely available through the repositories. Thereby, metadata and files are increasingly findable for both aggregators and researchers. The archives’ and repositories’ interfaces for metadata harvesting additionally enable this. Therefore, journals not only gain backup copy storage for their publications but also an increase in visibility and indexation. The direct deposition in OA repositories may be automatized as part of the publication workflow of a journal.

Where is the “Direct deposition in an OA repository” required?

The direct deposition of publication in an OA repository is a recommendation by Plan S.

Challenges and gaps in relation with the “Direct deposition in an OA repository”

In the OADJ Study¹⁰⁴ from 2021, there was a question formulated as “Does the journal or its publisher deposit articles (in a machine-readable community standard format such as JATS XML, and including complete metadata) into author designated or centralised Open Access repositories that fulfil the Plan S criteria?”. **Table 9** presents how the 1,619 journals responding to the survey provided their answers. The results were quite evenly divided between 33% responding “Yes”, 24% responding “No”, and 33% responding “Unknown”.

Yes	568	35 %
No	383	24 %
Unknown	542	33 %
No answer	126	8 %

Table 9: “Does the journal or its publisher deposit articles (in a machine-readable community standard format such as JATS XML, and including complete metadata) into author-designated or centralised Open Access repositories that fulfil the Plan S criteria?”

Technical support for “Direct deposition in an OA repository” among software

¹⁰³ <https://zenodo.org/>

¹⁰⁴ Bosman et al., 2021.

Metadata standard	OJS	Janeway	Lodel
Direct deposition in an OA repository	There is an actively maintained SWORD plug-in https://github.com/pkp/sword which is a protocol that for example DSpace uses.	Content can be harvested into OA repositories but no direct deposit available.	Content can be harvested into OA repositories but no direct deposit available.

OJS supports the deposition in OA repositories through a SWORD plug-in. In Janeway and Lodel, there is no direct deposit available, although content can be harvested and then placed in a repository.

Conclusions “Direct deposition in an OA repository”

Mirroring published content in open repositories has benefits for the resilience of the content and provides some assurance, in addition to, dedicated preservation services that the content will be openly available somewhere on the web for the foreseeable future even if the journal would cease to exist. At the moment, this practice is only a recommendation by Plan S and for obtaining the DOAJ Seal so it is not a basic barrier for compliance anywhere. From the survey results concerning this question, we could derive that for those global OAJs that responded around a third has compatibility with this. Based on the software overview, it looks like this functionality uses similar functions as for providing content to preservation services.

4.2.8 Metadata under CC0

Description of the standard “Metadata under CC0”

Using CC0¹⁰⁵ means to dedicate a work to the public domain. Works in the public domain can be copied, modified, distributed, etc. without asking permission or giving attribution. While it can be debated if metadata can be seen as “work” (and hence could be copyrighted at all), CC0 enables the easiest reuse and distribution of such information.

Where is “Metadata under CC0” required?

Plan S explicitly requires that metadata needs to be in the public domain.

Challenges and gaps in relation with “Metadata under CC0”

The 2023 DIAMAS survey included the following question “Does the IPSP release its metadata openly with a standard metadata description schema (MARC, MODS¹⁰⁶, DC, ONIX, JATS,

¹⁰⁵ <https://creativecommons.org/publicdomain/zero/1.0/deed.en>

¹⁰⁶ The acronym MODS stands for Metadata Object Description Schema

TEI)?”. The distribution of the answers from the 577 responding publishers that published academic journals is found in **Table 10**. Based on this information, only about 10% of institutional publishers that responded to the survey are compliant with this requirement.

Yes, under CC-BY or another Creative Commons licence	214	37.1%
Yes, under Creative Commons Public Domain Dedication (CC0)	58	10.1%
No	109	18.9%
Don't know	157	27.2%
Other (please specify)	24	4.2%
No response	15	2.6%

Table 10: “Does the IPSP release its metadata openly with a standard metadata description schema (MARC, MODS, DC, ONIX, JATS, TEI)?”

Metadata standard	OJS	Janeway	Lodel
Metadata under CC0	All metadata can be made available for harvesting via OAI-PMH. Metadata licence is the journal's policy decision, not prescribed by the platform. PKP recommends instead that journals add a statement somewhere on their website that metadata are released under CC0 (https://docs.pkp.sfu.ca/plan-s/en/#article-metadata-quality)	Metadata is available for harvesting over OAI-OMH and REST API. Journal sets the metadata licence policy.	The OE metadata is placed under CC0 and can be accessed via the indexing systems ¹⁰⁷ : – Coverage list – OAI-PMH repository – MARC record

Technical support for “Metadata under CC0” among software

¹⁰⁷ OpenEdition Team (2023, February 02). Licensing the metadata describing content published on OpenEdition. *Open Electronic Publishing*. <https://oep.hypotheses.org/3245>

The placement of metadata under CC0 is technically possible though OJS, Janeway and Lodel, although oftentimes the metadata licence depends on the journals' policies.

Conclusions “Metadata under CC0”

In order for metadata to be easily aggregated into different types of services, a CC0 licence removes any uncertainty about rights and potential legal obligations when the metadata is reused. From the scarce evidence there is on this standard, only a minority of OADJs have explicitly licensed their metadata CC0, even though it should be fairly easy to perform if the journal is already making openly structured metadata available through a publishing content management system.

4.2.9 Open Citations" standard compliance

Description of the standard “Open Citations" standard compliance”

Citations represent the foundational information sources that not only offer origin and rationale for established facts but also facilitate the acknowledgment and recognition of contributions to scholarship. Moreover, citations empower us to assess research and its influence. In essence, citations stand as the foremost instrument for the exploration, distribution and assessment of all forms of scholarly wisdom. However, citation information is typically not readily accessible, often encumbered by inconsistent and challenging-to-interpret licences, and generally lacks machine-readable formatting. Therefore, the Initiative for Open Citation (I4OC¹⁰⁸) aims to enhance the accessibility of structured, separable and open citation data. Structured implies that the information for every publication and citation instance is presented in standardised, machine-readable formats, enabling programmatic access. Separable indicates that citation instances can be examined independently without requiring access to the original bibliographic products, such as journal articles and books, where the citations originate. Open signifies that the data is openly available and can be freely accessed and reused. The inception of a worldwide, publicly accessible network of interconnected scholarly citation data could improve the discoverability of published content, irrespective of subscription-based or OA sources, and is particularly advantageous for individuals who lack access to commercial citation databases through academic institutions. The standard basically means that publishers should deposit structured versions of the reference list of published articles to Crossref, in conjunction with the deposit of all other content metadata.

Where is “Open Citations" standard compliance” required?

Making citation data publicly available, e.g. by including it into the Crossref metadata when registering DOIs, is required by EQSIP V1.0 and recommended by Plan S.

¹⁰⁸ <https://i4oc.org/>

Challenges and gaps in relation with “Open Citations standard compliance”

In the OADJ Study¹⁰⁹ from 2021, there was a question formulated as “Does the journal provide openly accessible data on citations according to the standards of the Initiative for Open Citations (I4OC)?”. **Table 11** presents how the 1,619 journals responding to the survey provided their answers. 24% of journals indicated “Yes”, 28% “No”, and 37% “Unknown. 11% did not answer the optional question.

Yes	390	24 %
No	450	28 %
Unknown	593	37 %
No answer	186	11 %

Table 11: “Does the journal provide openly accessible data on citations according to the standards of the Initiative for Open Citations (I4OC)?”

In the DIAMAS survey from 2023, responses were gathered from institutional publishers whose activities included publishing scholarly journals. The survey included the following question "Please consider the following statements, and mark those that are implemented at IPSP level: - Make references openly available according to the principles of I4OC (Initiative for Open Citations)". The results are found in **Table 12**. The total number of valid responses to this question was 573. One answer was adjusted since two incompatible alternatives were adjusted (respondent had selected both some and all journals, the all journals option was selected to remain). The main takeaway from these responses is that 43% of all journal publishers responding to the survey reported to be compliant with the requirements of the I4OC for all their journals.

Yes, for all journals	247	43.1%
Yes, for some journals	23	4.0%
No	76	13.3%
I don't know	123	21.5%
Not applicable	40	7.0%
No answer	64	11.2%

Table 12: IPSP making references openly available according to the principles of I4OC (Initiative for Open Citations)

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, of the 13,503 OADJs 2,164 (16%) indicated compliance with I4OC standards for open citations, while 4,587 (34%) indicated non-compliance. For 7,101 journals (52.6%), there is no information about this in the metadata.

¹⁰⁹ Bosman et al., 2021.

Technical support for “Open Citations standard compliance” among software

Metadata standard	OJS	Janeway	Lodel
Open Citations	A plug-in allows to transfer reference lists to Crossref. However, this citation data is unstructured unless the reference list is provided in structured JATS XML	Where structured references are supplied (via JATS), Janeway can deposit these with Crossref, thereby feeding into the respective OpenCitations Index	Supported

OJS enables the transfer of reference lists to Crossref, working best when the reference lists are provided as structured JATS XML. If metadata is supplied in the JATS XML format, it may be deposited in Crossref under Janeway. In this way, the metadata will be made available as part of the COCI Open Citations Index data set. Lodel provides support for the Open Citations standard.

Conclusions “Open Citations standard compliance”

Having citations provided among open metadata deposits enhances the possibilities of what can be done in terms of service provision and bibliometric research without relying on closed commercial databases for providing similar insight into which content is citing what other content. In principle, this should be fairly straightforward to achieve for journals that deposit metadata to Crossref as it is possible to deposit this metadata together with all other content-related metadata as part of the same process. From the studies done on the uptake of this standard among OADJs the results varied between 24%-43% shares among the respondents. Given that the uptake of fulltext JATS XML is not very high (cf. section 3.3.2) and that extra efforts are required to structure reference lists (i.e. citation data), it is not entirely clear whether these numbers indeed refer to structured citation data or “just” to the inclusion of unstructured reference lists into metadata. One can conclude that technologically there is already readiness to have the gap closed, the content management systems support this practice, but as fulltext JATS XML is required for best data quality broader uptake hinges on wider adoption of JATS XML.

4.3 Content

The content of publications is usually the main interest of human and machine readers when using scholarly literature. One of the core functionalities of OA, according to. e.g., the Berlin Declaration¹¹⁰, is access to the content. Hence, the content itself should be as accessible as possible and reuse should be facilitated easily. Also, digital preservation is listed as an important asset of OA in the Berlin declaration.

4.3.1 Human- and machine-readable information about the open access status, copyright holder and licensing in each publication in a standard non-proprietary format

Description of the standard “Human- and machine-readable information about the open access status, copyright holder and licensing in each publication in a standard non-proprietary format”

This standard overlaps with another more specific standard (Metadata about OA status) but includes a lot of requirements in addition to making it much more comprehensive. For both human and machine users, it is crucial to know under which conditions a scholarly work can be reused, i.e. who is the rightsholder and needs to be attributed. Hence, information on these aspects needs to be available in/on the actual documents but also in the respective metadata (compare section on metadata standards) in machine- and human-readable format. To ensure this information is presented in a comprehensible way, journals need a clear OA statement on their websites. Moreover, each article needs to display information on licence and copyright on the actual document but also on its landing page. Publishing systems, such as OJS and others, offer global copyright and licence settings for a given journal. During the production process, editors or (institutional) publishers/platforms need to make sure to include this information in output formats such as PDF, HTML, or (JATS) XML.

Where is “Human- and machine-readable information about the open access status, copyright holder and licensing in each publication in a standard in a non-proprietary format” required?

Plan S, DOAJ and EQSIP require information along these lines. Also, OpenAIRE promotes a tag indicating whether the provided content is OA.

¹¹⁰ <https://openaccess.mpg.de/Berliner-Erklaerung>

Challenges and gaps in relation with “Human- and machine-readable information about the open access status, copyright holder and licensing in each publication in a standard non-proprietary” format

In the OADJ Study¹¹¹ from 2021, there was a question included that was formulated as “Does the journal embed or display licensing information in its articles?”, where the alternatives were “Yes in PDF”, “Yes in HTML”, “Yes on the landing page”, and “No”. Overall 793 journals (49%) out of the 1,619 indicated at least one of the “Yes” alternatives. 117 journals (10.9%) had the information in PDF, 116 journals (7.2%) in HTML, 169 journals (10.4%) on the landing page. 683 journals (42.2%) answered an explicit “No”

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, there is a data field for reporting if a journal has “Machine-readable CC licensing information embedded or displayed in articles”. While being slightly more specific than the scope of the standard described here, it is one of the few data points there are. Of the 13.503 OADJs 6.158 (45.6%) journals indicated that they include CC licence information embedded or displayed in articles.

Technical support for “Human- and machine-readable information about the open access status, copyright holder and licensing in each publication in a standard non-proprietary format” among software

Content standard	OJS	Janeway	Lodel
Human- and machine-readable information about the OA status, copyright holder and licensing in each publication in a standard non-proprietary format	Landing page includes human readable information regarding licensing and copyright. This data can be made available in machine-readable using for example OAI-PMH, as an example see the OpenAIRE plug-in	Article pages include human readable information on licensing and copyright as well as machine readable DC and citation meta tags.	Supported: information is given online in the HTML version, and in machine-readable format (eg: CC licence in DC metadata)

Human- and machine-readable information about the OA status, copyright holder and licensing in each publication in a standard non-proprietary format is supported by all systems presented in this report. OJS offers this information human-readable on the landing pages and machine-readable via interfaces like OAI-PMH. In Janeway, the article pages include

¹¹¹ Bosman et al., 2021.

human as well as machine-readable information following the DC metadata standard. In Lodel as used by OE, this information is provided in the HTML versions and in the DC metadata.

Conclusions “Human- and machine-readable information about the open access status, copyright holder and licensing in each publication in a standard non-proprietary format”

There is no single data point that would be able to cover all of the ground that this standard entails, but from what can be gathered, OADJs are fairly well-equipped to cater to this requirement at least in terms of licensing and OA status information. When it comes to copyright holder information, we could not locate a data source that would be able to provide statistics on this element, so more investigation would be needed to make judgement on what the current practices for journals are in this regard, particularly concerning machine-readable information.

4.3.2 Full text in machine-readable format (JATS XML or equivalent (e.g. TEI))

Description of the standard “Full text in machine-readable format”

This standard mandates the publication of full text articles in a machine-readable format, such as XML, that serves as an open, flexible and convenient choice. It also requires the use of a standardised set of tags and attributes designed especially for scholarly articles, such as JATS (Journal Article Tag Suite), which is the most common standard in scholarly publishing, or TEI (Text Encoding Initiative).

Journal editors are required to generate XML files for each article, which can be accomplished using a professional tool (e.g. eXtyles), a combination of tools (e.g. OJS + Typeset) or by outsourcing the task. This process typically takes place before the online publication of the article and XML is commonly published alongside PDFs and other document formats.

XML’s high adaptability and openness make it a convenient choice for journals since it can be easily converted into various formats (e.g. HTML). Its open structure not only enhances the discoverability of the article but also simplifies the process of mining text, data and references. (JATS) XML is appropriate for long-term preservation of content due to its platform-independent nature, which ensures that it can be used in the future, regardless of changes in technology.

Where is “Full text in machine-readable format” required?

Plan S contains a strong recommendation for full text in machine-readable format, while it is required in the EQSIP V1.0 standard. The JATS XML format is, for instance, also mandatory at PubMed Central, JSTOR, Portico and Scientific Electronic Library Online (SciELO).

Challenges and gaps in relation with “Full text in machine-readable format”

In the OADJ Study from 2021, there was a question formulated as “Please indicate which formats of full text are available”. Responses from the 1,619 journals can be found in **Table 13**. From the results, only 13% report to provide XML-based output, however, 28% provide HTML which, depending on notation and implementation, is often better for machine readability than PDF.

XML	HTML	PDF	DOC	TXT	ePub	Mobi	[Other]	No answer
205	457	1,483	11	2	67	5	24	98
13 %	28 %	92 %	1 %	0 %	4 %	0 %	1 %	6 %

Table 13: “Please indicate which formats of full text are available”

In the DIAMAS survey from 2023, responses were collected from 577 institutional publishers whose activities encompass the publishing of scholarly journals. The survey included the following question “Which formats does the platform make content available in?”. The results are found in **Table 14**. Here, the results in terms of machine-readability are a bit improved from the earlier mentioned and conducted Diamond journal study, with 19% providing XML output and 49% HTML.

PDF	550	95.3%
HTML	229	39.7%
XML	111	19.2%
ePub	91	15.8%
Image or video formats (e.g. mp4, .mov)	72	12.5%
Data formats, e.g. csv	43	7.5%
Sound files (e.g. mp3, .wav)	40	6.9%
JSON	7	1.2%

Table 14: “Which formats does the platform make content available in? “

Technical support for having “Full text in machine-readable format” among software

Content standard	OJS	Janeway	Lodel
Full text in machine-readable format (JATS XML or equivalent (e.g. TEI))	OJS supports any format the journal is capable of creating	JATS XML files are supported and stubs (head metadata) can be generated.	Supported (TEI XML)

Full text in machine-readable format (JATS XML or equivalent e.g. TEI) are supported by OJS, Janeway and Lodel.

Conclusions “Full text in machine-readable format”

From the review, it can be concluded that only a minority of journals are currently doing copy editing and publishing in XML (only 19% of journal publishers that responded to the DIAMAS survey). Based on the scan of software support for XML publishing, all three pieces of software had full support for such content, suggesting that for journals already operating in any of those software environments the challenges in conducting XML copyediting and publishing lie elsewhere, likely due to the lack of resources or expertise, but that would still require further investigation to confirm since the motivations are not observable based on the existing survey materials.

4.3.3 Text and Data Mining is technically supported

Description of the standard “Text and Data Mining is technically supported”

Text and Data Mining (TDM) refers to the automated procedure of sifting through extensive textual or data sources. Its primary objectives encompass tasks such as information retrieval, pattern identification, relationship exploration, semantic analysis and the elucidation of how content aligns with concepts and requirements. TDM yields valuable insights crucial for academic studies, research and various other purposes.

Availability of metadata and full texts through standardised interfaces might help enabling TDM as easy harvesting for machines is facilitated. With regard to formats, PDFs are not optimal for TDM since information is usually not entailed in a structured way¹¹². Texts in machine-readable formats such as XML enable TDM in a better way. Providing full-text XML requires a respective markup of all elements within an article during production. There are different XML formats, the most common in journal publishing is JATS¹¹³ (see the next standard in this section for a closer look at that).

For example, OJS has a built-in solution for having an OAI-PMH interface for metadata harvesting. Setting this up is a one-time action. However, deeper support for TDM requires continuous work in the production workflow for every article. While full text XML flows can be (partly) automated, often additional manual work is required.

¹¹² Konrad, M. (2016, July 04). Data Mining PDFs – The Simple Cases. *WZB Data Science Blog*.

<https://datascience.blog.wzb.eu/2016/07/04/data-mining-pdfs-the-simple-cases/>

¹¹³ <https://jats.nlm.nih.gov/index.html>

Where is “Text and Data Mining is technically supported” required?

EQSIP V1.0¹¹⁴ underlines that the publishing platform should support TDM, such as automatic downloading, extraction of texts as well as metadata, and that this is included in the relevant policies.

Challenges and gaps in relation with “Text and Data Mining is technically supported”

In the OADJ Study¹¹⁵ from 2021, there was also a question formulated as “Does the journal allow, legally and technically, Text and Data Mining of the full text of articles by third parties?” Responses from the 1,619 journals can be found in **Table 15**. While this standard does not consider legal aspects in its scope, the responses can still give some relevant insight. 58% of responding OADJs provide unrestricted TDM access while a further 15% does so either on request or in some restricted capacity.

Yes, unrestricted	933	57.6 %
Yes, on request	126	7.8 %
Yes, restricted	115	7.1 %
No	214	13.2 %
No answer	231	14.3 %

Table 15: “Does the journal allow, legally and technically, Text and Data Mining of the full text of articles by third parties?”

As was mentioned in the introductory text to this standard, PDF as a format is not optimal for TDM, so some insight into what formats OADJs make their full text content available is relevant. In the OADJ Study¹¹⁶ from 2021, there was a question formulated as “Please indicate which formats of full text are available”. Responses from the 1,619 journals were presented in section 3.3.2, **Table 13** - repeated here as **Table 16** for convenience. 92% of OADJs make their content available as PDFs, with 13% and 28% doing so in either XML or HTML respectively, which are both superior to PDF in terms of their readiness to be machine-read and analysed at scale.

XML	HTML	PDF	DOC	TXT	ePub	Mobi	[Other]	No answer
205	457	1,483	11	2	67	5	24	98
13%	28%	92%	1%	0%	4%	0%	1%	6%

Table 16: “Please indicate which formats of full text are available”

For the DIAMAS survey from 2023, responses were gathered from 577 institutional publishers who were involved in publishing scholarly journals. The survey included the following

¹¹⁴ Armengou et al., 2023a.

¹¹⁵ Bosman et al., 2021.

¹¹⁶ Bosman et al., 2021.

question “Which formats does the platform make content available in?”. The results are found in **Table 17**. Here, PDFs were again the dominant format, 95%, with 59% having either XML or HTML.

PDF	550	95.3%
XML	111	19.2%
HTML	229	39.7%
ePub	91	15.8%
Image or video formats (e.g. mp4, .mov)	72	12.5%
Data formats, e.g. csv	43	7.5%
Sound files (e.g. mp3, .wav)	40	6.9%
JSON	7	1.2%

Table 17: “Which formats does the platform make content available in?”

Technical support for “Text and Data Mining is technically supported” among software

Content standard	OJS	Janeway	Lodel
Text and Data Mining is technically supported	OAI-PMH can be used for Text and Data Mining. LOCKSS and CLOCKSS optionally supported, with a registry page provided for scraping via the LOCKSS toolkit.	Article content can be mined via OAI-PMH, LOCKSS and CLOCKSS.	Supported but not fully open (all contents structured in TEI XML available on demand)

TDM is technically supported under OJS, Janeway and Lodel. In OJS and Janeway, TDM is possible via OAI-PMH, LOCKSS (Lots Of Copies Keep Stuff Safe)¹¹⁷ or CLOCKSS (Controlled LOCKSS)¹¹⁸ In the OpenEdition implementation of Lodel, structured TEI-XML documents are available on demand.

¹¹⁷ <https://www.lockss.org/>

¹¹⁸ <https://clockss.org/>

Conclusions “Text and Data Mining is technically supported”

Similar to some other standards, this overlaps a bit with some others that are also included in the review. Open harvesting of metadata (e.g. through OAI-PMH) facilitates article-level discovery and processing, and the use of machine-readable publication formats are both highly relevant to the composition of this standard. From the different results, we could gather that around half of OADJs seem to have technical readiness in terms of publication formats to enable TDM, with around two thirds of journals that allow this in an unrestricted way.

4.3.4 Deposited in a digital preservation service

Description of the standard “Deposited in a digital preservation service”

By depositing all content in a digital preservation service, journals ensure long-term archiving despite technical transformations or the closure of a journal or its publication platform. Hence, digital preservation aims to make content continuously available and usable, adapting to issues like changing or degrading technologies.

Widely-used preservation services include Portico and LOCKSS/CLOCKSS. Portico is run by the U.S. non-profit organisation ITHAKA, which also owns the JSTOR database and hosts archived digital objects on their servers in Princeton, as well as two copies each on different back-up servers¹¹⁹. LOCKSS was developed at Stanford University and uses a peer-to-peer network of libraries worldwide to store preserved content.¹²⁰ CLOCKSS is run by a non-profit organisation and uses the LOCKSS technology, maintaining 12 mirror repositories.¹²¹

Information on digital preservation status on all digital objects with an ISSN may be found in the Keepers Registry integrated in the ISSN Portal¹²². PKP provides a preservation network and a respective plug-in for OJS journals that are not included in other digital preservation services under the name PKP PN¹²³. The network consists of eight preservation nodes, distributed in North America and Europe. Project JASPER (Journals are Preserved forever)¹²⁴, led by DOAJ, particularly aims at archiving OA journals and is a collaboration of CLOCKSS, DOAJ, Internet Archive, Keepers Registry and PKP.

Journals have to set up the disposition in a suitable preservation service as part of their technical platform. After the initial set up and in view of regular updates, it represents a non-recurring process in the workflow as the deposition may be automatized, otherwise it

¹¹⁹ Portico Replication and Backup Policy, v. 1.3, <https://www.portico.org/preservation-policies/>

¹²⁰ <https://www.lockss.org/about>

¹²¹ <https://clockss.org/about/>

¹²² <https://keepers.issn.org/>

¹²³ <https://pkp.sfu.ca/pkp-pn/>

¹²⁴ <https://doaj.org/preservation/#open-access-journals-must-be-preserved-forever>

becomes a more prominent part of the publication workflow. Archiving content in a digital preservation service safe-guards a journal’s content from loss due to changing organisational structures, changes in technology or file formats.

Where is “Deposited in a digital preservation service” required?

EQSIP V1.0 as well as Plan S mandate the depositing in the digital preservation service and it is also required for the DOAJ Seal.

Challenges and gaps in relation with “Deposited in a digital preservation service”

The OADJ Study¹²⁵ from 2021 included a question “What digital archiving policy does the journal use?” for which respondents could tick any applicable options reflecting the situation of their own journal. The outcome of this question is presented in **Table 18**. 855 (53%) journals out of the total 1,619 respondents had ticked the “No policy in place” option, with a further 123 (8%) journals not selecting any option. The rest of the responses were distributed in a range between 0% and 11% of the listed service providers.

No policy in place	CINES ¹²⁶	CLOCKSS	LOCKSS	PKP PN	PMC/ Europe, PMC, PMC/ Canada	Portico	A National Library	Other	No answer
855	6	87	136	91	15	67	170	157	123
53%	0%	5%	8%	6%	1%	4%	11%	10%	8%

Table 18: What digital archiving policy does the journal use?

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, of the 13,503 OADJs 3,095 (22.9%) are indicated to be included in the coverage of a preservation service, and 1,273 (13.4%) journals with content being included in a national library preservation service. Since some journals are in both coverage, we calculated the share of journals not covered by either category of preservation and that figure was 9,519 journals (70.5%).

577 institutional publishers whose activities included publishing scholarly journals responded to the DIAMAS survey from 2023. The survey included a question concerning which, if any, preservation services the publisher's content was actively preserved in. The results are found in **Table 19** (publishers could choose multiple answer alternatives).

¹²⁵ Bosman et al., 2021.

¹²⁶ The acronym CINES stands for Centre Informatique National de l'Enseignement Supérieur.

	Yes	No	Don't know	No answer
Portico	40 (6.9%)	175 (30.3%)	76 (13.2%)	286 (49.6%)
CLOCKSS	61 (10.6%)	169 (29.3%)	82 (14.2%)	265 (45.9%)
LOCKSS	55 (9.5%)	171 (29.6%)	82 (14.2%)	269 (46.6%)
PKP PN	68 (11.8%)	164 (28.4%)	84 (14.6%)	261 (45.2%)
PubMed Central	23 (4%)	176 (30.5%)	75 (13%)	303 (52.5%)
National / institutional library or infrastructure	319 (55.3%)	73 (12.7%)	53 (9.2%)	132 (22.9%)
Other	48 (8.3%)	18 (3.1%)	30 (5.2%)	48 (83.4%)

Table 19: "Is the published content actively preserved in a digital preservation service?"

Here is a quick overview of the distribution: 140 publishers (24.3%) did not provide any "Yes" answer to any of the service alternatives, 320 (55.5%) gave 1 "Yes" answers, 75 (13%) gave 2 "Yes" answers, 30 (5.2%) gave "Yes" answers, 11 (1.9%) gave 4 "Yes" answers, and 2 (0.3%) gave 5 "Yes" answers".

Another of the questions included in the DIAMAS survey relevant for this section was one querying what type of challenges (if any) these organisations were having with "Archiving, backing up or preserving content and software". The results for the 577 publishers are summarised in **Table 20**. Please note that publishers could indicate more than one challenge, so the total number of responses and % is higher than the simple total of respondents.

Financial constraints	126	21.8%
Lack of human resources	143	24.8%
This is not a challenge	132	22.9%
Other	7	1.2%
Lack of expertise	74	12.8%
Administrative constraints	38	6.6%
No answer	134	23.2%

Table 20: "Archiving, backing up or preserving content and software"

Content standard	OJS	Janeway	Lodel
Deposition in a digital preservation service	Supported with plug-ins including PKP PN service https://pkp.sfu.ca/pkp-pn/ , Archivematica, LOCKSS/CLOCKSS, and others.	Janeway supports archiving in CLOCKSS, LOCKSS, Portico and the Library of Congress.	Supported (French national digital archive CINES)

Technical support for “Deposited in a digital preservation service among software

Deposition in a digital preservation service is technically supported by OJS, Janeway and Lodel. OJS, for example, includes a plug-in for the PKP PN. Under Janeway, the long-term digital preservation of content is possible in CLOCKSS, LOCKSS, Portico or the Library of Congress. Under Lodel, OE makes use of the French national digital archive Cines.

Conclusions “Deposited in a digital preservation service”

Of all the standards, this is the one with potentially the most serious consequences if a journal is not compliant - meaning that the entire content of the journal becomes inaccessible should anything happen to the original publisher’s website and its files. From the available evidence, it seems that around half of OADJs are not enrolled in any preservation service, which is concerning for the integrity of the scholarly record. The DIAMAS survey provided valuable insight into why this might be the case, suggesting that financial constraints and lack of human resources are contributing factors to lacking enrollment. From the overview of software features in publishing content management systems, preservation is at this stage a built-in feature, which means that it is not a major technical obstacle as long as one is running an up-to-date version of the software and has enabled the feature.

4.4 Website features

Websites are the “shop windows” for journals. However, a nice look is not everything. There are several technical functionalities that can enhance journals’ findability on the web and better website usability for both human and machine users. Article landing pages not only provide direct gateways to the content (compare content section above), but also offer the opportunity to link to other research outputs or underlying materials relating to the presented research. Moreover, usage statistics can be presented there.

In addition to the technical requirements and standards presented, T6.3 “Single Sign-On access to services and content” of the CRAFT-OA project is working on a single sign-on access

to services and content. The aim is to facilitate the IPSPs and IPTPs (Institutional Publishing Technology Provider) to connect their services and to easily connect with the EOSC. Since this might be of interest to the readers of this report, the basic concept of the Authentication and Authorization Infrastructure (AAI) is plotted in the Annex 1.

4.4.1 SEO

Description of the standard “SEO”

Search engine optimization (SEO) is a process which journals may use to increase their visibility in search engines. This may include issues ranging from using a suitable domain name, implementing certain meta-tags in the HTML or including certain keywords to optimising a website for mobile use and strategically linking networks to other web pages. In general, optimising a journal's website's accessibility¹²⁷ will positively affect SEO. On an article level, the inclusion of meta-tags in the HTML may be used to optimise indexing in Google Scholar and search engines or aggregators. While updating the journal website and the content management system are recurring tasks, the setup of a web page, its structural features and a suitable article template may only be required once. Increasing their visibility in public search engines may benefit journals with reaching a larger number of potential authors as well as readers.

Where is “SEO” required?

SEO is a requirement in the EQSIP V1.0 where the section “Visibility, Indexation, Communication, Marketing and Impact” states that “IPSPs make sure that visibility of publications in search engines (general and academic) and aggregators is improved by using search engine optimization techniques ...”¹²⁸.

Challenges and gaps in relation with “SEO”

The DIAMAS survey from 2023 distributed to 577 institutional publishers publishing academic journals included the question “Is the IPSP satisfied with the level to which its published content is included in scholarly search engines and different indexes?”. 248 (43%) publishers selected the “Our content is already very well indexed” response, 292 (50.6%) the “We would like to see (better) indexing in these search engines”, and 37 (6.4%) provided no answer to this question.

¹²⁷ <https://www.w3.org/WAI/fundamentals/accessibility-intro/>

¹²⁸ Armengou et al., 2023a, p. 14.

Website feature requirement	OJS	Janeway	Lodel
SEO	<p>SEO depends partly on the chosen theme plug-in. Thereby, the default theme performs very well in accessibility https://docs.pkp.sfu.ca/accessibility-statement/en/. There are also Google Scholar and DC plug-ins to add Highwire and DC meta-tags. Also possibility to add custom HTML meta-tags to the page header.</p>	<p>Some limited SEO is available in HTML and meta tags.</p>	<p>Supported (HTML metadata, search interface)</p>

Technical support for “SEO” among software

SEO measures are supported by OJS, Janeway and Lodel to varying degrees.

Conclusions “SEO”

SEO is an important topic for publishers in order to ensure that relevant content is found by interested users both through specific scholarly literature discovery tools as well as general purpose web search engines. From the results of the DIAMAS survey, we could perceive that around half of journal publishers would like to see improved discovery and indexation for their published content. Through the scan of technical support for SEO among publishing software, we could see that there is some support available, but there is not one specific approach or technical solution that ensures good SEO. Rather, both machine-readable metadata as well as the choice of visual theme might have implications for how well the content is harvested by various service providers.

4.4.2 Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review (where relevant)

Description of the standard “Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review (where relevant)”

This standard takes a variety of potential forms, only a selection of which can be outlined here to illustrate it. Alerting services may be used by a journal to announce, e.g. new publications or news relevant to the organisation. This could be implemented using an (automated) Rich Site Summary (RSS) feed or even a newsletter. It could also include (automatically) sharing this information on a Social Media platform. Having regular alerting services active as a journal may increase its visibility and enable interested parties to follow its activity both from a reader as well as an author or even reviewer’s perspective. Post-publication evaluation and commenting could, for instance, be realised by linking to a collaborative annotations platform, like Hypothesis¹²⁹. Community interaction and engagement with a journal’s publications may be increased in this way.

Support for multimedia can mean that videos are supported by a website but also, for instance, that animated or interactive graphs are enabled. By allowing for a wider range of media to be supported, a journal may lower the threshold for readers and add to the range of research communication for authors. There are many forms of open peer review¹³⁰, one of them being that in the review process all parties are aware of each other’s names and the reviews are published together with the articles. In this way, a constructive conversation between researchers is supported and the peer-review process gains added transparency. Adding alerting services, automated sharing to social media, post-publication evaluation and commenting, support for multimedia and open peer-review may be implemented during the technical set-up of a webpage and maintained alongside it.

Where are “Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review” required?

EQSIP V1.0 specifies that a suitable publishing platform includes the functionalities to allow for alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review.¹³¹

Challenges and gaps in relation with “Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review”

The DIAMAS survey from 2023 included the question “Does the IPSP [institutional publishing service providers] have a newsletter or social media or networking profiles to inform the

¹²⁹ <https://web.hypothes.is/>

¹³⁰ cf. Ross-Hellauer, T. (2017). What is open peer review? A systematic review [version 2; peer review: 4 approved]. *F1000Research* 2017, 6:588, <https://doi.org/10.12688/f1000research.11369.2>

¹³¹ Armengou et al., 2023a, p. 14.

community about updates?”. 379 (65.7%) publishers provided the answer “Yes”, 168 (29.1%) “No”, 16 (2.7%) “Don’t know”, and 23 (4%) did not provide any response.

Investigating the data from OADJs based on DOAJ’s publicly available journal-level metadata file, of the 13,503 OADJs 99 (0.7%) indicated to be implementing “Open peer review” processes, and 13 (0.1%) journals “Post-publication peer review”, “Open Peer Commentary”, or “Community review” processes.

Technical support for “Alerting services, sharing to social networks, post- publication evaluation and commenting, support for multimedia and open peer review” among software

Website feature requirement (where relevant)	OJS	Janeway	Lodel
Alerting services	Features to alert all registered users when a new issue or an announcement has been published are part of OJS Core.	The core of Janeway supports ATOM+RSS and email alerts for readers.	Supported (RSS)
Sharing to social networks	AddThis plug-in		
Support for multimedia			
Open peer review, including post-publication evaluation and commenting	Sharing of author/reviewer identities during peer review (and direct communication between the two parties) can be opted for in the set-up of the peer review process in OJS Core. Hypothes.is and Disqus plug-ins can be used for post-publication commenting.	Open peer review is supported in Core with metadata deposited via Crossref. Commenting is available via plug-in.	

Support for alerting services, sharing to social networks, post-publication evaluation and commenting, multimedia and open peer review (where relevant) is available with OJS, Janeway, and Lodel to a varying degree.

Conclusions “Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review”

The features covered in this section cover a large span of different aspects that all can be implemented differently since there is no strict standard for these. Based on the overview of OADJs, only to a minimal degree are they engaging with open peer review and commenting practices. Two thirds of the journal publishers responding to the DIAMAS survey reported having some way of notifying users of new content, which indicates some room for growth in this regard. Technically, these aspects seem to be well catered to in modern publishing content management software, where also multimedia files can be provided in addition to text files.

4.4.3 Unique URLs for landing pages

Description of the standard “Unique URLs for landing pages”

Unique URLs for landing pages enable a unique linking to both the journal and the article level. This makes all published items uniquely identifiable but also means that this should be rooted in the journal’s web page structure. It would therefore not be possible to make all articles available as e.g. a list of downloadable articles under the same URL. Each article requires its own unique URL to fulfil this standard. As with other URL-based modes of identification, they have to be updated on levels like published metadata if the unique URL for a landing page changes. This is a recurring task for journals while a suitable web site structure can be relied on after one initial set up.

Metadata on article landing pages improves the visibility (SEO) of the journal. This makes it easier for interested readers to discover the journal's content through online searches.

Where is a “Unique URLs for landing pages” required?

EQSIP V1.0 states that “each published item (article, chapter, book, etc.) has a dedicated unique URL (landing page)...”¹³². Similarly, DOAJ requires unique URLs for landing pages in their guidelines.

Challenges and gaps in relation with “Unique URLs for landing pages”

Since this is a basic requirement for inclusion in the DOAJ, we can derive that all 13,503 journals included in the journal-level metadata set fulfil this criteria.

¹³² Armengou et al., 2023a, p. 12.

Website feature requirement	OJS	Janeway	Lodel
Unique URL for landing pages	Supported in OJS core	Supported in Janeway core.	Supported

Technical support for “Unique URLs for landing pages” among software

Unique URLs for landing pages are supported in the OJS core and Lodel.

Conclusions “Unique URLs for landing pages”

The nature of this standard is fairly simple from a technical perspective and is not a very challenging standard to comply with for OADJs. It is supported in all publishing content management systems we looked into.

4.4.4 URLs linking to related research objects

Description of the standard “URLs linking to related research objects”

In the case of journals and published articles, related research objects may include objects such as research data, samples, code or videos. These can be embedded in the website or presented as external links to the respective research object, which could be stored in a repository like Zenodo. In this way, connections between research objects become more transparent and the different types of research objects are highlighted, thus enabling the transparency of connections beyond the referenced citations in an article.

To conform with this standard, journals have to set up the structure of their website and ideally also their submission process accordingly. While this may be a single step for a journal, keeping the URLs up to date is a recurring task.

Where are “URLs linking to related research objects” required?

EQSIP V.1.0 requires linking to related research objects.

Challenges and gaps in relation with “URLs linking to related research objects”

In the OADJ Study¹³³ from 2021, there was a question formulated as “Does the journal require linking to data, code, and other research outputs that underlie the publication and are available in external repositories?”. **Table 21** presents how the 1,619 journals responding to the survey provided their answers. 25% of journals answered “Yes” while 49% rescinded with “No”, suggesting that this was at least at the time of the survey not a mainstream practice among OADJs.

¹³³ Bosman et al., 2021.

Yes	402	25%
No	791	49%
Unknown	277	17%
No answer	149	9%

Table 21: Does the journal require linking to data, code, and other research outputs that underlie the publication and are available in external repositories?

Website feature requirement	OJS	Janeway	Lodel
URLs linking to related research objects	Research objects can be linked using the concept of external Galley in OJS. However, no direct support for linking research objects.	Research objects can be uploaded and hosted directly in Janeway (with component DOI) or linked out using a typeset file/galley.	Not supported (possible to manually add links in notes)

Technical support for “URLs linking to related research objects” among software

URLs linking to related research objects are partly supported under OJS through an external Galley. Janeway supports the uploading and hosting of related research objects and the linking to outside resources. Lodel does not support it but links can manually be added in the notes.

Conclusions “URLs linking to related research objects”

In the spirit of open science, materials associated with publications should be interlinked and discoverable in order to enable transparency and to facilitate further use of the data. Linking to such objects can happen in many different ways but is best handled as part of structured metadata collected during the manuscript submission process. From the data that is available about OADJs practices, only around 25% of the journals require linking to related research objects, showing substantial room for growth since the technical requirements for this are not that advanced, as the links can be stored and displayed in many different ways.

4.4.5 COUNTER

Description of the standard “COUNTER”

Established in 2002, Counting Online Usage of Networked Electronic Resources (COUNTER) offers a Code of Practice for publishers, vendors, and libraries that enables the comparison of usage data on electronic resources. The standard allows for a consistent view of usage statistics in terms of value, status and impact.¹³⁴ The COUNTER Code of Practice has been continuously developed. According to the COUNTER Code of Practice V5.0.2, the aim is “to balance changing reporting needs with the need to make things simpler so that all content providers can achieve compliance and librarians can have usage statistics that are credible, consistent and comparable”¹³⁵. The COUNTER reports follow a specified formatting structure and a predefined set of element names, and have to be submitted in tabular as well as JSON¹³⁶ format. The COUNTER Project makes supporting mechanisms available, like a validation tool, documentations and audit support.

Specifically with regards to OA publishing, PKP reports that COUNTER “provides rules on what should be counted as a view, including specific rules for robot usage and multi-click abuse. OJS filters metrics through these rules. It should also be noted that PKP is part of the COUNTER Bots and Crawlers Working Group.”¹³⁷ This working group was formed in response to challenges measuring open digital resources’ usage data.¹³⁸ Accessing COUNTER statistics gives journals insights into the usage statistics and thereby also a better grasp of their readership.

Where is “COUNTER” required?

In accordance with EQSIP V1.0, a publishing platform should adhere to current interoperability standards, such as COUNTER.

Challenges and gaps in relation with “COUNTER”

In the data sources analysed in this report, no empirical evidence could be gathered regarding the use of COUNTER among OA journals.

Technical support for “COUNTER” among software

¹³⁴ Shepherd, P. T. (2010). *COUNTER: current developments and future plans. The E-Resources Management Handbook*. UKSG. <https://www.uksg.org/sites/default/files/5-Shepherd-W1G177781L48WR13.pdf>

¹³⁵ <https://cop5.projectcounter.org/en/5.0.2/00-foreword.html>

¹³⁶ The acronym JSON stands for JavaScript Object Notation.

¹³⁷ <https://docs.pkp.sfu.ca/learning-ojs/en/statistics>

¹³⁸ Greene, J. W. (2017). Developing COUNTER standards to measure the use of Open Access resources. *Qualitative and Quantitative Methods in Libraries*, 6(2). <https://www.qqml-journal.net/index.php/qqml/article/view/410/404>

Website feature requirement	OJS	Janeway	Lodel
COUNTER	Supported in OJS core	Janeway has a stricter counting method than COUNTER, it does not have COUNTER reports.	Supported (COUNTER 5)

COUNTER is supported in the OJS core and Lodel. Janeway has a different method for counting usage statistics and does not provide COUNTER reports.

Conclusions “COUNTER”

This standard proved to be challenging to collect evidence as neither DOAJ nor any of the past surveys have included data points on this. Fortunately, the scan of software features provided some assurance that both OJS and Lodel are compliant with COUNTER and Janeway offers similar stricter reporting that can be used for analytics.

5 CASE STUDIES

This section includes case studies of four journal platforms that each host multiple journals, providing another perspective on the gap analysis rather than how it is perceived from the individual journal's point of view.

5.1 Journal.fi

5.1.1 Introduction

Journal.fi¹³⁹ is an OA publishing service provided by the Federation of Finnish Learned Societies (TSV). Launched in January 2017, the Journal.fi site currently features 140 Finnish scholarly journals. The service is using the OJS 3.2 software. TSV also operates a similar service, Edition.fi¹⁴⁰, for the publication of OA books launched in 2020. TSV also distributes subsidies granted by the Ministry of Education and Culture to learned societies for their publishing and international activities¹⁴¹.

5.1.2 Background

Development of the platform took place between 2015-2016 via the KOTILAVA project¹⁴², a joint effort by TSV and the National Library of Finland to support Finnish scholarly journals in their transition to immediate OA. The KOTILAVA project was part of the Open Science and Research Initiative¹⁴³ (2014-2017), a project of the Ministry of Education and Culture for promoting Finnish open science in extensive cooperation between ministries, universities, research institutes and research funders.

The two main goals of the KOTILAVA project, as outlined in a 2014 report Finnish Scientific Journals and Open Publishing: A Study of Possible Funding Models¹⁴⁴ (in Finnish), were 1) to develop an OJS platform for editing and publishing OA journals, and 2) to create a new consortium-based funding model for Finnish OA journals. While the project resulted in the launch of the Journal.fi platform, a sustainable collective OA funding model has not been established despite continuous effort.

¹³⁹ <https://journal.fi/>

¹⁴⁰ <https://edition.fi/>

¹⁴¹ <https://www.tsv.fi/en/grants/publishing-and-international-activities-scientific-societies>

¹⁴² <https://web.archive.org/web/20230327004401/https://kotilava.fi/19-elokuu-2016-1247/kotilava-%E2%80%93-finnish-academic-journals-towards-immediate-open-access>

¹⁴³ <https://avointiede.fi/en/policies/policies-open-science-and-research-finland/open-science-and-research-initiative>

¹⁴⁴ <https://edition.fi/tsv/catalog/book/157>

- In 2019, a National policy and executive plan on OA to scholarly publications¹⁴⁵ was published by National Open Science Coordination (AVOTT). One objective of the policy was that the research community creates a jointly funded publishing model that enables immediate OA to research articles published in Finland.
- In 2022, an AVOTT working-group published Proposals for New Supplementary Funding Models for Domestic Scientific Periodicals to Enable Immediate Openness: Final Report of the Working Group Appointed by the National Steering Group for Open Science and Research¹⁴⁶ (in Finnish).
- In 2023, TSV working-group published a Proposal for the Funding of Open Domestic Scientific Journals¹⁴⁷ (in Finnish).

According to Late et al.¹⁴⁸, learned societies publish around 70% of 402 peer-reviewed publication channels in Finland, including serials and book publishers. Commercial publishers produce only 2.6% of Finnish journals and books. A thorough study of the peer-reviewed journal landscape in Finland is provided by Linna et al.¹⁴⁹. The study found that of the 336 peer-reviewed journals identified, 53% were publishing immediate OA, with a further 6% as delayed OA, and 2% of journals offering a hybrid OA option. Diamond OA is the dominant model of OA publishing among journals, with the study only identifying seven journals in the country that ask for an APC. Looking at what is visible through the lens of indexing in DOAJ, Finland has 66 journals in DOAJ of which almost all (59) are diamond journals. Of 66 DOAJ indexed journals, 55% are published on Journal.fi platform.

5.1.3 Journal.fi service

The Journal.fi service is free for learned societies who are TSV members¹⁵⁰ while other publishers are charged a nominal fee. For all publishers using the platform, TSV provides guidance related to online journal publishing, offering short training and email support for new journals.

Certain criteria are applied for accepting journals to the platform, as outlined on the TSV's Journal.fi service website¹⁵¹ (in Finnish). Separate criteria apply to TSV members and other publishers:

¹⁴⁵ Open Science Coordination in Finland, Federation of Finnish Learned Societies (2019). Open access to scholarly publications. National Policy and executive plan by the research community in Finland for 2020-2025 (1). *Responsible Research Series*. <https://doi.org/10.23847/isbn.9789525995343>

¹⁴⁶ https://avointiede.fi/sites/default/files/2022-10/Loppuraportti-rahoitusmalliehdotukset-1022_0.pdf

¹⁴⁷ https://www.tsv.fi/sites/tsv.fi/files/media/ehdotus_avointen_kotimaisten_tiedelehkien_rahoytukseksi.pdf

¹⁴⁸ Late, E., Korkeamäki, L., Pölonen, J., & Syrjämäki, S. (2020). The role of learned societies in national scholarly publishing. *Learned Publishing*, 33, 5-13. <https://doi.org/10.1002/leap.1270>

¹⁴⁹ Linna, A.-K., Holopainen, M., Ikonen, A., & Ylönen, I. (2020). Kotimaiset tieteelliset julkaisusarjat ja avoimuus. *Informaatiotutkimus*, 39(4), 4-32. <https://doi.org/10.23978/inf.98656>

¹⁵⁰ https://www.tsv.fi/en/toiminta_en/membersocieties

¹⁵¹ <https://tsv.fi/palvelut/avoimen-julkaisemisen-palvelut/journalfi>

- TSV member society journals:
 - Articles must be openly available with a maximum delay of one year from the date of publication.
 - Journals based on subscription models from member societies can also use the service for manuscript reception and editorial work.
 - Journal is not required to be peer-reviewed or have a JUFO classification¹⁵².
- Other than TSV member society journals:
 - Journal must be immediately and completely open. It must be published regularly, at least once a year.
 - Journal must present scientific research results. This can be demonstrated if the publication series is classified at JUFO levels 1, 2, or 3¹⁵³, or if it uses TSV Label for peer-reviewed scholarly publications¹⁵⁴.
 - Publisher of the journal/yearbook must be Finnish, or it must be a joint publication with at least one Finnish entity as a publisher.

Instructions for using the Journal.fi and Edition.fi service are provided on the Instructions for TSV's Open Publishing Services¹⁵⁵ (in Finnish). On this page, guidance is provided on the following topics:

- Instructions Applicable to Both Services
 - Joining the Service
 - Making Support Requests
 - Customizing the Appearance and Navigation of Your Site
 - Using Creative Commons licences
 - Self-Archiving: Listing in the SHERPA/RoMEO Service
- Journal.fi Service
 - Learning OJS Guide
 - Article Visibility and Journal.fi
 - DOI Identifiers
 - ORCID Identifiers
 - Joining the DOAJ Service and Submitting Article Metadata
 - Peer Review Badge
 - Pre-Publishing Articles (Forthcoming plug-in)
 - Collecting Funder Information
 - Statistics
 - Rapid Publication plug-in
 - Importing Old Issues
 - Delayed (Embargo) Publishing and Paid plug-ins
- Edition.fi Service
 - Learning OMP Guide
 - Publishing and Managing Books

¹⁵² <https://www.julkaisuforum.fi/en>

¹⁵³ <https://julkaisuforum.fi/en/evaluations>

¹⁵⁴ <https://tsv.fi/en/services/label-for-peer-reviewed-scholarly-publications>

¹⁵⁵ <https://docs.google.com/document/d/e/2PACX-1vR6Xf8ugrjnIjmf79cW-lgeu-bYnGpGgiAGlAlxb7su6lslpC00FO8wHd-5pOI0cyicKeuG7fsRcJ8g/pub>

Journal.fi platform, as well as Edition.fi, already offer interfaces to a wide range of international services, such as Crossref, DOAJ, ORCID, OpenAIRE and Google Scholar (see **Figure 2**). These platforms also facilitate integration of publication information to national services, and are able to also support VIRTAs publication information service and National Research Information Hub. Regarding DOIs, journals using the Journal.fi service for publishing can apply to join the agreement between TSV and Crossref, through which the journal can obtain DOI identifiers for its articles. This service is currently free for TSV member societies.

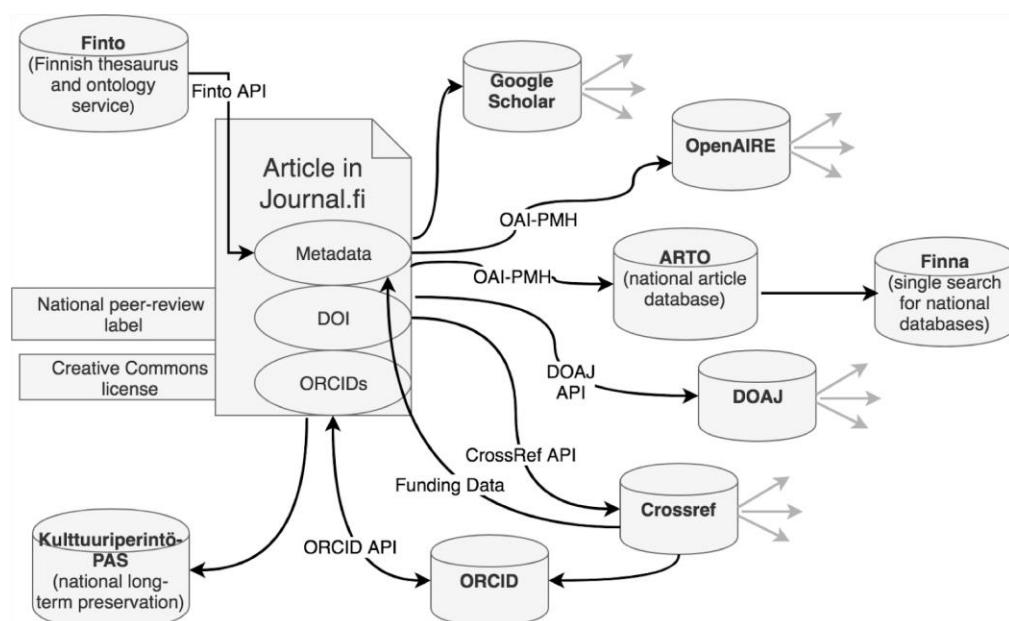


Figure 2: Current Journal.fi data integration

The contents, OA and linguistic diversity of the platform have been briefly summarised in a 2021 study “Who are the users of national open access journals?”. The case of the Finnish Journal.fi platform¹⁵⁶ published in *Learned Publishing*: “In 2020, Journal.fi platform hosted 98 journals, of which 85% provide immediate open access and 15% have an embargo period. These journals publish in a variety of languages, however the national languages — Finnish and Swedish — and English are most common. The journals represent all scientific fields, however, we estimate that the vast majority (around 85%) specialise in the SSH. Almost all journals on the Journal.fi platform are peer-reviewed. As of 17 November 2020, Journal.fi platform had a total of 47,970 articles published between 1883 and 2020 (also including material published before the establishment of the platform in 2017). Roughly 77% of all articles are in Finnish, 19% in English, 3% in Swedish and only 1% are in other languages.”¹⁵⁷

¹⁵⁶ Pölönnen, J., Syrjämäki, S., Nygård, A.-J., & Hammarfelt, B. (2021). Who are the users of national open access journals? The case of the Finnish Journal.fi platform. *Learned Publishing*, 34(4), 585–592.

<https://doi.org/10.1002/leap.1405>

¹⁵⁷ Pölönnen et al., 2021.

5.1.4 Compliance with technical standards and their adoption among journals

This section goes into closer detail regarding the technical features implemented into the software environment of Journal.fi. As previously mentioned, the service is running OJS where each journal can, to a certain degree, customise their instances of the software environment and select which standards and practices they want to adhere to. The review is structured similarly to how the standards were presented earlier in this report in section 3, by being grouped into the same four main categories. A full table containing the support and adoption information is provided in **Table 22**.

Identifiers

All journals on the platform have an ISSN and have registered it into OJS where it is a required information field regarding the publication. Some other identifiers are not as comprehensively adopted even though there is software support for them:

- Only 39 out of 139 journals are using the ORCID plug-in, and while it is possible that some journals are manually entering ORCID IDs into the metadata, this is not a recommended practice since that can lead to incorrect ORCIDs being entered. The plug-in automates and validates the data, making the preferred alternative;
- 103 out of the 139 journals are using DOIs;
- 9 out of 139 journals are using the OJS funding plug-in for registration of funder DOIs/PIDs.

These are the ones where there would be room for growth in terms of identifier adoption without any changes to the technical environment. A plug-in is available for registration of RORs and ARKs in the metadata but so far they have not been put into use. Handle identifiers are not supported in OJS and are not in use in Journal.fi.

Metadata

With regards to the metadata standards, essentially everything is technically supported on Journal.fi except KBART. Most of the, like the metadata exchange for harvesting, having the OpenAIRE plug-in enabled in OJS, enabling the mass data export plug-in in OJS, including metadata about OA status and placing metadata under CC0, are also comprehensively adopted by journals since they are back-end features that do not require manual intervention from the journal editors to keep running. Regarding metadata that should be registered outside of the actual journal platform, only 63 out of the 139 journals have registered their self-archiving policies into SHERPA/RoMEO, this despite organising awareness campaigns and workshops on this specific topic. Direct deposition into OA repositories is something that is currently not implemented for any journal in the portal, and unfortunately, there are no available statistics on the comprehensiveness of citation data registration into Crossref.

Content

For the standards included in the content category, all of them are supported and available to journals by the OJS software environment on Journal.fi. There is some variation regarding the adoption among journals. Only a couple of journals are providing their content in JATS XML, likely due to the technical expertise and extra steps needed in processing to make it happen for all content. Regarding preservation, a choice of service provider is still being deliberated, comparing national and international providers, but this will likely be resolved soon on the portal-level. The copyright and licensing information would need to be looked into closer, but on an aggregate level one can see that about half of the journals have defined a default copyright holder and default licence in the system, suggesting that there might be some room for improvement in that regard.

Website features

The standards included under website features are softer around the edges in terms of specificity, but the OJS implementation used on Journal.fi does have support for all of them at least on a basic level outside of post-publication commenting/evaluation processes for which there is currently no available plug-in. In terms of those where adoption can be assessed quantitatively, a plug-in for adding social sharing features has been installed in 88 out of the 139 journals, and some elements of open peer review have been adopted by a couple of individual journals. Research objects are mostly linked to from the full-text documents rather than having them as separate structured metadata.

5.1.5 Conclusion

Journal.fi provides comprehensive readiness for journals taking the most central identifiers relevant for journals into use - the challenge seems mainly to create awareness and motivation for journals to expand their current practices. For metadata, many of the central standards are comprehensively available and taken into use by all journals since they are turned on by default and require minimal interaction with the editorial staff to be kept running. However, the degree to which journals on the portal have registered their self-archiving policies into SHERPA/RoMEO could be improved as under half of all journals have submitted this data to the service. When looking at the content standards, the very low adoption of JATS XML is not positive for the machine-readability and format-independence of content but it is at the same time understandable that journals operating with modest financial means might not have access or the resources to hire the technical expertise needed to make this happen on an ongoing basis. Preservation aspects for content are still not completely resolved at the time of writing, the technical readiness is there but the choice of service provider is still at the evaluation stage. When it comes to website features, it was

harder to make quantitative judgement on adoption regarding all aspects, but the technical readiness is there to cater to different types of needs regarding this.

Overall, one could conclude that Journal.fi is technically very mature and aligned with the many of the demands stated by the standards included in this analysis. The major challenge is obtaining higher adoption for the standards that are not purely back-end features that can be set to run without intervention in the background, and towards this goal different types of training materials would likely be helpful.

Identifiers	Supported?	Adopted?
ISSN/eISSN identifier	Yes	A required field in OJS
ORCID identifier	Yes	32/139 journals are using the ORCID plug-in. OJS also supports a metadata field for ORCID identifiers by default and some journals might fill in ORCID IDs manually (not recommended).
ROR identifier	No, we are not using the existing plug-in currently	No
ARK identifier	Supported with a plug-in, https://github.com/yasielpv/pkp-ark-pubid	No
Handle identifier	No	No
DOI identifier	Yes	103/139 journals are using DOIs
Funder DOIs/PIDs	Yes	9/139 journals are using OJS Funding plug-in
Support of PIDs for authors/funders	Yes (ORCID and Funder DOI)	Yes, ORCID for authors
Metadata	Supported?	Adopted?
Metadata exchange for harvesting (e.g. DC, OpenAIRE, OAI-PMHetc.)	Yes	OAI-PMH is active by default. Journal.fi journals are required to have CC0 metadata
OpenAIRE Guidelines	Yes	OpenAIRE plug-in is turned on site wide
Mass metadata export (as CSV files, ONIX XML feeds or in any	Yes	Import and export plug-ins are turned on by default. Mass imports have been used to import old content.

other established format)		
KBART	No	-
Metadata about OA status	Yes	For example, the OpenAIRE OAI-PMH feed includes the OA status of each individual article.
Registering of self-archiving policy	Not something that requires platform integration	Yes, 63/139 of Journal.fi journals are registered in SHERPA/RoMEO
Direct deposition in OA repository	There is an actively maintained SWORD plug-in https://github.com/pkp/sword which is a protocol that for example DSpace uses.	No
Metadata under CC0	All metadata can be made available for harvesting via OAI-PMH. Metadata licence is the journal's policy decision, and journal.fi requires that all journals have a CC0 metadata policy	Not something that requires platform compatibility, but journal.fi requires that all journals have a CC0 metadata policy
Open Citations standards compliance	A plug-in allows to transfer reference lists to Crossref. However, this citation data is unstructured unless the reference list is provided in structured JATS XML	Unknown
Content	Supported?	Adopted?
Human- and machine-readable information about the OA status, copyright holder and licensing are provided in each publication in a standard in a non-proprietary format	Yes	OA status is article based and can be determined for all content. 75/139 journals have defined a default copyright holder 68/139 journals have defined a default licence
Text and Data Mining is technically supported	Yes	Available via OAI-PMH

Full text in machine-readable format (JATS XML or equivalent (e.g. TEI))	Yes	OJS supports any publishing format. Publishers decide the format they want to use. JATS XML is used by very few - probably 2/139 - because most journals lack the skills and resources to produce JATS XML.
Deposited in a digital preservation service	Yes	OJS supports digital preservation but we are still in the process of deciding which service we can use. The digital preservation service provided by the national library is the one we aim to use, but there have been some delays with it. We are also looking into Portico and PKP PN, but the General Data Protection Regulation (GDPR) implications of these services are somewhat unclear. The servers are not situated in the EU.
Website features	Supported?	Adopted?
SEO	A very wide concept, but in general OJS and the default OJS theme is SEO friendly	100/140 journals use the journal.fi theme which is based on the OJS default theme. The remaining journals use the OJS default theme.
Alerting services, sharing to social networks,	Yes	AddThis-plug-in in 88/139 journals
Post-publication evaluation and commenting,	No	
Support for multimedia,	Yes	
Open peer review (where relevant)	Yes	Some elements of open peer review adopted by a couple of journals.

Unique URL for landing pages	Yes	Yes
URLs linking to related research objects	Yes	Research objects usually linked in the full text.
COUNTER	Yes	Yes

Table 22: Support and adoption information for journal.fi

5.2 HRČAK – Portal of Croatian scientific and professional journals

5.2.1 Introduction

HRČAK¹⁵⁸ is the central Portal of Croatian scientific and professional journals that serves as a publishing platform for 530+ OA journals and provides access to 287,000+ full text articles with an average 70,000+ daily visitors. Since its launch in 2006, HRČAK has become the prominent OA platform in Croatia, publishing and supporting OA and good publishing practices (e.g. usage of ORCID identifiers, publishing associated datasets and linking papers to them, open and machine-readable format JATS XML). The Portal was built and has been maintained by SRCE – University of Zagreb University Computing Centre in collaboration with the experts from the field of information and library science in Croatia, and the representatives from journal editorial boards.

5.2.2 Scope and mode of operation

HRČAK is freely accessible for both publishers and readers, allowing unrestricted use of the journals' content without any associated costs. Journals in HRČAK need to have connections to Croatia, but these journals accept manuscripts from global contributors, enabling diverse contributions from scholars worldwide.

To be accepted into HRČAK, journals must not only have connections to Croatia but also fulfil specific criteria. These include having a defined review policy, the use of open licences and providing transparent ethics statements. The criteria have been established by the HRČAK advisory board, a body composed of representatives from journal editorial boards in HRČAK, the National and University Library in Zagreb, CROASC (Croatian Association for Scholarly Communication) and the Ministry of Science and Education. After acceptance into HRČAK, the journal's editors are accountable for the content's accuracy, while the HRČAK team at SRCE oversees the functionality, security and design of the Portal.

¹⁵⁸ <https://hrcak.srce.hr/en>

In addition to the Portal, HRČAK operates its own instance of OJS¹⁵⁹, allowing editors to manage the editorial process through this platform. Over 50 journals are actively using HRČAK's instance of OJS. The synchronisation of OJS with HRČAK Portal facilitates easy journal publishing across both platforms. Furthermore, HRČAK offers support to publishers for the publication of conference proceedings through the OMP¹⁶⁰, which was released in September 2022.

5.2.3 Compliance with technical standards and their adoption among journals

HRČAK was developed and is sustainably evolving in accordance with needs and trends of scientific publishing. From the beginning, a strong emphasis has been placed on the Portal's technical features, enabling interoperability, harvesting and distribution of journals on HRČAK. The technical standards met by HRČAK will be presented through several categories recognized in this document, focusing specifically on the HRČAK Portal developed within SRCE. It won't cover the OJS instance of HRČAK, since OJS has already been addressed in other sections of the document. The emphasis will be on the degree of adoption among journals, aiming to analyse the underlying reasons for such levels of adoption.

Identifiers

HRČAK mandates and encourages the use of PIDs, where ISSN is required, while the use of DOI and ORCID identifiers is strongly encouraged. All journals in HRČAK have an assigned ISSN, although not all journals have DOI identifiers. For measurement purposes, it's important to note that HRČAK distinguishes between DOIs for journal, issue and article. During the period of November 2022 to November 2023, 280 journals have entered DOI numbers for at least one article. While lacking supporting data for the hypothesis, it can be speculated that the low number of DOIs is potentially influenced by the limited financial resources and lack of understanding regarding the importance of DOIs. Additionally, the non-mandatory nature of the DOI field in HRČAK might lead some journals to overlook entering this data.

In terms of authors' IDs, HRČAK has implemented a procedure to link articles to authors' ORCID profiles, mandating verification and authentication by the respective authors, resulting in 76,000+ (out of possible 562,000) links between articles and ORCID profiles. The relatively low number of linked ORCIDs (around 14%) should not be attributed to a technical gap. On one hand, the challenge arises from the presence of articles on HRČAK that were originally published in the 19th or 20th century, making it impossible to link them to ORCID. On the other hand, the linking process between articles and ORCID requires action from both editors and authors, adding email addresses for all authors on each article and the authorization of

¹⁵⁹ <https://hrcak.srce.hr/ojs/>

¹⁶⁰ <https://pkp.sfu.ca/software/omp/>

ORCID by authors themselves. So, the gap is primarily connected to organisational and editorial factors, along with a lack of awareness among editors and authors regarding the importance of PIDs.

Metadata

Each journal on HRČAK is described by a set of specified metadata. When publishing articles, editors are required to input metadata for each article, with the option to include multilingual metadata. The metadata is structured according to the MODS and DC schemas and accessible for harvesting through the OAI-PMH interface¹⁶¹. The OAI-PMH interface was implemented in 2006 and is fully compliant with the OpenAIRE Guidelines for Literature Repository Managers v3.0¹⁶². As a result, HRČAK is registered as a data source on the OpenAIRE portal. Furthermore, HRČAK exposes metadata in Highwire Press meta-tags required for indexing in Google Scholar and it also provides the KBART-compliant list of journals, available in two formats: tab-separated¹⁶³ and HTML¹⁶⁴. There is a specified metadata policy¹⁶⁵ stating that the metadata published on HRČAK can be reused in any medium without prior permission. It can be affirmed that the adoption of technical standards regarding openness and the ability to disseminate metadata is comprehensive because there is no policy allowing journals to restrict access to their metadata.

Content

In addition to the metadata policy, HRČAK has a defined data policy indicating that papers published on HRČAK can be stored, distributed and used in other ways in accordance with the usage rights and the licences granted by the editorial of the journal. The usage rights and licences are displayed in a human-readable format on the journals' web pages within HRČAK and are accessible in a machine-readable format through the OAI-PMH interface. These rights and licences can be customised at the level of the journal, issue and article. That information is mandatory on HRČAK, ensuring that all journals have provided it.

Furthermore, besides the obligatory publishing of full-text in PDF format, since 2017, journals in HRČAK have the option to publish the machine-readable full-text JATS XML of articles. This option allows journals to become accessible for text, data and reference mining purposes. Despite the advantages of JATS XML, publishing it on HRČAK faced a challenge as editors were forced to create XML files themselves. That required use of expensive tools or outsourcing, leaving editors to pay based on their publication frequency. An alternative to that was manual

¹⁶¹ <https://hrcak.srce.hr/oai/?verb=Identify>

¹⁶² OpenAIRE (2012). *OpenAIRE Guidelines for Literature Repository Managers 3.0*. <https://cordis.europa.eu/docs/projects/cnect/5/283595/080/deliverables/001-D44OpenAIREGuidelinesv3DELIVERABLE.pdf>

¹⁶³ <https://hrcak.srce.hr/kbart>

¹⁶⁴ <https://hrcak.srce.hr/kbart-html>

¹⁶⁵ <https://hrcak.srce.hr/en/politike>

article rewriting that proved to be a time-consuming task with need for specialised personnel. Considering these circumstances, the adoption of this feature in HRČAK has been relatively low, with less than 10 journals using it. In practice, only journals regularly using this format in HRČAK were those already preparing XML files for PubMed Central, which mandates the use of JATS. In September 2023, HRČAK introduced a new feature enabling automated creation of full-text JATS XML from Word (DOCX) documents without prior knowledge of the standard. Within the initial two months following its release, 20 journals out of 537 have adopted the feature, resulting in the publication of 220+ articles in JATS XML format. Although the feature is currently exclusive to journals in HRČAK, it will be released as a standalone tool on the EOSC Marketplace in 2024, enabling other European journals to use it.

Website features

Every article on HRČAK has a unique HRČAK ID and a landing page presenting metadata about the article, the full-text PDF document, and, for those articles with created JATS XML, the full-text in HTML format. The article's metadata such as information about the authors and their affiliations, article's title, journal's title, DOI, publication date, keywords et al. are included in Highwire Press meta-tags for a better visibility through Google Scholar.

End users have the option to register on HRČAK to subscribe to journals. The subscription enables them to receive newsletters whenever the subscribed journal releases a new issue.

Articles on HRČAK can be linked to the original dataset and can be accompanied by supplementary files.

5.2.4 Conclusion

Implementing and maintaining HRČAK according to the most important technical standards enhances visibility and accessibility of the published journals, amplifying their reach and impact. For instance, HRČAK has been a registered data provider in OpenAIRE since 2015. Also, collaborative efforts between SRCE, the National and University Library in Zagreb, and CROASC have resulted in substantial progress regarding automated crawling to Web of Science (WoS) and Scopus. Having said that, some of the implemented features still have limited use, such as usage of available PIDs (DOI and ORCID), which can be attributed to editors' lack of education on their significance and optimal usage. Significant strides have been made within HRČAK to advance machine-readability as well. That includes enabling the publication of JATS XML format and holding workshops to educate and inform editors about the benefits and options available for creating content in JATS. In 2023, HRČAK took a significant leap forward by implementing the automated creation of full-text JATS XML, which will be available through EOSC Marketplace in 2024. This feature underscores HRČAKs

commitment to advance technological capabilities and facilitate access to machine-readable content for publishers across Europe.

Despite the successful compliance with most mandatory technical standards, as depicted in the table below, addressing the educational needs of editors remains an ongoing challenge for HRČAK. By actively investing in editors' education, HRČAK fosters not only a high-quality platform, but the scholarly discourse within the platform as well.

Identifiers	Supported?	Adopted?
ISSN/eISSN identifier	Yes	All journals in HRČAK have ISSN/eISSN
ORCID identifier	Yes	76,000+ links between articles and ORCID profiles, verified and authenticated by their respective authors
ROR identifier	No	-
ARK identifier	No	-
Handle identifier	No	-
DOI identifier	Yes	260/537 journal have DOI assigned on the level of journal
Funder DOIs/PIDs	No	-
Support of PIDs for authors/funders	Yes (ORCID IDs)	Authenticated collection of ORCIDs and linking papers to ORCID profiles
Metadata	Supported?	Adopted?
Metadata exchange for harvesting (e.g. Dublin Core, OpenAIRE, OAI-PMH s etc.)	Yes	All metadata is accessible via OAI-PMH
OpenAIRE Guidelines	Yes	Compatibility with OpenAIRE Guidelines for Literature Repositories v3
Highwire Press	Yes	Metadata is exposed through Highwire Press meta-tags

Mass metadata export (as CSV files, ONIX XML feeds or in any other established format)	No	-
KBART	Yes	Journals are available in KBART compliant list: tab-separated and HTML
Metadata about OA status	Yes	HRČAK requires to publish in OA
Registering of self-archiving policy	Feature does not require platform integration	Not required by HRČAK
Direct deposition in OA repository	No	-
Metadata under CCO	Yes	All metadata is freely accessible via OAI-PMH and may be re-used in any medium without prior permission
Open Citations standards compliance	No	-
Content	Supported?	Adopted?
Human- and machine-readable information about the OA status, copyright holder and licensing is provided in each publication in a standard in a non-proprietary format	Yes	All journals display their terms of use on their page, also accessible via the OAI-PMH interface
Text and Data Mining is technically supported	Yes	20+ journals that publish JATS XML facilitate Text and Data mining
Full text in machine-readable format (JATS XML or equivalent (e.g. TEI))	Yes	20+ journals publish in JATS XML
Deposited in a digital preservation service	Yes	SRCE manages the in-house long-term preservation of journals within HRČAK. The

		National and University Library in Zagreb collects the legal deposit of digital publications.
Website features	Supported?	Adopted?
SEO	Yes	All journals undergo the HRČAK SEO optimization
Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review (where relevant)	Only alerting services are supported	12,573 subscribers
Unique URL for landing pages	Yes	All journals, issues and articles have unique landing page
URLs linking to related research objects	Yes	The functionality is still not used as recommended.
COUNTER	No	-

Table 23: Support and adoption information for HRČAK

5.3 ZRC SAZU journals

5.3.1 Introduction

Založba ZRC¹⁶⁶ (ZRC publishing house), as the publishing unit of the Scientific Research Centre of the Slovenian Academy of Sciences and Arts¹⁶⁷ (abbreviated in Slovenian as ZRC SAZU), a public and non-profit institution, is the publishing house of 18 institutes of ZRC SAZU and one of the three largest Slovenian scientific publishers. The mission of Založba ZRC includes three closely related tasks: the production and publication of academic and scholarly literature, the sale of books and journals and the promotion of the institute's scientific achievements. It specialises in academic and expert literature (monographs and journals), mainly in the Humanities, but its publishing programme also includes some social and natural sciences, as well as publications in the field of social medicine and related sciences.

¹⁶⁶ <https://zalozba.zrc-sazu.si/en/predstavitev>

¹⁶⁷ <https://www.zrc-sazu.si/en>

Slovenia started the implementation of open science principles in September 2015 with the National Strategy of OA to Scientific Publications and Research Data in Slovenia 2015–2020¹⁶⁸ (in Slovenian). The strategy stipulates that, between 2015 and 2020, each beneficiary should provide OA to all peer-reviewed scientific publications relating to the results of nationally funded research. While the strategy has not been implemented, it has encouraged OA publishing. The Slovenian Research and Innovation Agency (ARIS)¹⁶⁹, which has the largest share of co-financing of the scientific press, implemented this condition in their tenders for Slovenian academic journals in 2018. Principles of Plan S have been incorporated into national legislation in 2022¹⁷⁰: from 2023, all research results more than 50% funded by the state must be OA. Additionally, as a member of the EU, Slovenia is bound by the legislation and policies of the ERA.¹⁷¹

In 2015, ZRC SAZU launched the establishment of electronic publishing, the transition of journals to the OJS and the OA publishing policy as part of Digital Research Infrastructure for the Arts and Humanities (DARIAH)¹⁷², the Digital European Research Infrastructure for the Arts and Humanities, in which ZRC SAZU has been involved since 2011. The project was related to the above mentioned national strategy. In 2015, a common OJS for all the journals of the ZRC SAZU institutes (which had previously operated completely independently, including online editions, with the Založba ZRC only providing support for the printing of the journals), was established at the Založba ZRC and has since been available on the ZRC SAZU OJS platform. Initially, 6 journals joined the platform, but today all 13 journals published by the ZRC SAZU institutes are included on the OJS platform, with complete archives. ZRC SAZU also maintains two other OJS installations. The first¹⁷³, hosts two journals of the Slovenian Academy of Sciences and Arts, while the second¹⁷⁴, currently hosts three scientific journals published by learned societies.

When the OJS platform for ZRC SAZU journals was established, the first phase focused on the content standards of open science publishing (e.g. defining editorial and peer review policies, ethical standards, copyright, licensing) and the basic legalities set by OJS (each article has its own landing page with metadata about the article and DOI identifier). During these phases, the editorial teams of the journals worked with the editor-in-chief of Založba ZRC, and the technical details required the help of someone who was familiar with OJS and could provide

¹⁶⁸ <https://www.arrs.si/sl/dostop/strategija.asp>

¹⁶⁹ <http://www.arrs.si/en/>

¹⁷⁰ <http://dirrosdata.ckk.uni-lj.si/en/nacrt-s/>

¹⁷¹ <https://dirrosdata.ckk.uni-lj.si/en/evropska-unija/> as well as Pogačnik, A. (2022). Slovenske znanstvene revije v letu 2021, *Organizacija znanja*, 27(1-2). <https://doi.org/10.3359/oz2227002>

¹⁷² <http://www.dariah.si/dariah-si-eng/>

¹⁷³ <https://ojs.sazu.si/>

¹⁷⁴ <https://ojs-gr.zrc-sazu.si/>

advice and technical support for the transition to the platform. External technical support was obtained.

The editor-in-chief and managing director of Založba ZRC, and since 2023, also the assistant editor-in-chief for journals, are responsible for publishing policy. The work of the publishing house is supervised by a publishing council of 12 researchers (elected for a period of 4 years). The publishing house also actively participates in two ZRC SAZU commissions: the Open Science Commission and the Ethics and Integrity Commission. The preparation of the journal itself and the peer review process are entirely the responsibility of the individual institutes and their researchers (the journals do not have a professional editorial board). Založba ZRC has a total of 10 FTE. It also has a sales department with a bookshop (four employees), a promotion/events department (3 employees), and a designer who also liaises with printers. The typesetting and formatting of the journals is largely done by external staff invited by the individual editorial boards.

In accordance with OA publishing policy, journals are published on an OJS (version 3.3.0.10) platform. Since 2019, the Založba ZRC has been publishing its monographs in OA through an OMP platform¹⁷⁵; currently, over 800 monographs (more than half of the total 1993-2024 production) have been published in the OMP. General technical support for OJS and OMP is provided by two external technical assistants. Training on how to use OJS has been provided to all journal editors and their external technical assistants.

In the journals published by the Založba ZRC, all articles are made fully OA immediately after publication, authors retain all rights, articles published in the journals are currently available under licences CC BY (five journals), Creative Commons Attribution Share-Alike (CC BY-SA) (four journals) and Creative Commons Attribution NonCommercial NoDerivates (CC BY-NC-ND) (three journals). Authors may also publish their work online (e.g. in institutional repositories or on their own websites). The journals publish articles in Diamond OA: there is no charge to authors or readers. All journal content is stored in the repository of the Digital Library of Slovenia¹⁷⁶. All 13 ZRC SAZU journals are indexed in Scopus, eight of them also in DOAJ and five of them also in Web of Science (two in Science Citation Index-Expanded (SCIE), two in Arts & Humanities Citation Index (AHCI), one in Emerging Sources Citation Index (ESCI)).

5.3.2 Compliance with technical standards and their adoption

The version of OJS (3.3.0.10) used by the Založba ZRC technically supports almost all the standards discussed in D3.2. **Identifiers** such as ISSN, ORCID, ROR, DOI, are implemented.

¹⁷⁵ <https://omp.zrc-sazu.si/zalozba>

¹⁷⁶

<https://www.dlib.si/results/?euapi=1&query=%27keywords%3dzrc+sazu%27&sortDir=ASC&sort=date&pageSi ze=25&ftype=znanstveno+%c4%8dasopisje&flocation=ZRC+SAZU&fformattypeserial=journal>

Each journal has its own ISSN and eISSN, each article has its own DOI, authors and editors are linked to their ORCID. The ROR identifier has only recently been implemented and is not yet in use. The ARK identifier and funder DOIs/PIDs plug-ins are not installed and the Handle identifier is not supported in the version of OJS used.

For harvesting **metadata** exchange the OAI-PMH, DC and OpenAIRE plug-ins are available in the current version of OJS used by Založba ZRC, mass metadata export and metadata about OA are also supported. OpenAIRE compliance is enabled with the plug-in, but is not yet in use, the KBART plug-in is not available in OJS 3.3.0.10.

Information about the **content**, such as human- and machine-readable information about OA status, copyright holder and licensing, is provided on landing pages in each publication in a standard non-proprietary format, TDM is technically supported with OAI-PMH.

Full text in machine-readable format is also supported, articles are published in PDF. Both issues in 2022 of the journal *Two homelands*¹⁷⁷ were published also in XML format (exporting in JATS XML via InDesign). This was a test project for publishing journals also in XML, but the chosen method of preparing the text in XML format did not turn out to be suitable. All content of ZRC SAZU journals will be deposited in a digital preservation service using the PKP PN plug-in in 2024, from 2015 all publicly funded journals are obliged to deposit the digital content of their journals in the Digital Library of Slovenia, DLib¹⁷⁸.

Unique URLs are in use for landing pages for all articles. There is all of the metadata about the article (title, abstract, article's DOI, licence, copyright and references) and author (ORCID, affiliation, email address).

The COUNTER plug-in is enabled, but this does not mean that a journal is COUNTER compliant. By registering in the OJS system, users have the possibility to receive notifications about the publication of new articles/journal issues, and for the journals (6 out of 13) that use OJS for the editorial process, the notification is also done via the OJS system.

5.3.3 Conclusion

A review of the standards on the OJS platform used by ZRC SAZU for publishing suggests that most of the standards discussed in this document have been implemented.

As open publishing and its standards are constantly evolving, it is essential to keep abreast of new developments in the field (both content-related and technical) in order to provide good support to editorial teams. An important role in this is played by (1) the education of editors about open publishing and its importance, and the possibilities offered by the OJS for

¹⁷⁷ <https://ojs.zrc-sazu.si/twohomelands/issue/view/841>

¹⁷⁸ <https://www.dlib.si/?&language=eng>

achieving open publishing standards, visibility and discoverability, and (2) the assistance of technical staff who can provide support in the technical implementation of the standards.

Identifiers	Supported?	Adopted?
ISSN/eISSN identifier	Yes	Yes
ORCID identifier	Yes (needs to be upgraded)	No, ORCIDs are entered in author metadata manually
ROR identifier	Yes (installed)	No
ARK identifier	Yes (plug-in not installed)	No
Handle identifier	No Not supported in OJS	No
DOI identifier	Yes	Yes
Funder DOIs/PIDs	Yes (Funding plug-in not installed)	No
Support of PIDs for authors/funders	Yes (ORCID and Funder DOI)	No
Metadata	Supported?	Adopted?
Metadata exchange for harvesting (e.g. DC, OpenAIRE, OAI-PMH s etc.)	Yes (OAI-PMH)	Yes, DC 1.1 metadata in oai (DC Indexing plug-in)
OpenAIRE Guidelines	Yes, installed OpenAIRE plug-in 2.0	No
Mass metadata export (as CSV files, ONIX XML feeds or in any other established format)	Yes	Yes
KBART	No, it is not in our version of OJS	No
Metadata about OA status	Yes	Yes
Content	Supported?	Adopted?
Human- and machine-readable information about the OA status, copyright holder and licensing is provided in each publication in	Yes	Yes

a standard in a non-proprietary format		
Text and Data Mining is technically supported	Yes	Yes
Full text in machine-readable format (JATS XML or equivalent (e.g. TEI))	Yes	No
Deposited in a digital preservation service	Yes (PKP PN installed, possible upgrade)	
Website features	Supported?	Adopted?
SEO	Yes	No
Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review (where relevant)	Yes	No
Unique URL for landing pages	Yes	Yes
URLs linking to related research objects	Yes	No
COUNTER	Yes	No

Table 24: Support and adoption information for ZRC SAZU journals

5.4 OpenEdition

5.4.1 Introduction

The visibility of OpenEdition's (OE) OA journals in DOAJ represents an interesting use case for analysing the gap between diamond publishers and infrastructures focusing on some limitations, not on the publishers' side, but on the infrastructures'.

OpenEdition Journals¹⁷⁹ (OEJ), is the OE service dedicated to the publication of journals, which, for the most part, belong to the diamond publishing category¹⁸⁰. Having identified

¹⁷⁹ <https://journals.openedition.org/>

¹⁸⁰ To date, in November 2023, OEJ contains 620 journals, 520 of which are in open access, and all of these can be considered as diamond journals. As this particular information about the business model is not recorded yet as such, the number of actual diamond journals on OEJ can only be estimated. The proportion of diamond journals on OEJ can be estimated to be 83%.

DOAJ as a major player for OA publishing visibility, OEJ made a significant effort over the years to increase the amount of its journals indexed in the DOAJ. Although the general result of such action is rather satisfying, a deeper analysis of the numbers shows however that there are still some bottlenecks limiting a smoother and broader integration of OA journals in the broader ecosystem.

5.4.2 Context

OE's first interactions with DOAJ go back to 2007. In September 2007, the first OEJ journals integrated DOAJ's collections¹⁸¹ and, in December of the same year, OE was the first French organisation to become a member of the DOAJ. From this date, referencing journals on DOAJ became a specific objective of OEJ. Assigned to different departments and job positions over the time, journals' indexing on DOAJ thus remained a permanent activity within the organisation since 2007. The direct reference to DOAJ in major policy documents related to OA (Plan S, French Committee for Open Science criteria), only confirmed OE's will to pursue its effort in this direction.

In this context, OEJ regularly updates its platform's features to match DOAJ's requirements: creation of a webpage dedicated to the description of publishing policies on the journal's website; addition of new metadata fields compliant with DOAJ's criteria (e.g., open licences field). OE also provides an exhaustive documentation about DOAJ's submission process¹⁸², and concretely supports the journals to achieve their submission.

5.4.3 OEJ journals on DOAJ: a tentative assessment

Statistics

The table below gives the main figures regarding the referencing of OEJ journals on DOAJ from 2007 to 2023. The table reports the number of OA journals on OEJ (both active and inactive), the number of active journals already registered on DOAJ, and the ratio between the two figures for active journals. The last two columns give more details about the proportion of French journals on DOAJ.

Table 25 allows us to see that the actions conducted by OEJ translated into a regular increase of OE's journals indexed on the DOAJ and overall a good ratio of OE's journals' submissions validated by the DOAJ. Moreover, OEJ also allows for a good representation of French journals in the DOAJ (in 2023, approx. 73% of French journals on DOAJ come from OEJ).

Year	OA journals on OE	OEJ on DOAJ	Ratio Active OEJ /	Total journals on DOAJ	Total French Journals on
------	-------------------	-------------	--------------------	------------------------	--------------------------

¹⁸¹ <https://leo.hypotheses.org/37>.

¹⁸² <https://objs-fr.hypotheses.org/2807>

	Total Actives			OEJ on DOAJ		DOAJ
2007			14		3,032	56
2008					3,812	73
2009	98		21	21%	4,506	82
2010	142		63	44%	5,936	118
2011	172		66	38%	7,372	136
2012	219		115	52%	8,521	175
2013	251		127	51%	9,804	
2014	296		133	45%	10,144	
2015	290				10,965	
2016	333				9,455	
2017			172		10,753	
2018					12,438	
2019					14,150	
2020					15,677	
2021	462		210	45.5%		
2022	501		224	44.7%		
2023	520	462	237	51%	20,114	322

Table 25: OpenEdition's OA journals in DOAJ collections (2007-2023)

Despite their positive aspect, figures also show that the share of OE's journals referenced on DOAJ remains stable (around 50%), while it would be possible to assume that the support provided would allow for a progressive improvement of the submissions' validation pace. In fact, for several years the workflow effectively limited the validation to 2 journals per month. Following that pace, the validation of existing and new OEJ journals would have approximately required 10 years.

Moreover, beyond OE's specific case, and even considering the differences regarding the definition of a "French" journal, the number of OA French journals on DOAJ appears to be globally very low (the mir@bel platform¹⁸³ references 2,290 OA French journals, from which

¹⁸³ <https://reseau-mirabel.info/>

only approx. 10% are indexed on the DOAJ). Given the effort of the French research community to be referenced in DOAJ, the numbers seem to show that some infrastructural limitations do exist.

In the context of OEJ, indeed, the technical and human support provided to the journals allow to find reasons for this discrepancy beyond the limitations of the platform's features or of the publishers' awareness about the DOAJ's requirements. Identified limitations appear to be rather related to either legal specificities, or retroactivity management, or disciplinary practices.

Legal specificities

Author's rights retention

To conform with OA definition, DOAJ's criteria indicates that "authors must retain all their copyright", which corresponds to the definition of copyright licensing in common law countries. In this context, the author authorises the publisher to use the content only within the limits of a licence, and retains in this way his/her copyright. However, in contract law countries, such as France, more protective of authors, the author's work use is by default assumed to have been authorised on a non-exclusive basis. In this context, the mention "cession de droits d'auteur" (approximative translation: "author's rights transfer") does not necessarily mean that the transfer is exclusive; further information is required to determine if the transfer is or is not exclusive. Therefore, the French mention "cession de droits d'auteur à titre non exclusif" (approximative translation: "author's rights non exclusive transfer") ensures authors' rights remain protected and is actually independent from the licence chosen for the publication. As a consequence, it is arguable that OEJ French journals using this mention respect the French legal context, do not transfer the full copyright to the publisher, and do not contradict the OA principles.

"All rights reserved" mention

According to DOAJ's specifications, the mention "all rights reserved" is always in contradiction with the OA principles, and even more so whenever it appears in a publication under a Creative Commons (CC) licence. The mention appears however in some OEJ's publications, with reference to images or other contents reproduced within the publication and having a different authorship than the text of the publication. In a specific example, an article is published under a Creative Commons Attribution (CC BY) licence and the final page of the publication specifies that the CC licence applies only to the text, while all the other elements fall under the "all rights reserved" protection. According to legal experts, it seems indeed possible to consider that the CC licence does not apply exhaustively to all the publication's content without infringing the OA principles.

Retroactivity management

The update of the admission criteria due to changes of policy implies a retroactive full update

of journals' legacy collections to ensure they remain indexed. Such an update, especially when regarding the legal aspects mentioned above, raises implementation issues which further hinder the process of journals' validation against the new criteria. If the creation of a new metadata field implies an update of the information system at various levels, the retroactive update of journals' information related to authorship can be even more challenging as it requires decisions that may contradict previous arrangements and therefore time to explain and implement the changes.

Disciplinary practices

A more recent change in DOAJ's criteria affected the publication of special issues¹⁸⁴. The abuse of this form of publication by predatory publishers led to refuse journals where all content is published as special issues. It is, however, a long established tradition in some Social Sciences and Humanities fields to publish exclusively in this format, which is precisely chosen in order to publish extensive thematic issues. It is not possible to evaluate at this stage the impact of this new criteria, but it illustrates how the setting of acceptable criteria for OA publications can potentially contradict bibliodiversity.

In conclusion, while the issues quickly described here are different in terms of nature and of impact, they can, all combined, explain how visibility of OA journals remains structurally limited. Indeed, the legal blocking points mentioned above are in themselves generic, related to cultural important differences, and can have therefore a general effect. Furthermore, as they impact, in this use case, a French publishing platform dealing mostly with diamond journals, they are probably a good indication about the effort that the infrastructures themselves must make to increase diamond journals visibility in the context of CRAFT-OA.

In fact, the issue related to retroactivity shows that the building, not only of services for the diamond community, but of the diamond community itself, comes with the responsibility for the leading organisations to establish policies that are consensual, inclusive and stable enough for the whole community. The last example about the disciplinary specificities, although it will probably have a more limited impact, stresses also for better coordination of the infrastructures and organisations involved in CRAFT-OA so that they can ensure the preservation and promotion of the bibliodiversity which characterises the diamond publishing community.

Precisely, DOAJ and OE took the opportunity of this use case study to start addressing the issues listed. Firstly, it has been confirmed that the validation pace of new journals could be easily increased, as it did not correspond to any rule in place at DOAJ. Secondly, the discussion about the French legal specificities has been opened, and the two organisations will continue exchanging in order to find an agreement on this matter. In that sense, the T3.2 activities and

¹⁸⁴ See DOAJ (2023, November 02). New criteria for special issues. *DOAJ News Service*.
<https://blog.doaj.org/2023/11/02/new-criteria-for-special-issues/>

the use case study proved to be efficient to start improving coordination of the diamond publishing ecosystem.

6 DISCUSSION

The review of individual standards through different datasets has provided a lot of new insight about the technical status of OADJs and where such journals stand in terms of meeting key mandatory and recommended standards. We could perceive some notable differences in the trajectories of development and adoption for the standards that we inspected, which included among other things the following observations:

- Some standards practices are still emerging, making it hard for technical development and actual practice to rapidly develop (e.g. how use and validation of RORs should fit into journal workflows).
- Other standards are technically not all that challenging, having support integrated into commonly used publishing content management systems, but awareness and adoption still has room to grow (e.g. long-term preservation, ORCIDs). Similar observations and conclusions were also found by Frantsvåg & Strømme¹⁸⁵ concerning lack of enrollment in long term preservation.
- Some standards require ongoing technical expertise to process content correctly (e.g. JATS XML publishing), which can be challenging based on what we know about the resourcing of many OADJs from previous surveys of the landscape.
- Progressive open science approaches (e.g. open peer review or commentary of content) pose challenges of adoption because they are disruptive to the social practices and established practices within scholarly communities, not because there would not be software support for such practices available.
- Standards that cost money for every time they are complied with (e.g. use of DOIs) can create obstacles for equitable opportunities for different journals to be visible and participate in the research arena.
- Many standards would almost automatically be fulfilled if the journal would reside on a modern publishing content management system rather than something self-built or a generic website framework not tailored to publishing specifically.

It is particularly this last point that emerged as particularly common, where we already at the start of the introduction to the gap analysis could see that the software environment for OADJs is very heterogeneous and essentially got confirmation to that as the standards and their uptake was reviewed. It would be possible for OADJs to comply with many standards just by changing software and after that changing nothing else, however, journal migration is not an easy task and can be either a technically complex process or/and a very labour-intensive manual process so it can not be put forward as an easy solution. The distributed nature of OADJs into smaller journals published by organisations that are not professionally

¹⁸⁵ Frantsvåg & Strømme, 2019.

dedicated to publishing is a challenge for having technical expertise available to every outlet to consult, support and perform more technical work when it comes to journal operations. That is why solutions should be as accessible, guided and managed as possible to ensure that as many as possible can be compliant with the various technological standards that are important for journal publishers active in the scholarly landscape.

The look at journal portals through the case studies was a very useful supplementary perspective and provides information for what type of support materials and interventions might make sense to communicate both to operators of such portals as well as to journals residing on such portals. One might have the idea that journals residing on shared and well-maintained technical infrastructures would all be meeting the highest of recommended technical standards, but a lot of the challenge seems to come with getting actual adoption even though the technical readiness is there waiting to be used. In such situations, portals can create their own requirements and recommendations for what type of processes and standards journals should adhere to if they are part of the platforms, but it goes without saying that introducing such aspects might not always be perceived positively if they are not well-motivated and come bundled with support, and some idea for how potential added resource needs for journals should be catered for.

For the purposes of D3.3 and the design and production of training materials, there should be valuable evidence within this deliverable, collected and interpreted together for the first time, in order to make informed decisions about which type of training materials would best fit with each type of gap and standard that was observed.

7 CONCLUSIONS AND OUTLOOK

With regards to overcoming technical challenges for reaching compliance, the evidence of the gap analysis overall suggests that parts of the OADJs would likely be better off switching to a modern publishing content management system rather than attempting to reach compliance with individual standards building upon some alternative technological solutions. The increasingly advanced and sophisticated digital and interconnected scholarly journal landscape is moving forward from just provision of PDFs to weaving in metadata as well as content into a wider web of global information about research. The second challenge of actually making use of the technological possibilities that wait to be utilised in a chosen software solution is a problem of a completely different nature that likely requires a multi-pronged approach where training materials and support are one crucial element, and it is here where the next task in CRAFT-OA WP3 will take over from where this task leaves off. Just as there are benefits of scale to technology development and standardisation, there are also benefits in establishing best practices and recommended work processes. In this way, adoption is made as easy as possible for those with the capacity to change their ways of working.

In order to overcome the gaps discussed in the literature and observed in our data analysis, CRAFT-OA WP3 will develop training materials for journals and journal platforms. However, we do not aim to duplicate work and will make use of existing resources as much as possible. A toolkit for OA journals¹⁸⁶ was recently launched as a collaboration between the Open Access Scholarly Publishing Association (OASPA) and DOAJ, which also contains pages that give advice for how journals can implement technical standards¹⁸⁷. Also, PKP is providing online tutorials on how journals can become compliant with the Plan S implementation guidelines¹⁸⁸ and provide guidance on how to configure the software to fulfil the requirements of DOAJ indexation¹⁸⁹. Also, the DOAJ itself provides advice for journals¹⁹⁰. Thus, one can assert that for many technical requirements both guidance and technical solutions are already available, but nonetheless, implementation is lacking. Hence, the main challenge for CRAFT-OA seems to be bringing the knowledge about such guidance and built-in technical solutions to editors and (technical) staff operating journals and platforms at institutions. To achieve this, two crucial aspects need to be considered.

On the one hand, the target group needs to be approached. This need is closely linked to T7.2 of the CRAFT-OA project, which aims to establish a sustainable network of IPTPs. The

¹⁸⁶ <https://www.oajournals-toolkit.org/>

¹⁸⁷ <https://oaspa.org/information-resources/principles-of-transparency-and-best-practice-in-scholarly-publishing/>

¹⁸⁸ <https://docs.pkp.sfu.ca/plan-s/en/>

¹⁸⁹ <https://docs.pkp.sfu.ca/doaj/en/>

¹⁹⁰ <https://doaj.org/apply/guide/#advice>

international project consortium of CRAFT-OA provides an excellent basis to communicate within various national and disciplinary networks and beyond. Therefore, close exchange is also needed with T7.4, which is dedicated to project communication. Moreover, the collaboration with the DIAMAS project will be of great value in approaching the right audience. The main focus of DIAMAS are IPSPs. However, in institutional publishing often there is no clear distinction between IPTPs and IPSPs. Therefore, it is important to make use of the network around the DIAMAS project, too.

On the other hand, awareness needs to be built that the implementation of technical standards is not an end in itself, but that each technical standard enables functionalities the journals can benefit from. As elaborated in sections on the individual standards in our gap analysis, this is possible for each and every standard.

T3.3 therefore will work to deliver education, training and a toolkit along these lines - focussing on stakeholders actually operating journals/publishing infrastructures and raising awareness on the utilities of the individual standards. Thereby, we will take into account specifically the gaps we identified in this report. However, special needs from the IPTPs network may also be addressed. The materials such as hands-on checklists, self-learning courses and the toolkit will be conceptualised and delivered throughout the remaining project runtime. Any workshops will be coordinated in accordance with WP2 (Technical coordination and frameworks for community support) and the resulting training materials will be available long-term through WP7.

8 LIST OF REFERENCES

All references and websites mentioned in the document were last checked for availability on 21.12.2023.

Ancion, Z., Borrell-Damián, L., Mounier, P., Rooryck, J., & Saenen, B. (2022). *Action Plan for Diamond Open Access*. Zenodo. <https://doi.org/10.5281/zenodo.6282403>

Armengou, C., Edig, X. van, Laakso, M., & Umerle, T. (2023). *CRAFT-OA Deliverable 3.1 Report on Standards for Best Publishing Practices and Basic Technical Requirements in the Light of FAIR Principles* (under EC review). Zenodo. <https://doi.org/10.5281/zenodo.8112662>

Armengou, C., Redhead, C., & Rooryck, J. (2023a). *DIAMAS Deliverable 3.5 Extensible Quality Standard in Institutional Publishing (EQSIP)_V1.0* (under EC review). Zenodo. <https://doi.org/10.5281/zenodo.7923916>

Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, P.-C., & Proudman, V. (2021). *OA Diamond Journals Study. Part 1: Findings*. Zenodo. <https://doi.org/10.5281/zenodo.4558704>

Bosman, J., Frantsvåg, J. E., & Kramer, B. (2021a). *OA Diamond Journals Study. Dataset* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4553103>

Brun, V., Torny, D., & Pontille, D. (2023). *DIAMAS Deliverable 3.3 Report on the gap analysis results (1.0)* (under EC review). Zenodo. <https://doi.org/10.5281/zenodo.10083615>

COAR (2023, September 28). Persistent Identifiers: Addressing the challenges of global adoption. *COAR Confederation of Open Access Repositories*. <https://www.coar-repositories.org/news-updates/persistent-identifiers-addressing-the-challenges-of-global-adoption/>

Corcho, O., Eriksson, M., Kurowski, K., Ojsteršek, M., Choirat, C., Sanden, M. van de, & Coppens, F. (2021). *EOSC interoperability framework – Report from the EOSC Executive Board Working Groups FAIR and Architecture*. <https://data.europa.eu/doi/10.2777/620649>

de Castro, P., Herb, U., Rothfritz, L., & Schöpfel, J. (2023). *Failed PIDs and unreliable PID implementations*. Knowledge Exchange. <hal-04245144>

de Jonge, H., Kramer, B. Michaud, F., & Hendricks, G. (2023, September 06). Open funding metadata through Crossref; a workshop to discuss challenges and improving workflows. *Crossref Blog*. <https://www.crossref.org/blog/open-funding-metadata-community-workshop-report/>

DOAJ (2023, November 02). New criteria for special issues. *DOAJ News Service*. <https://blog.doaj.org/2023/11/02/new-criteria-for-special-issues/>

Frantsvåg, J. E., & Strømme, T. E. (2019). Few Open Access Journals Are Compliant with Plan S. *Publications*, 7(2): 26. <http://dx.doi.org/10.3390/publications7020026>

French, A., Hendricks, G., Lammey, R., Michaud, F., & Gould, M. (2023, September 07). Open Funder Registry to transition into Research Organization Registry (ROR). *Crossref Blog*. <https://www.crossref.org/blog/open-funder-registry-to-transition-into-research-organization-registry-ror/>

Gorraiz, J., Melero-Fuentes, D., Gumpenberger, C., & Valderrama-Zurián, J. C. (2016). Availability of digital object identifiers (DOIs) in Web of Science and Scopus. *Journal of Informetrics*, 10(1), 98-109. <https://doi.org/10.1016/j.joi.2015.11.008>

Greene, J. W. (2017). Developing COUNTER standards to measure the use of Open Access resources. *Qualitative and Quantitative Methods in Libraries*, 6(2). <https://www.qqml-journal.net/index.php/qqml/article/view/410/404>

Khanna, S., Ball, J., Alperin, J. P., & Willinsky, J. (2022). Recalibrating the scope of scholarly publishing: A modest step in a vast decolonization process. *Quantitative Science Studies*, 3(4): 912-930. https://doi.org/10.1162/qss_a_00228

Khanna, S., Raoni, J., Smecher, A., Alperin, J., Ball, J., & Willinsky, J. (2022a). *Details of publications using software by the Public Knowledge Project* (Harvard Dataverse; Version 3) [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/OCZNVY>

Konrad, M. (2016, July 04). Data Mining PDFs – The Simple Cases. *WZB Data Science Blog*. <https://datascience.blog.wzb.eu/2016/07/04/data-mining-pdfs-the-simple-cases/>

Kramer, B., & de Jonge, H. (2022). The availability and completeness of open funder metadata: Case study for publications funded by the Dutch Research Council. *Quantitative Science Studies*, 3(3), 583-599. https://doi.org/10.1162/qss_a_00210

Laakso, M., Matthias, L., & Jahn, N. (2021). Open is not forever: A study of vanished open access journals. *J Assoc Inf Sci Technol*. 72, 1099-1112. <https://doi.org/10.1002/asi.24460>

Laakso, M., & Multas, A.-M. (2023). European scholarly journals from small- and mid-size publishers: mapping journals and public funding mechanisms, *Science and Public Policy*, 50(3), 445-456. <https://doi.org/10.1093/scipol/scac081>

Late, E., Korkeamäki, L., Pölönen, J. & Syrjämäki, S. (2020). The role of learned societies in national scholarly publishing. *Learned Publishing*, 33, 5-13. <https://doi.org/10.1002/leap.1270>

Linna, A.-K., Holopainen, M., Ikonen, A., & Ylönen, I. (2020). Kotimaiset tieteelliset julkaisusarjat ja avoimuus. *Informaatiotutkimus*, 39(4), 4-32. <https://doi.org/10.23978/inf.98656>

Liu, J. (2021). Digital Object Identifier (DOI) Under the Context of Research Data Librarianship. *Journal of eScience Librarianship*, 10(2): 4. <https://doi.org/10.7191/jeslib.2021.1180>

Martínez-Galindo, F. J., Rubio, F., Hernández-San-Miguel, J. & Fernández Burguete, S. (2019). Plan S: Challenges and Opportunities in Spain. *Insights*, 32(1). <https://doi.org/10.1629/uksg.467>



OpenAIRE (2012). *OpenAIRE Guidelines for Literature Repository Managers 3.0*. <https://cordis.europa.eu/docs/projects/cnect/5/283595/080/deliverables/001-D44OpenAIREGuidelinesv3DELIVERABLE.pdf>

OpenEdition Team (2023, February 02). Licensing the metadata describing content published on OpenEdition. *Open Electronic Publishing*. <https://oep.hypotheses.org/3245>

Open Science Coordination in Finland, Federation of Finnish Learned Societies (2019). Open access to scholarly publications. National Policy and executive plan by the research community in Finland for 2020–2025 (1). *Responsible Research Series*. <https://doi.org/10.23847/isbn.9789525995343>

Pogačnik, A. (2022). Slovenske znanstvene revije v letu 2021, *Organizacija znanja*, 27(1-2). <https://doi.org/10.3359/oz2227002>

Pölönen, J., Syrjämäki, S., Nygård, A.-J., & Hammarfelt, B. (2021). Who are the users of national open access journals? The case of the Finnish Journal.fi platform. *Learned Publishing*, 34(4), 585–592. <https://doi.org/10.1002/leap.1405>

Rathemacher, A., Ragucci, M., & Doellinger, S. (2022). Don't Wait, Automate! Industry Perspectives on KBART Holdings Automation. *The Serials Librarian*, 82(1-4), 91–97. <https://doi.org/10.1080/0361526X.2022.2019545>

ROR (2023, October 12). How ROR and the Open Funder Registry Overlap: A Closer Look at the Data. *ROR Blog*. <https://ror.org/blog/2023-10-12-ror-funder-registry-overlap/>

Ross-Hellauer, T. (2017). What is open peer review? A systematic review [version 2; peer review: 4 approved]. *F1000Research*, 6:588, <https://doi.org/10.12688/f1000research.11369.2>

Schirrwagen, J. (2019, April 26). Open Journal Systems (OJS) sets new standards to achieve OpenAIRE compliance with JATS. *OpenAIRE Blog*. <https://www.openaire.eu/blogs/open-journal-systems-ojs-sets-new-standards-to-achive-openaire-compliance-with-jats>

Shepherd, P. T. (2010). *COUNTER: current developments and future plans*. *The E-Resources Management Handbook*. UKSG. <https://www.uksg.org/sites/default/files/5-Shepherd-W1G177781L48WR13.pdf>

Scholastica (2022). *The State of Journal Production and Access 2022: Report on survey of independent academic publishers*. *Scholastica Report*. <https://lp.scholasticahq.com/journal-production-access-survey-2022/>

van Ballegoie, M., Meares, S., & Wilson, K. (2017). Deep Dive into KBART. *The Serials Librarian*, 72(1-4), 15–25. <https://doi.org/10.1080/0361526X.2017.1309826>

9 LIST OF WEBSITES

<https://arks.org/about/>

<https://avointiede.fi/en/policies/policies-open-science-and-research-finland/open-science-and-research-initiative>

https://avointiede.fi/sites/default/files/2022-10/Loppuraportti-rahoitusmalliehdotukset-1022_0.pdf

<https://clockss.org/>

<https://clockss.org/about/>

<https://coara.eu/coalition/working-groups/>

<https://cop5.projectcounter.org/en/5.0.2/00-foreword.html>

<https://creativecommons.org/publicdomain/zero/1.0/deed.en>

<https://diamasproject.eu/>

<http://dirrosdata.ck.uni-lj.si/en/nactr-s/>

<https://dirrosdata.ck.uni-lj.si/en/odprta-znanost-v-evropski-uniji/>

<https://www.dlib.si/?&language=eng>

<https://doaj.org/apply/guide/#advice>

<https://doaj.org/apply/seal/#criteria>

<https://doaj.org/csv>

<https://doaj.org/oai?verb=Identify>

<https://doaj.org/preservation/#open-access-journals-must-be-preserved-forever>

<https://docs.google.com/document/d/e/2PACX-1vR6Xf8ugrjnljmf79cW-lqeu-bYnGpGgiAGIalxb7su6lslpC00FO8wHd-5pOl0cyicKeuG7fsRcJ8g/pub>

<https://docs.pkp.sfu.ca/accessibility-statement/en/>

<https://docs.pkp.sfu.ca/doaj/en/>

<https://docs.pkp.sfu.ca/learning-ojs/en/statistics>

<https://docs.pkp.sfu.ca/plan-s/en/>

<https://docs.pkp.sfu.ca/plan-s/en/#article-metadata-quality>

<https://edition.fi/>

<https://edition.fi/tsv/catalog/book/157>

<https://eosc-portal.eu/eosc-interoperability-framework>

<https://explore.openaire.eu/>

<https://github.com/ajnyga/funding>

<https://github.com/BirkbeckCTP/janeway/issues/3168>

<https://github.com/ojsde/openAIRE>

<https://github.com/pkp/orcidProfile>

<https://github.com/pkp/sword>

<https://github.com/UB-Heidelberg/kbartExport>

<https://github.com/withanage/ror>

<https://github.com/yasielpv/pkp-ark-pubid>



<https://guidelines.openaire.eu/en/latest/>
<https://hdl.handle.net/>
<https://hrcak.srce.hr/en>
<https://hrcak.srce.hr/en/kbart>
<https://hrcak.srce.hr/en/kbart-html>
<https://hrcak.srce.hr/en/politike>
<https://hrcak.srce.hr/oai/?verb=Identify>
<https://hrcak.srce.hr/ojs/>
<https://i4oc.org/>
<https://jats.nlm.nih.gov/index.html>
<https://journal.fi/>
<https://journals.openedition.org/>
<https://julkaisufoorumi.fi/en/evaluations>
<https://keepers.issn.org/>
<https://leo.hypotheses.org/37>
<https://oai-openedition.readthedocs.io/en/latest/index.html>
<https://oai-openedition.readthedocs.io/en/latest/qdc.html#dcterms-accessrights>
<https://oaspa.org/information-resources/principles-of-transparency-and-best-practice-in-scholarly-publishing/>
<https://objs-fr.hypotheses.org/2807>
<https://ojs.sazu.si/>
<https://ojs.zrc-sazu.si/twohomelands/issue/view/841>
<https://ojs-gr.zrc-sazu.si/>
<https://omp.zrc-sazu.si/zalozba>
<https://openaccess.mpg.de/Berliner-Erklaerung>
<https://pkp.sfu.ca/ojs/>
<https://pkp.sfu.ca/pkp-pn/>
<https://pkp.sfu.ca/software/ojs/>
<https://pkp.sfu.ca/software/omp/>
<https://prep.labs.crossref.org/>
<https://reseau-mirabel.info/>
<https://ror.org>
<https://web.archive.org/web/20230327004401/https://kotilava.fi/19-elokuu-2016-1247/kotilava-%E2%80%93-finnish-academic-journals-towards-immediate-open-access>
<https://tsv.fi/en/services/label-for-peer-reviewed-scholarly-publications>
<https://tsv.fi/palvelut/avoimen-julkaisemisen-palvelut/journalfi>
<https://web.hypothes.is/>
<http://www.arrs.si/en/>
<https://www.arrs.si/sl/dostop/strategija.asp>

<https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/>
https://www.coalition-s.org/technical-guidance_and_requirements/
<https://www.coar-repositories.org/news-updates/persistent-identifiers-addressing-the-challenges-of-global-adoption/>
<http://www.dariah.si/dariah-si-eng/>
<https://www.dlib.si/?&language=eng>
<https://www.dlib.si/results/?euapi=1&query=%27keywords%3dzrc+sazu%27&sortDir=ASC&sort=date&pageSize=25&ftype=znanstveno+%c4%8dasopisje&flocation=ZRC+SAZU&fformat=typeserial=journal>
https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai_node.html
<http://www.doaj.org/oai.article?verb=Identify>
<https://www.dona.net/handle-system>
<https://www.helsinki-initiative.org/>
<https://www.iso.org/obp/ui/#iso:std:iso:26324:ed-1:v1:en>
<https://www.julkaisufoorumi.fi/en>
<https://www.lockss.org/>
<https://www.lockss.org/about>
<https://www.niso.org/standards-committees/kbart>
<https://www.niso.org/standards-committees/kbart/kbart-content-providers>
<https://www.oajournals-toolkit.org/>
<https://www.openedition.org/26973>
<http://www.openarchives.org/>
<https://www.portico.org/preservation-policies/>
<https://www.sherpa.ac.uk/romeo/>
<https://www.tsv.fi/en/grants/publishing-and-international-activities-scientific-societies>
<https://www.tsv.fi/en/grants/publishing-and-international-activities-scientific-societies/application-instructions>
https://www.tsv.fi/en/toiminta_en/membersocieties
https://www.tsv.fi/sites/tsv.fi/files/media/ehdotus_avointen_kotimaisten_tiedelehtien_rah_oitukseksi.pdf
<https://www.w3.org/WAI/fundamentals/accessibility-intro/>
<https://www.zrc-sazu.si/en/>
<https://zalozba.zrc-sazu.si/en/predstavitev>
<https://zeitschriftendatenbank.de/startseite>
<https://zenodo.org/>

10 LIST OF FIGURES

Figure 1: The three types of DOAJ data used for the analysis.....	20
Figure 2: Current Journal.fi data integration	87

11 LIST OF TABLES

Table 1: Received and rejected journal application counts for the DOAJ 2018-2022.....	21
Table 2: Categories of reasons for journal application rejection to the DOAJ	22
Table 3: DIAMAS survey responses for the question what software the institutional publisher is using	26
Table 4: DIAMAS survey responses for questions concerning how the institutional publishers services and / or technical infrastructure are maintained and updated?.....	27
Table 5: “Metadata, PIDs, supplying and enriching metadata, or making metadata available for use”	45
Table 6: “Trying to achieve and maintain interoperability with other services”	45
Table 7: “Does the IPSP release its metadata openly with a standard metadata description schema (MARC, MODS, DC, ONIX, JATS, TEI)?”	47
Table 8: Are the journal’s metadata compliant with the OpenAIRE standard?	50
Table 9: “Does the journal or its publisher deposit articles (in a machine-readable community standard format such as JATS XML, and including complete metadata) into author-designated or centralised Open Access repositories that fulfil the Plan S criteria?”	58
Table 10: “Does the IPSP release its metadata openly with a standard metadata description schema (MARC, MODS, DC, ONIX, JATS, TEI)?”	60
Table 11: “Does the journal provide openly accessible data on citations according to the standards of the Initiative for Open Citations (I4OC)?”	62
Table 12: IPSP making references openly available according to the principles of I4OC (Initiative for Open Citations)	62
Table 13: “Please indicate which formats of full text are available”	67
Table 14: “Which formats does the platform make content available in?”	67
Table 15: “Does the journal allow, legally and technically, Text and Data Mining of the full text of articles by third parties?”	69
Table 16: “Please indicate which formats of full text are available”	69
Table 17: “Which formats does the platform make content available in?”	70
Table 18: What digital archiving policy does the journal use?	72
Table 19: “Is the published content actively preserved in a digital preservation service?” ...	73
Table 20: “Archiving, backing up or preserving content and software”	73
Table 21: Does the journal require linking to data, code, and other research outputs that underlie the publication and are available in external repositories?	81
Table 22: Support and adoption information for journal.fi	93
Table 23: Support and adoption information for HRČAK.....	99
Table 24: Support and adoption information for ZRC SAZU journals	104
Table 25: OpenEdition’s OA journals in DOAJ collections (2007-2023)	106

12 ANNEX

BASIC AUTHENTICATION & AUTHORISATION INFRASTRUCTURE CONCEPTS

Authentication and Authorisation are mechanisms to ensure proper access to resources. **Authentication** consists of proving the identity of a user, which is often done by requesting a secret password ("something the user knows"), possession of a hardware device ("something the user has", such as security key), biometrics ("something the user is", such as a fingerprint) or a combination of these. On the other hand, **Authorisation** consists of what a user is allowed to do in a system, which is normally implemented with associated user permissions, many times organised in groups or roles.

An Authentication and Authorisation Infrastructure (AAI) is responsible for implementing such mechanisms in a secured and trusted manner. Different roles can be identified in this infrastructure, such as the **Identity Providers** (idP), which operate as authentication servers that authenticate users, and **Service Providers** (SP), which represent the services and applications that need to be accessed in a secure and trusted manner.

The support of standards for the authentication and authorisation of users is relevant for publishing platforms, such as Open Journal Systems¹⁹¹, Janeway¹⁹² or Lodel¹⁹³. For example, Janeway supports OIDC, whereas OJS and Lodel do not seem to support any, and would require some effort for their integration.

PROTOCOLS FOR AAI

There are multiple protocols and standards that can be used for authentication and authorisation. The most relevant industry-standard mechanisms for authentication and authorisation include OIDC, SAML and certificates (X.509).

- **OIDC** (OpenID Connect¹⁹⁴) is an authentication protocol, which securely verifies the user that is connected to a system through, for example, a Web browser or mobile app. It works by adding an identity layer on top of the OAuth 2.0 framework¹⁹⁵, which offers authorisation delegation. OIDC works on the Web and it has become the de facto standard for modern Web authentication.
- **SAML** (Security Assertion Markup Language¹⁹⁶) is an open standard published by OASIS for exchanging authentication and authorization data between different

¹⁹¹ <https://pkp.sfu.ca/software/ojs/>

¹⁹² <https://janeway.systems/>

¹⁹³ <https://lodel.hypotheses.org/>

¹⁹⁴ <https://openid.net/developers/how-connect-works>

¹⁹⁵ <https://oauth.net/2/>

¹⁹⁶ <https://www.oasis-open.org/standard/saml/>

parties. Its syntax is based on XML, and allows Web applications to transfer information between the Identity Provider and the Service Provider.

- **X.509** is a standard that defines the format of public key certificates. Public key cryptography is based on the notion of asymmetric cryptography, in which different but mathematically bound keys are used for both encryption and decryption. The identity of the certificate owner is bound to the certificate using a digital signature, which requires an implicit or explicit trust on the entity that signs the certificate. X.509 certificates are widely used on the Web to protect communications (via the TLS and HTTPS protocols) and ensure secure access to information resources. Thanks to the use of public key cryptography, the authentication of the identity of users is possible.

AAI IN CRAFT-OA

Despite the open nature of Diamond OA publishing, some of the processes involved, such as the editorial process, still require these mechanisms to operate safely. Publishing software always needs to implement security policies, including authentication and authorisation mechanisms to ensure appropriate access to its different components. Additionally, the CRAFT-OA project aims to support the integration of Diamond OA content, publishers and IPSPs with the EOSC¹⁹⁷. This will be done through plug-ins that enable technical interoperability with the EOSC Catalogue, implement EOSC interoperability frameworks for publishing/accessing research products, and enable the inclusion of metadata enrichments from the EOSC catalogue. Similarly, publishing services for OJS will be federated under a common Single Sign-On access¹⁹⁸, for secure and user-friendly access across the related publishing and EOSC services. This link between publishing platforms and the EOSC is enabled with a community-driven AAI.

EOSC AAI

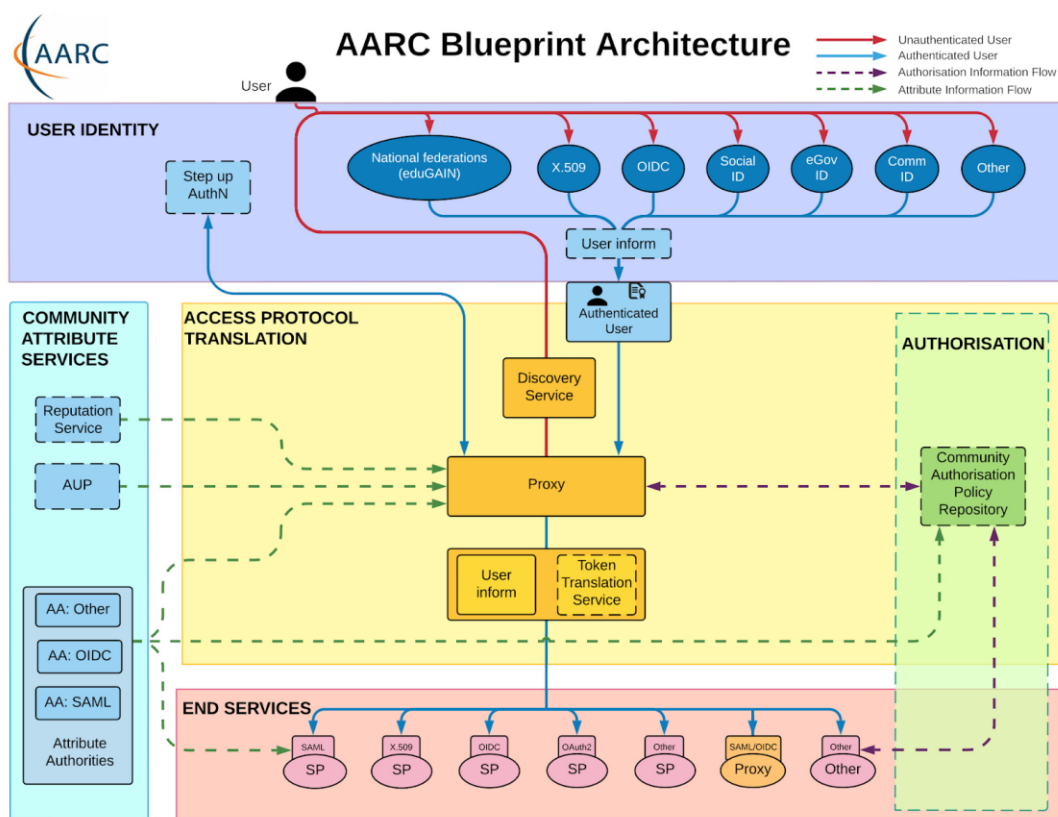
EOSC runs on a Federated environment of systems and services. The purpose of the Authentication and Authorisation Infrastructure in EOSC is to support the FAIR principles for data and services while enabling high-trust collaborations to be established and maintained with little or no friction to the end user. EOSC AAI architecture is based on the **AARC Blueprint Architecture**¹⁹⁹, depicted in **Annex Figure 1**. The figure shows the data flow from the user, since the moment it is authenticated by an Identity Provider (such as those adhered to eduGAIN²⁰⁰), releasing user attributes for authentication and authorisation, to the access to the end services offered by an SP (Service Provider).

¹⁹⁷ European Open Science Cloud, <https://eosc-portal.eu/>

¹⁹⁸ A mechanism that allows users to login once and be able to access multiple related applications and services, without the need to enter user credentials for each individual application or service

¹⁹⁹ <https://aarc-project.eu/architecture/>

²⁰⁰ <https://edugain.org/>



Annex Figure 1: AARC BPA

Implementing and configuring an AAI service compatible with EOSC AAI (and, herein, with AARC BPA) is a complex task. However, this is facilitated by services such as the **EGI Check-in** service²⁰¹, which will be used in the project. EGI Check-in is an Identity and Access Management system and an AARC Proxy service that connects federated Identity Providers (IdPs) with SPs. EGI Check-in service is a registered member of the EOSC AAI Federation and this means that all the services integrated with it will be accessible to all EOSC users in an interoperable and trusted manner: it will allow users to authenticate with their home organisation (typically, a research institute participating in eduGAIN) as well as with an academic account (e.g. ORCID), a social account (e.g. GitHub, Google, LinkedIn, etc.) and others. EGI Check-in provides a simple, integrated method to ensure EOSC users the use of CRAFT-OA services according to their defined access policies.

Note that EOSC currently uses SAML, but there is a work in progress to move from SAML to OIDC Federation 1.0²⁰².

²⁰¹ <https://www.egi.eu/service/check-in/>

²⁰² https://openid.net/specs/openid-connect-federation-1_0-28.html