D3.1

Report on standards for best publishing practices and technical requirements in light of the FAIR principles

Submission Date 2023.06.30	Version 1.0 – Submitted version
Due Date 2023.06.30	PU: Public: X SEN: Sensitive: R-UE/UE-R: EU classified:
Deliverable Title	Report on standards for best publishing practices and technical requirements in light of the FAIR principles
Deliverable No.	D3.1
Lead beneficiary	DOAJ
Contributing WP	WP3
Type	Report
	HORIZON-INFRA-2022-EOSC-01 Grant Agreement: 101094397

DOI: 10.5281/zenodo.8112662

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	Clara Armengou (https://orcid.org/0000-0003-2124-5165), Xenia van Edig (https://orcid.org/0000-0003-2510-0529), Mikael Laakso(https://orcid.org/0000-0003-3951-7990), Tomasz Umerle(https://orcid.org/0000-0002-7335-0568)
Abstract	This report reviews technical standards to better integrate journals into search engines, indexes, library catalogues, and discovery services. It analyses five policies and service-related documents which represent different perspectives within the scholarly communication ecosystem to extract mandatory and recommended technical standards. Overall, the document aims to provide insights into the technical standards and best practices necessary for achieving FAIR compliance in scholarly publishing, with a focus on the challenges faced by OA journals and platforms. It also mentions the development and implementation of training materials (future tasks) to assist journals in meeting the reviewed standards.

Version and Revision History

Version	Date	Author/Reviewer/Contributors	Comments
0.1	2023.04.19	Clara Armengou (DOAJ)	First draft
0.2	2023.06.14	Authors: Clara Armengou (DOAJ), Xenia van Edig (TIB), Mikael Laakso (TSV); Tomasz Umerle (IBL PAN) Contributors: Arnaud Gingold (AMU) (https://orcid.org/0009-0000-0466-5542), Ljiljana Jertec Musap (SRCE) (https://orcid.org/0000-0002-0059-9899), Tabea Klaus (UGOE) (https://orcid.org/0000-0002-2791-1053), Ignacio Lamata Martinez (EGI) (https://orcid.org/0000-0002-2964-6337), Isabella Meinecke (FHH) (https://orcid.org/0000-0001-8337-3619), Leonidas Pispiringas (OpenAIRE AMKE) (https://orcid.org/0000-0002-7111-094X), Jadranka Stojanovski (UNIZD) (https://orcid.org/0000-0001-7399-522X), Barbara Svetina (ZRC SAZU) (https://orcid.org/0000-0002-9636-4102)	Version submitted to reviewers
0.3	2023.06.21	Reviewers: Johan Rooryck (ESF), Jeroen Bosman (UU)	Review
0.4	2023.06.28	Authors: Clara Armengou (DOAJ), Xenia van Edig (TIB), Mikael Laakso (TSV)	Comments from reviewers integrated
0.5	2023.06.30	Contributors: Tabea Klaus (UGOE) (https://orcid.org/0000-0002-2791-1053), Theresa Waldmann (UGOE) (https://orcid.org/0009-0002-2994-6660)	Final formal check
1.0	2023.06.30		Submitted version

Disclaimer

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.

List of Acronyms

A	Accessible
AMU	Aix-Marseille University
ANR	French National Research Agency
APC	Article Processing Charge
APDOA	Action Plan for Diamond Open Access
API	Application Programming Interface
ARK	Archival Resource Key
BASE	Bielefeld Academic Search Engine
CARNET	Croatian Academic and Research Network
CC	Creative Commons
CC BY	CC Attribution License
CC BY-NC	CC Attribution-NonCommercial License
CC BY-NC-SA	CC Attribution-NonCommercial-ShareAlike License
CC BY-SA	CC Attribution-ShareAlike License
CC0	CC Zero
CLOCKKS	Controlled LOCKSS
COPE	Commitee on Publication Ethics
COUNTER	Counting Online Usage of NeTworked Electronic Resources
CRAFT-OA	Creating a Robust Accessible Federated Technology for Open Access
CRediT	Contributor Roles Taxonomy
CSV	Comma-Separated Values
D	Deliverable

DC	Dublin Core
DIAMAS	Developing Institutional Open Access Publishing Models to Advance Scholarly Communication
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identifier
DTD	Document Type Definition
EGI	EGI Foundation
eISSN	electronic International Standard Serial Number
EOSC	European Open Science Cloud
ePub	Electronic Publication
EQSIP	Extensible Quality Standard in Institutional Publishing
ESF	European Science Foundation
EU	European Union
F	Findable
FAIR	Findable, Accessible, Interoperable, Reusable
FHH	Freie und Hansestadt Hamburg – Staats- und Universitätsbibliothek Hamburg
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
I	Interoperable
I4OC	Initiative for Open Citations
IBL PAN	The Institute of Literary Research of the Polish Academy of Sciences
ID	Identifier
ISSN	International Standard Serial Number

JATS XML	Journal Article Tag Suite XML
JEL	Journal of Economic Literature Classification System
KBART	Knowledge Base And Related Tools
LOCKSS	Lots of Copies Keep Stuff Save
M	Mandatory
MA	Mandatory if Applicable
MARC	Machine-Readable Cataloguing
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Scheme
NIFO	National Interoperability Framework Observatory
NISO	National Information Standards Organization
O	Optional
OA	Open Access
OADJ	Open Access Diamond Journals
OADJS	Open Access Diamond Journals Study
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OJMO	Open Journal of Mathematical Optimization
OJS	Open Journal Systems
ONIX	Online Information Exchange
OpenAIRE AMKE	OpenAIRE AMKE
OPERAS	Open Scholarly Communication in the Social Sciences and Humanities
ORCID	Open Researcher and Contributor Identifier

PDF	Portable Document Format
PID	Persistent Identifier
R	Recommended (OpenAIRE)
R	Reusable (FAIR)
ROR	Research Organisation Registry
RSS	Really Simply Syndication
SEO	Search Engine Optimisation
SRCE	University of Zagreb University Computing Centre
SSH	Social Sciences and Humanities
TCP/IP	Transmission Control Protocol/Internet Protocol
TEI	Text Encoding Initiative
TIB	TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek Hannover
TSV	Federation of Finnish Learned Societies
UGOE	Georg-August-Universität Göttingen
UNIZD	University of Zadar
URL	Uniform Resource Locator
UU	Utrecht University
W3C	World Wide Web Consortium
W3C PROV-O	W3C Provenance Family of Specifications-Onthology
WCAG	W3C Web Content Accessibility Guidelines
WP	Work Package
XLS	Excel Spreadsheet
XML	Extensible Markup Language

ZRC SAZU	Znanstvenoraziskovalni center SAZU (Research centre of the Slovenian Academy of Sciences and Arts)

Table of Content

1 Executive Summary

"Diamond OA Journals have an enormous potential to establish and sustain an open scholarly communication. This is uncovered by the “Open Access Diamond Journals Study” (OADJS)." From 17 000 to 29 000 Open Access Diamond Journals (OADJ) published worldwide are responsible for publishing 8-9% of the world's scientific articles. This makes up 45% of Open Access (OA) publishing in general.

To develop this potential serious challenges need to be overcome, mastered and robust support needs to be provided to the Diamond OA Journals community. In 2022 the broadly supported “Action Plan for Diamond Open Access” (APDOA) was published as the follow up to the OADJS to outline the most pressing issues demanding swift action from the community. APDOA argues that the Diamond Open Access “is held back by challenges related to the technical capacity, management, visibility, and sustainability of journals and platforms”.

Associated projects Developing Institutional Open Access Publishing Models to Advance Scholarly Communication (DIAMAS) and CRAFT-OA embody this action to support institutional Diamond OA publishing. While DIAMAS focusses on developing non-technical standards and best practices, CRAFT-OA specifically targets the OADJ technology development. CRAFT-OA’s Work Package 3 (WP3) is responsible for Task 3.1 providing a technical standards’ and best publishing practices overview, Task 3.2 preparing a gap analysis to understand the challenges that OADJ’s face when aiming to comply with the standards and best practices and with/in Task 3.3 offering targeted training to narrow this gap.

This deliverable is related to the Task 3.1 and offers an overview of the technical standards and best publishing practices which is intended to be reused by the community and also to guide the gap analysis and training to be offered through WP3.

We argue that the OADJs find the current dispersion and multiplicity of requirements and standards particularly difficult both to monitor and adhere to due to the OADJs’ insufficient resources and lack of collaborative workflows. This deliverable aims to alleviate this burden through identifying key requirements and policy documents (see 2. *Definition and scope*), organising the standards they mention (see *4. Technical standards for each of the FAIR principles* and *5. Other recommended technical standards*) and showcasing best publishing practices exemplifying the implementation of standards or adherence to the requirements (see *6. Examples illustrating several or all of the basic technical standards and best publishing practices*).

The scope of this deliverable is impacted by the source documents we decided to concentrate on. We focus on two policy documents: a key, widely supported OA publishing policy paper Plan S and the Extensible Quality Standard in Institutional Publishing (EQSIP) compiled in DIAMAS, and two documents originating from key service providers in the OA publishing (IPSPs): the DOAJ Seal from the Directory of Open Access Journals and the OpenAIRE Guidelines for Literature Repository Managers v4. As this report aims to contribute to the interoperability of Diamond Open Access publishing, especially in the context of the European Open Science Cloud (EOSC), the EOSC interoperability framework was reviewed, but no concrete standards above those mentioned in the other documents were extracted. Chapter *4. Technical standards for each of the FAIR principles* represents the overview of technical standards. These documents identify in the view of how they contribute to the OADJ's findability, accessibility, interoperability or reusability (as defined by the Findable, Accessible, Interoperable, Reusable (FAIR) principles). The standards expected or recommended by these documents serve as a good representation of what is now considered quality OA publishing by the community.
However, there are also other standards worth mentioning (see and *5. Other recommended technical standards*) which are better suited to be discussed outside of the FAIR principles framework. We especially recognise that the larger context for OADJs is the EOSC which is developing its interoperability framework . It is important to be mindful of the interoperability challenges as they are stated in the EOSC ecosystem as onboarding of OADJ in EOSC is recognised by CRAFT-OA as the key factor for their visibility and sustainability. Additionally, as we use the FAIR principles as an important framework structuring this report we need to recognise that the FAIR principles development and implementation have their own dynamic that in some respects may not correspond to the standards development specific for the Diamond publishing. This is because while some OA publishing standards fall under the FAIR compliance, others do not. FAIR compliance is an important factor, but it is not a deciding factor for the inclusion of standards in this report.

There are already a number of actors in the OADJ ecosystem that comply with the discussed standards and can be regarded as illustrations of the best publishing practices. Out of many examples we are focusing on 1. two DOAJ-indexed journals Open Journal of Mathematical Optimization (OJMO) and Arheološki vestnik showcasing a platform for a single DOAJ-compliant journal, 2. two publishing platforms based on the Open Journal Systems (OJS) software (TIB[1] Open Publishing and HRČAK - Portal of Croatian scientific and professional journals which illustrate the effects of choosing a software which intends to support OA publishing, 3. workflow for the OpenEdition's journals FAIR assessment illustrating the operationalisation of FAIR principles in the editorial process.

[1] TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek Hannover

Based on the standards and best practices overview, this deliverable closes with a set of conclusions. The report emphasises the significance of interoperability in facilitating the discoverability, reuse, and reproducibility of research outputs. To our knowledge, it is the first endeavor to systematically gather and compare the distinct requirements established by the chosen policies and services with one another. Through the adoption of these standards and best practices, publishers can play a vital role in ensuring that research outputs are easily discoverable not only by peers but also by the general public. Furthermore, implementation facilitates accessibility for all stakeholders, promotes interoperability with diverse services, and enables the seamless reuse of research outputs in new research endeavors or policy-making decisions.

2 CONTEXT

The purpose of WP3 of the CRAFT-OA project is to enable the upskilling of publishing platforms, publishers, and stand-alone Diamond OA journals. It will achieve this goal by

A. reviewing standards for best publishing practice and technical standards to better integrate journals into search engines, indexes, library catalogues, and discovery services
B. identifying challenges and obstacles for OA journals and platforms to comply with such standards
C. identifying, developing, and implementing relevant help measures (i.e. training materials, workshops, etc.) for Diamond OA journals, publishers, and platforms to comply with agreed standards.

This deliverable (D 3.1) forms the foundation for the upcoming Task and Deliverable 3.2 "Challenge and gap identification for OA journals and platforms to comply with standards". By reviewing existing technical interoperability standards and the degree of implementation (mandatory/recommended) required by key stakeholders, we can construct a framework to compare the alignment of individual journals and larger groups of journals. The latter half of the WP3 project will be used to prepare training materials and a self-assessment toolkit. These will be based on the gap analysis between stakeholder requirements and the current implementation level of journals.

The information collected and organised in this deliverable will also serve as a starting point for WP5, Task 5.1, "State-of-the-art of citation indexes and aggregators in the publishing domain".

This deliverable is of general interest and utility to anyone interested in interoperability standards in scholarly publishing, but we are particularly targeting the following stakeholder groups and the functions and goals listed:

- **Publishing platforms, publishers, and stand-alone Diamond OA journals**
  - Supporting processes of professionalisation
  - Increasing interoperability, discoverability, and technical sophistication
  - Supporting cultural change while doing so
  - Ensuring knowledge transfer to researchers as authors
- **Funders and other policymakers**
  - Establishing funding compliance criteria

- **Aggregators (EOSC[2], European Union (EU) catalogues, global (non-)commercial aggregators etc.)**
    - Facilitated harvesting
    - Optimised metadata
    - Improving functionality according to FAIR
    - Improving showcase function of EU research area (ERA) results by demonstrated high quality

[2] The European Open Science Cloud (EOSC) is a European initiative that aims to create a seamless and open ecosystem for sharing and accessing scientific data and services. The EOSC envisions a future where researchers, innovators, and citizens have easy and open access to a wide range of digital resources for research, collaboration, and innovation.

3 DEFINITION AND SCOPE

We have chosen to group the reported standards and best practices according to the FAIR principles[3]. Thereby, we focus on technical aspects and leave editorial standards out. However, this distinction is somewhat subjective. In this section, we define how we have used the FAIR principles in conjunction with standards for interoperability and publishing best practices, reflecting on existing definitions and prior work in this area.

3.1 The FAIR principles

FAIR is an abbreviation of Findable, Accessible, Interoperable, and Reusable. While the FAIR principles were initially developed for research data and the respective metadata, they can also be applied to other areas of the scholarly communications ecosystem.

The main emphasis of the FAIR principles is on machine-actionability, as humans rely heavily on computational support, especially when working with research data. One main aspect of machine-actionability is interoperability.

When applying the FAIR principles to scholarly publications, we have used them as general principles rather than strictly codified standards in themselves as they are with, e.g. data management and data sharing (see, e.g. FAIR Data Maturity Model Working Group 2020). This is due to both the wide range of activities involved in journal publishing and the mix of intertwined technical and organisational aspects. These aspects need to be considered to enable ongoing interoperability and to facilitate flexible human- and machine-based usability in an evolving landscape. It is also important to clarify how publication venues and infrastructures can enable and work towards the FAIR principles through their actions and services, not how individual authors can implement FAIR in their own scholarly works.

3.2 Standards for interoperability

In the following section, we define the terms ‘standard’ and ‘interoperability’ and also discuss our use of these terms/concepts for the purpose of this report.

3.2.1 Standards

The Oxford reference dictionary defines a standard in the following way:
“A specification of the design of particular goods or components. Examples range from the gauges of screw-on nuts and bolts to the voltages of electronic equipment. Technical

[3] https://www.go-fair.org/fair-principles/

standards are needed to ensure compatibility. The use of technical standards is an example of network externalities, where everybody gains from their equipment being compatible with everybody else's. In the early days of a new product, there is a considerable competitive advantage to a firm that can get its own design accepted as the technical standard."[4]

Overall we take a very inclusive view of what counts as a technical standard, as it is not just limited to specifications issued by formal standards issuing bodies, and encompassing almost any sort of technical implementation. Standards and standardisation can be fairly straightforward if stakeholders can agree between multiple available alternatives without competitive powers at play, and the resulting adoption of the standard is rapid and comprehensive. Where it becomes more difficult is standardising while products and services are still maturing and being adopted at various stages of implementation, where iteration among alternatives is still active (de Vries, 2005), and where commercial interests of certain key players may steer standardisation processes in a particular direction. In the context of scholarly journal publishing, technologies continuing to develop and mature at the same time as standards are being set has been challenging. It has been necessary for publishers and technical content platforms to update their interfaces and enrich their metadata records regularly to keep up with the many requirements for implementing standards set by different organisations (more about these in the next section).

Interoperability

There is often a lot to gain through the standardisation of interoperability as such standards (particularly those based on open specifications) often carry network externalities. It also enables the design of complementary services and products, based on the assumption that the standard will be widely adopted. For interoperability, we share the pragmatic view of the EOSC definition that any specification, regardless of size or scope, that enables the communication of data in the realm of computer systems falls under the umbrella of interoperability.

The European Commission's National Interoperability Framework Observatory (NIFO) defines interoperability in the following way:

"Interoperability is a key factor in making a digital transformation possible. It allows administrative entities to electronically exchange meaningful information in ways that are understood by all parties. It addresses all layers that impact the delivery of digital public services in the EU, including: legal, organisational, semantic and technical aspects." (NIFO, n.d.)

[4] https://www.oxfordreference.com/display/10.1093/oi/authority.20110803102805121

This report aims to contribute to the interoperability of Diamond Open Access publishing, especially in the context of the European Open Science Cloud (EOSC). The EOSC interoperability framework (EOSC IF) uses a straightforward definition of interoperability as the "ability of different information technology systems and software applications to communicate and exchange data". (Corcho et al. 2021, p.11; also appears in Jardim & Martins, 2016). However, interoperability, when applied, can carry many meanings depending on the perspective of the actor, as the EOSC Interoperability Framework wiki observes:

"This is a complex subject, where interoperability might mean "metadata" to one group of stakeholders, but might mean the "exposure of a particular format or API" to another group" (EOSC Future, wiki.eoscfuture.eu (n.d.)a).

We treat all standards and best practices discussed in this report as contributions to the ongoing discussion about the interoperability of the EOSC ecosystem. As CRAFT-OA aims to provide sustainable solutions for the onboarding of Diamond journals in EOSC, this report is an attempt to define the interoperability of Diamond publishing in view of existing standards required, expected or recommended by key documents and services shaping the EOSC ecosystem and beyond.

The remainder of this report will build upon the definitions and descriptions of standards and interoperability provided in this section.

In the following analysis, we include only technical standards, technology-enabled functionality, and best practices required or recommended by policies and services that we deem of key importance for publication platforms.

3.3 Background

Science policy within the European Union has included a strong push for more open access publishing for many years but long-term strategy and precise mechanisms to reach those goals had not been clearly defined. That changed on the 23rd May 2023 when the Council of the European Union (2023) signed a policy document titled "On high-quality, transparent, open, trustworthy and equitable scholarly publishing". This includes the following:

"ENCOURAGES Member States and the Commission to invest in and foster interoperable, not-for-profit infrastructures for publishing based on open source software and open standards, in order to avoid the lock-in of services as well as proprietary systems, and to connect these infrastructures to the EOSC" (Council of the European Union, 2023, p.16).

This is a clear signal against what has been a mainstream practice, at least in expenditure proportion and publication volume.
Regarding scholarly journals, the conclusions call for investment in Diamond OA journal publishing rather than models based on article processing charges (APCs).

Our knowledge of the current status of Diamond OA journals is largely informed by the 2021 OA Diamond Journals Study, which provided a comprehensive overview of the technical diversity present in the Diamond OA journals landscape (Bosman et al., 2021)[5]. The democratisation of scholarly publishing, through digitalisation and openness, has made it possible for organisations and individuals that are not professional publishers to be active in this space. While this is a positive development, it also creates substantial technical challenges as these new providers may lack sufficient resources and expertise to manage the technical design and maintenance of the platform used.

The Council's conclusions also state that the technical aspects of journal implementation require improvement. It is common for journals to be hosted on content management systems designed for generic blogs (e.g. WordPress), organisational sub-pages, or custom-built basic Hypertext Markup Language (HTML) pages designed decades ago. Technical standards and best practices have evolved significantly during the last two decades, which means that uploading a full-text document to the web is not enough. There is a growing expectation and need for scholarly communication to be interconnected, which is enabled by the creation, implementation, and adoption of technical standards. Increased technical sophistication has created a demand for easy-to-use publishing platforms for managing and automating all key aspects of content and associated metadata, something which software such as Open Journal Systems (OJS) is providing as one of the major players in this area. But even with journals using OJS, tens of thousands of such journals use versions of the software that are outdated, unsupported and officially end-of-life (anything below version 3.0) (Khanna, Raoni, Smecher et al. 2022). Even if free modern tools exist, there are gaps in facilitating their adoption and continued maintenance.

Scarce resources, lack of expertise and an overwhelming variety of expectations (recommendations, requirements, standards, etc.) are likely the main reasons for the low degree of implementation of technical standards and best practices (Bosman et al., 2021). These gaps often lead to Diamond OA journals lacking visibility, discoverability, and

[5] The study analysed how OA diamond journals comply with industry standards exemplified by Plan S technical requirements. The “Scientific and editorial quality” requirements are not in the scope of this report. With regard to persistent identifiers, long-term digital preservation, machine-readable metadata in CC0, as well as the several requirements regarding copyright and licensing, the study discovered considerable gaps between the operations of the surveyed journals and Plan S compliance in most areas.

recognition and hence to a potential competitive disadvantage compared to their commercial counterparts.

In 2022, Science Europe, cOAlition S, Open Scholarly Communication in the Social Sciences and Humanities (OPERAS), and the French National Research Agency (ANR) released an "Action Plan for Diamond Open Access" (Ancion et al. 2022). It has since been endorsed by a variety of organisations.[6] The action plan calls for further development and expansion of a sustainable, community-driven Diamond OA scholarly communication ecosystem. The plan addresses the areas of efficiency, quality standards, capacity building, and sustainability. All these areas are closely connected to the implementation of technical standards and best practices by players in the Diamond OA journal landscape.

The fast rate of growth in the range of actors involved in OA scholarly publishing in recent years has been matched by the pace of development and introduction of new technical standards relevant to this space. Metadata schemas, accessibility, machine-readability, persistent identifiers (PIDs), long-term preservation, and indexing are all areas that have been rapidly evolving in terms of the possibilities and expectations for publishing outlets. Research funders are also involved, requiring specific technical interoperability standards to be implemented in the outlets where they allow grantees to comply with their mandates (cOAlition S, 2019).

Overall there is a need to identify what key publishing practices and technical standards form

1. the core requirements for digital scholarly journal publishing, as well as
2. recommended standards that should be implemented to adhere to the best practices available today and that are likely to become standards in future.

Implementing technical best practices and standards can potentially increase the visibility, findability, and long-term accessibility of publications. Another important aspect is enhanced interoperability between systems (e.g. between diamond OA journals and search engines, indexes, etc.).

3.3.1 Methodology/ Approach

The following policy- and service-related documents regarding technical standards and best practices are used for this deliverable.

[6] The list of endorsing organisations is available at https://www.scienceeurope.org/our-resources/action-plan-for-diamond-open-access/

We decided to include five policy- and service-related documents of technical standards/best practices in our report. These documents are in many of their elements forward-facing and act as aspirational targets for their target stakeholders. They do not simply depict what the majority of technology platforms, journals, and publishers already implement, but rather what the authoring bodies perceive as the currently achievable ideal to meet their current minimum and recommended requirements. When looked at together, these documents provide a unique perspective of the scholarly communication ecosystem:

- **DOAJ inclusion criteria and DOAJ extra quality criteria (Seal).** DOAJ (The Directory of Open Access Journals) is the largest **index** of peer-reviewed, fully open access scholarly research journals. DOAJ is a unique and extensive index of diverse open access journals from around the world, driven by a growing community, committed to ensuring quality content is freely available online for everyone. Thereby, DOAJ is a vital part of the global open access infrastructure. It is grounded on a global community, and DOAJ's standards have become a standard for open access publishing.
- **Plan S requirements for publication venues.** Launched in September 2018, Plan S is an Open Access publishing initiative that has garnered support from cOAlition S, a consortium of international research funding and performing organisations. As per Plan S, scientific publications that arise from publicly funded research grants must be published in Open Access journals or platforms that comply with the initiative's guidelines, beginning in 2021. The technical implementation guidelines issued by Plan S, reflect the changing policies by major **funders**.
- **OpenAIRE Guidelines for Literature Repository Managers v4.** OpenAIRE is a major **aggregator** of information on scholarly outputs of European research. In 2018, OpenAIRE A.M.K.E was established as a Non-Profit Partnership to provide a sustainable open scholarly communication infrastructure to support research in Europe. Its mission is to ensure the continuity of this infrastructure for the long term. In the course of the CRAFT-OA project, OpenAIRE will onboard further journals into the EOSC.
- **The Extensible Quality Standard for Institutional Publishing (EQSIP).** EQSIP V1.0 provides a **vision** for the future development of the diamond OA journal landscape. The EQSIP seeks to ensure the quality and transparency of governance, processes and workflows in institutional publishing and addresses the seven core components of scholarly publishing outlined in the Diamond Open Access Action Plan, which were subsequently revised and modified by the DIAMAS project team.
- **The EOSC Interoperability Framework Guidelines (**EOSC Future *(n.d.)b*)**.** The aim of the European Open Science Cloud (EOSC) is to establish a virtual environment that allows for the sharing and accessing of research data across scientific disciplines and geographical borders. This pan-European initiative has led to the development of the

EOSC Portal, which acts as a gateway to this environment by offering a centralised point of entry to a broad range of research resources and services. The EOSC Interoperability Framework was developed by the Interoperability Task Force of the EOSC Executive Board FAIR Working Group, with participation from the Architecture WG.

We analysed these policies and service-related documents to extract the required (i.e. mandatory) and recommended technical standards/requirements. We compiled a comprehensive list from this information in a spreadsheet. The spreadsheet served as a data source for the clustering presented in 4.1, 4.2 and 4.3. Some of the requirements appeared in more than one of the analysed documents.

We found that the standards varied widely in terms of granularity, meaning that the different documents analysed address different scales (from broad meta requirements to single metadata fields). While policy documents like Plan S and EOSIP V1.0 provided rather broad recommendation "frameworks" to be implemented through the adoption of actual standards (e.g. specific metadata schemes/ Document Type Definitions (DTDs)), services like OpenAIRE requested specific technical standards (Dublin Core (DC) metadata provided through an Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface).

While the DOAJ, Plan S, OpenAire, and EOSIP V1.0 requirements and recommendations are fairly detailed and relatively stable in their formulations, the EOSC interoperability guidelines are still under active development by different working groups, making mapping to FAIR principles difficult. Rather than excluding this comprehensive emerging source completely, we have instead opted to treat its mapping at a higher level. We did not mix in detailed content from the current version of the EOSC technical interoperability guidelines but related the mapping of these guidelines as contribution towards "Interoperability" as part of the FAIR principles.

4 TECHNICAL STANDARDS FOR EACH OF THE FAIR PRINCIPLES

It is important to note that the mapping across the four FAIR principles is an interpretive task since the observed elements are not always completely independent and separable; technical standards are interrelated and can and do often contribute towards enabling more than one of the FAIR principles. Furthermore, the FAIR principles themselves are also interrelated: If aspects regarding one of the principles are implemented, it is likely the performance of another principle is improved as well. Our methodology for considering these circumstances has been to map and indicate which FAIR principle each standard primarily supports. We also noted which other FAIR principles might be gaining support as they are being implemented.

This section is organised into four subsections, each dedicated to one of the four FAIR principles. Each section contains a table listing the standards identified to support the corresponding principle, and indicating which actor either recommends or requires its implementation for compliance. We have organised the text in each subsection to start with a broader description of the general and fundamental standards.

4.1 F – Findable

Concerning research data, Findability is the ability to locate data and is one of the first steps in enabling the reuse of data. However, this is also true for other scholarly information, or assets, such as research papers. Findable assets are, among other things, equipped with globally unique and PIDS, described with rich metadata and registered or indexed in a searchable resource.[7]

Findability highlights two things:

1. The emphasis on rich metadata and globally unique and persistent identifiers clearly and explicitly included in the metadata, and
2. The registration and indexation of a publication in repositories, search engines, aggregators, etc.

The metadata aspect of FAIR principles implementation is not, of course, limited to Findability, e.g. the principle “F2” explicitly underlines the value of contextual metadata for the richness of metadata.[8]

[7] Cf. https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/
[8] https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/

Metadata, i.e. "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource", (NISO, 2004), and **PIDs** play a key role in findability, as they support users, both humans and computers, in the discovery of resources.[9] Findability relies heavily on metadata, which should be as complete and rich as possible in documenting information about the actual resource (research data, publications etc.).

PIDs are especially important for the findability of a publication. There are many permanent identifiers available. The ISSN (International Standard Serial Number), and its electronic version (eISSN), are common identifiers (IDs) in the publishing industry. DOIs (Digital Object Identifiers) are common permanent identifiers in academia, and Open Researcher and Contributor Identifier (ORCID)[10] IDs are acknowledged as a reliable means of identifying researchers. Research Organisation Registry (ROR[11]) IDs are emerging PIDs for research organisations.

Rich metadata and PIDs are not enough to make a resource findable. While DOI and deposition of good metadata in the process is strategically by far the most important action for making content findable, we can differentiate three additional general scenarios: one, a service-led process of registering a publication (through assessment of whether it adheres to its inclusion criteria, e.g. DOAJ, Scopus, etc.); second, depositing a publication in a institutional, subject, or generic repository (e.g. Zenodo, etc.) which also can include adherence to a set of rules while depositing a file (although they would mostly relate to basic metadata elements); third, being indexable by services such as BASE (Bielefeld Academic Search Engine) or Google Scholar (which mainly depends on the technological features of a publication tool, e.g. a OAI-PMH protocol, Search Engine Optimisation (SEO), Uniform Resource Locator (URL) for articles, etc., compare section 5.1).

Table 1 presents the technical standards and practices supporting Findability.

Standard/Requirement	Basic/Mandatory	Recommended
ISSN/eISSN ID	DOAJ	
ORCID ID		DOAJ Seal
ROR ID	EQSIP V1.0	
Archival Resource Key (ARK) ID		DOAJ Seal

[9] In this document the term "resource" refers to any type of research output, e.g. preprints, research data and datasets, software, code, methodology, protocol, model, algorithm, exhibition, strategy, policy, as well as formal publications (journals and journal articles, books and book chapters, conference proceedings, etc.).
[10] https://orcid.org/
[11] https://ror.org/

Handle ID		DOAJ Seal
DOI ID	Plan S	DOAJ Seal
Funder DOIs	Plan S	
Support of PIDs for authors/funders		Plan S
Metadata formats for harvesting (e.g. DC, OpenAIRE, etc.)	EQSIP V1.0, OpenAIRE Guidelines for Literature Repository Managers v4, Plan S, DOAJ	
Full text in machine-readable format	DOAJ	Plan S
Title	OpenAIRE Guidelines for Literature Repository Managers v4	
Creator	OpenAIRE Guidelines for Literature Repository Managers v4	
Contributor	OpenAIRE Guidelines for Literature Repository Managers v4	
Funding Reference	OpenAIRE Guidelines for Literature Repository Managers v4	
Alternate ID		OpenAIRE Guidelines for Literature Repository Managers v4
Related ID		OpenAIRE Guidelines for Literature Repository Managers v4
Embargo Period Date	OpenAIRE Guidelines for Literature Repository Managers v4	
Language	OpenAIRE Guidelines for Literature Repository Managers v4	
Publisher	OpenAIRE Guidelines for Literature Repository Managers v4	
Publication Date	OpenAIRE Guidelines for Literature Repository Managers v4	

Resource Type	OpenAIRE Guidelines for Literature Repository Managers v4	
Description	OpenAIRE Guidelines for Literature Repository Managers v4	
Format		OpenAIRE Guidelines for Literature Repository Managers v4
Resource ID	OpenAIRE Guidelines for Literature Repository Managers v4	
Source		OpenAIRE Guidelines for Literature Repository Managers v4
Subject	OpenAIRE Guidelines for Literature Repository Managers v4	
Size		OpenAIRE Guidelines for Literature Repository Managers v4
Resource Version		OpenAIRE Guidelines for Literature Repository Managers v4
Citation Metadata: Title Citation Volume Citation Issue Citation Start Page) Citation End Page Citation Edition Citation Conference Place Citation Conference Date		OpenAIRE Guidelines for Literature Repository Managers v4
Alerting services, sharing to social networks, post-publication evaluation and commenting, support for multimedia and open peer review (where relevant)	EQSIP V1.0	

Unique URL for landing pages	DOAJ	
SEO	EQSIP V1.0	
URLs linking to related research objects	EQSIP V1.0	
Metadata exchange protocols (OAI-PMH, Application Programming Interfaces (APIs))	EQSIP V1.0, OpenAIRE Guidelines for Literature Repository Managers v4, Plan S	DOAJ Seal

Table 1: Technical standards and practices supporting Findability

The role of metadata for the findability of a resource, is universally recognised and emphasised in the documents analysed (DOAJ Seal, EQSIP, Plan S, OpenAIRE Guidelines for Literature Repository Managers v4). OpenAIRE Guidelines for Literature Repository Managers v4 detail the expected requirements and recommendations related to metadata which – together with other requirements – allow the content to be indexed by OpenAIRE metadata aggregator. There are already a number of tools which are OpenAIRE-compliant[12] – e.g. DSpace or OJS – which facilitate aggregation by OpenAIRE and through this also the uptake of scientific output by EOSC.

The role of PIDs is either required or recommended by e.g. Plan S, DOAJ Seal or EQSIP, with the emphasis on PIDs for authors, resources and funders. Finally, the technological features of publishing platforms for the findability of publications – e.g. data exchange protocols, SEO, URL configuration – are emphasised as a prerequisite for the findability of scientific output by, e.g. indexes or metadata aggregators.

4.2 A – Accessible

Users need to know how they can access the resources found. To achieve maximum accessibility, metadata needs to be retrievable through their IDs using a standardised communication protocol. This protocol should be openly available, free to use, and universally implemented. If required, such a protocol should include an authentication and authorization procedure.

To enable long-term accessibility to a resource its metadata needs to remain accessible even if the corresponding resource is no longer available.

[12] OpenAIRE compliance is described in OpenAIRE Guidelines (https://guidelines.openaire.eu/en/latest/)

Table 2 presents the technical standards and practices supporting Accessibility.

Standard/Requirement	Basic/Mandatory	Recommended
File Location	OpenAIRE Guidelines for Literature Repository Managers v4	
Metadata about OA status	Plan S	DOAJ Seal
Access Rights	OpenAIRE Guidelines for Literature Repository Managers v4	
Deposited in a digital preservation service	EQSIP V1.0, Plan S	DOAJ Seal
Registering of self-archiving policy		Plan S, DOAJ Seal
Direct deposition in OA repository		Plan S, DOAJ Seal

Table 2: Technical standards and practices supporting Accessibility

Information on accessibility should be available in the metadata. The OpenAIRE Guidelines for Literature Repository Managers v4 require the element "file location" to enable this. Plan S requires "Machine-readable information on the Open Access status and the licence embedded in the article, in standard non-proprietary format." (cOAlition S 2019). DOAJ requires journals to "(...) display an open access statement indicating that it fulfils the DOAJ definition of open access." (DOAJ, n.d.). Embedding the licence in the journal article's full text is only a requirement for the DOAJ Seal (DOAJ, 2023). These requirements are also in-line with the oai_dc element "Access Right" required by the OpenAIRE Guidelines for Literature Repository Managers v4.

Articles should not only be accessible in the short-run, but also long-term. Therefore, Plan S requires "deposition of content with a long-term digital preservation or archiving programme". EQSIP also emphasises this aspect: "The published content is deposited in a digital preservation service (Lots of Copies Keep Stuff Save (LOCKSS), CLOCKSS, Portico, Internet Archive, national libraries and other public preservation services etc.)." (EQSIPv1 2023, p.13). DOAJ has a similar requirement for its Seal. Clear policies on self-archiving/deposit enhance the accessibility of scholarly information. Furthermore, automated workflows to deliver metadata and full text into institutional or subject-specific repositories increase both accessibility for users searching for scholarly information and findability.

4.3 I – Interoperable

Interoperability of technical systems is a prerequisite for optimal machine-based communication. It takes a seamless multidirectional exchange of metadata to unfold the full potential of institutional publishing platforms and support value-added services. Interoperability heavily relies on standards that ensure sustainability and can be adapted to recent development trends. Metadata should use a formal, accessible, shared, and widely applicable language for knowledge representation. Metadata should adopt vocabularies that adhere to the FAIR principles and incorporate qualified references to other metadata.

With regard to technical standards, open specifications are recommended. In this respect, the EOSC Interoperability Framework (Corcho *et al.*, 2021) identifies the standards and guidelines that systems must follow to enhance the ability of users to connect services in more powerful, efficient and functional combinations. This framework includes the Resource description framework, the Metadata framework, the Authentication and Authorisation framework and many others.

Table 3 presents the technical standards and practices supporting interoperability.

Standard/Requirement	Basic/Mandatory	Recommended
Metadata exchange protocols (OAI-PMH, APIs)	EQSIP V1.0	DOAJ Seal
OpenAIRE Guidelines	EQSIP V1.0	Plan S
Human- and machine-readable information about the open access status, copyright holder and licensing is provided in each publication in a standard in a non-proprietary format	EQSIP V1.0	DOAJ Seal, Plan S
Knowledge Base And Related Tools (KBART)	EQSIP V1.0	
Counting Online Usage of NeTworked Electronic Resources (COUNTER)	EQSIP V1.0	
Metadata under CC Zero (CC0): reuse guaranteed	Plan S	
Text and Data Mining is	EQSIP V1.0	

technically supported		
Open Citations standards compliance	EQSIP V1.0	Plan S
Mass metadata export (as Comma-Separated Values files (CSV), ONIX[13] XML[14] feeds or in any other established format)	EQSIP V1.0	
Metadata exchange protocols (OAI-PMH, APIs)	EQSIP V1.0, OpenAIRE Guidelines for Literature Repository Managers v4, Plan S	
Full text in machine-readable format		Plan S

Table 3: Technical standards and practices supporting Interoperability

There are two crucial dimensions of interoperability:

1. proper representation of publications' metadata (universally required or recommended by the documents analysed by us), including usage data (see COUNTER); and
2. interoperability of the contents (full-text interoperability recommended by Plan S).

The key recommendations are:

- Proper implementations of metadata formats and schemas to facilitate interoperability with different services dedicated to aggregation and indexing of publications
- Standardised data exchange protocols
- Use of text markup for machine-readability of contents

Interoperability standards are related to many standards related to findability, accessibility and reusability. This demonstrates that certain standards are key to different aspects of the publishing ecosystem and that publishers should pay closer attention to them.

[13] Online Information Exchange
[14] Extensible Markup Language

4.4 R – Reusable

The reusability principle refers to the ability of a resource to be reused by a wide range of stakeholders. According to the "EOSC Model Specification and Guidelines" (https://doi.org/10.15497/rda00050), the principle involves using existing resources in alternative contexts or for different purposes, which may include republishing and creating derivative works. To facilitate this, the resources intended for such purposes should be accurately described and carefully curated.

For example, the provision of rich, high-quality metadata that are described with accurate and relevant attributes is crucial for achieving content reusability. Metadata must encompass licensing information in a standard, machine-readable format, as well as details on the provenance of the resource. In addition, the resource and accompanying metadata should conform to a machine-understandable community standard.

The principle of reusability aligns closely with the principle of findability, especially when describing research data and other scholarly outputs with rich metadata, as this contributes to content discovery and subsequent reusability. The stated objectives for achieving reusability are summarised in the technical standards listed in the table below.

Table 4 presents the technical standards and practices supporting Reusability.

Standard/Requirement	Basic/Mandatory	Recommended
Metadata under CC0	Plan S	
Licence Condition		OpenAIRE Guidelines for Literature Repository Managers v4
A human- and machine-readable information about the open access status, copyright holder, and licensing in a non-proprietary format	EQSIP V1.0, Plan S, DOAJ	DOAJ
Full text in machine-readable format (Journal Article Tag Suite XML (JATS XML) or equivalent (e.g. TEI)	EQSIP V1.0, DOAJ	Plan S
Open Citations standards compliance	EQSIP V1.0	Plan S

Description	OpenAIRE Guidelines for Literature Repository Managers v4	
Format		OpenAIRE Guidelines for Literature Repository Managers v4
Resource Type	OpenAIRE Guidelines for Literature Repository Managers v4	
Coverage		OpenAIRE Guidelines for Literature Repository Managers v4
Size		OpenAIRE Guidelines for Literature Repository Managers v4 (optional)
Geo Location		OpenAIRE Guidelines for Literature Repository Managers v4 (optional)
Audience		OpenAIRE Guidelines for Literature Repository Managers v4 (optional)

Table 4: Technical standards and practices supporting Reusability

The selected standards clearly emphasise the use of machine-readable standards and worldwide, royalty-free, non-exclusive, irrevocable licences that are also machine-readable (Plan S). It is important that these licences apply not only to the resource itself but also to its metadata, supplementary materials, references and citations. These licences facilitate the sharing and adaptation of the resource while ensuring proper attribution is given to the author(s) and contributors. Plan S recommends the default use of the Creative Commons Attribution (CC BY) 4.0 licence.

Furthermore, the principle of reusability dictates that metadata should include information about the origin of the specific output type (e.g. research data), such as the location or method of research data collection or generation. This is especially applicable in publications that publish data papers, i.e. papers that describe datasets and are published in a form of scholarly article. For instance, such information can take the form of a formal document that outlines how research data will be collected, organised, documented, stored, shared, and preserved throughout a research project or study (Data Management Plan), or an informal

description of the workflow that resulted in the resource, providing specific details regarding the individuals or entities responsible for its generation or collection, along with information on the processing steps that were carried out and the software used (code). Provenance should be represented and generated in different systems and contexts, such as the World Wide Web Consortium Provenance Family of Specifications Onthology (W3C PROV-O)[15].

Finally, machine-readable formats and the utilisation of cross-domain languages is essential to maximise the usability of metadata. Examples of machine-readable formats include XML structured in accordance with the Journal Article Tag Suite (JATS) standard (https://jats.nlm.nih.gov/) for semantic and structured article formatting or the Text Encoding Initiative (TEI) standard (https://tei-c.org/). Open formats not only facilitate metadata search but also enable content mining and reference mining, which then leads to the reuse of the resource itself and related materials.

[15] https://www.w3.org/TR/prov-o/

5 OTHER RECOMMENDED TECHNICAL STANDARDS

This section includes technical interoperability standards that we deem important but that do not naturally fall directly under any specific FAIR principle and other technical best practices. These are presented in Table 5.

Category	Standard	Basic	Recommended
Long-term preservation	Versioning	Plan S	
Metadata	Conflict of Interest statement in the metadata/ CoI in JATS	EQSIP V1.0	
Metadata	Machine-Readable Cataloguing (MARC) records to libraries	EQSIP V1.0	
Metadata	Contributor Roles Taxonomy (CRediT) in JATS	EQSIP V1.0	
Outputs/long-term preservation	Full text content is tagged in the XML JATS or equivalent (e.g. TEI) format and provided in multiple digital formats (Portable Document Format (PDF), HTML, XML, ePub, etc.), at least one of which is suitable for preservation.	EQSIP V1.0	DOAJ Seal
Platform features	article/chapter-level metrics, such as visits, views, downloads, citations	EQSIP V1.0	
Platform features	publication-level metrics	EQSIP V1.0	
Platform features	Tables of contents or structures that allow direct access to articles/chapters in as few clicks as possible are provided	EQSIP V1.0	
Platform features	publishing infrastructure: maintained, regularly backed up and protected from viruses and malware	EQSIP V1.0, DOAJ	
Platform features	mailing lists, content alerts, notifications, Really Simply Syndication (RSS)/Atom feed or	EQSIP V1.0	

	other mechanism		
Platform features	geographical spread of visitors	EQSIP V1.0	
Platform features	unique URL for landing pages	DOAJ	
Platform features	searching, browsing, navigation	EQSIP V1.0	
Platform features	support for both individual and bulk uploads of manuscripts		Plan S
Platform features	high resolution figures and well-constructed tables	EQSIP V1.0	
Platform features	analytics software and methods used to generate and collect metrics	EQSIP V1.0	
Platform features	accessibility guidelines (e.g. W3C Web Content Accessibility Guidelines - WCAG)	EQSIP V1.0	

Table 5: Other recommended technical standards and practices

Most of the technical standards that do not fall directly into any of the FAIR principles categories are related to platform features. These standards are recommended to help the user of the resource better understand, navigate, search and view the content. For example, the addition of Altmetrics and analytics software to generate and collect metrics are basic standards encouraged in EQSIP V1.0. One aspect which is not reflected widely in the documents analysed (only in ESQIP v1 2023) is the Web Content Accessibility Guidelines (WCAG). While web accessibility aims for barrier-free access for humans, the WCAG consists of various technical guidelines, e.g. to support assistive technologies such as screen readers.

We also found some metadata-related standards that are important to highlight, such as MARC[16] records to libraries, CRediT[17] in JATS[18] and Conflict of Interest documentation, all of which form part of the basic recommendations of EQSIP V1.0.

EQSIP V1.0 specifies full-text content tagged in the XML JATS or equivalent (e.g. TEI) format and provided in multiple digital formats (PDF, HTML, XML, ePub, etc.), at least one of which is suitable for preservation.

[16] https://www.loc.gov/standards/marcxml/
[17] https://credit.niso.org/
[18] https://jats.nlm.nih.gov/

5.1 Academic search engines and other discovery platforms

While requirements of academic search engines and discovery platforms were not central to the analysis presented, they play an important role in the discoverability of diamond OA journals. We observed several overlaps in requirements, especially for those platforms focusing on the discoverability of open content. The following list is not complete, but presents prominent examples of search engines and discovery platforms.

Google Scholar[19] is probably the best-known search engine for academic content. It suggests the inclusion of certain metatags for search engine optimization (Google Scholar, n.d.). However, there are also other academic search engines which are very relevant.

BASE[20] is a non-commercial endeavour and indexes more than 300 million documents from more than 10,000 content providers. 60% of its content is available open access (BASE, (n.d.)a). To become indexed in BASE, three basic criteria must be met: (1) the source has to contain academic content; (2) at least some documents from the source are available as open access (full texts free of charge, without registration); (3) the metadata of the documents are provided via a valid OAI-PMH interface. To enable indexing, content providers need to provide DC metadata (oai_dc) at the OAI-PMH. Oai_dc metadata is also requested by OpenAIRE (compare tables 1-4). BASE recommends that each record of a content provider's "OAI interface should have metadata for a document that is as complete as possible and use standardised vocabularies" (BASE, (n.d.)b). It emphasises that the more complete metadata is provided, the easier it will be to find documents from the respective source in BASE.

Another discovery service worth mentioning is **GoTriple**[21]. Developed and maintained by OPERAS, GoTriple is a multilingual discovery platform for Social Sciences and Humanities (SSH). It indexes publications, datasets, authors, and projects in this area, providing semantic enrichments in 11 languages and a set of additional services. GoTriple collects metadata through OAI-PMH and database dumps; metadata respects a data model aligned with major aggregators (OpenAIRE, BASE) and is structured according to the schema.org ontology. To date, GoTriple references more than 10 million documents.

Another important player is the research information database **Dimensions**[22]. Dimensions aims to make connections between research information (datasets, grants, publications,

[19] https://scholar.google.de/
[20] https://www.base-search.net/
[21] https://www.gotriple.eu/
[22] https://www.dimensions.ai/

patents, policy documents, etc.) explorable. For its publication metadata, Dimensions crawls Crossref and PubMed and retrieves the OA status from Unpaywall.

According to its website, **CORE**[23] offers users access to the most extensive compilation of open access research papers available globally. It diligently collects and indexes research materials from various repositories and academic journals. At present, CORE extensively harvests research papers from diverse sources, including institutional and subject repositories, as well as open access and hybrid journals. CORE's collection comprises a staggering 266 million open access articles, sourced from a network of 11,000 data providers. CORE uses the Open Archives Initiative (OAI) base URLs for harvesting.

The Lens[24] is an online search platform for both patents and scholarly literature. Renowned for its extensive scope, The Lens has surpassed the combined breadth and depth of two prominent commercial databases, namely Web of Science and Scopus. Operating as an agglomeration database, The Lens consolidates bibliometric data from various sources, including PubMed and Crossref, into a single unified platform with deduplicated content and a consistent search syntax.

It should be mentioned that many of the mentioned services populate their indexes by:

1. Ingesting data from registries and other indexes using APIs (such as Crossref[25], Datacite[26], Pubmed[27], OpenAlex[28], Semantic Scholar[29], etc.)
2. Harvesting repositories using OAI-PMH
3. Ingesting data from publisher platforms
4. Web indexing.

Hence, diamond journals, publishers, and platforms can gain visibility by providing appropriate metadata to the search service or its sources. For instance, there are many search services that by default ingest all scholarly publications metadata in Crossref. Registering DOIs with rich metadata then guarantees visibility via those search engines.

[23] https://core.ac.uk/
[24] https://www.lens.org/
[25] https://www.crossref.org/
[26] https://datacite.org/
[27] https://pubmed.ncbi.nlm.nih.gov/
[28] https://openalex.org/
[29] https://www.semanticscholar.org/

6 EXAMPLES ILLUSTRATING SEVERAL OR ALL OF THE TECHNICAL STANDARDS AND BEST PUBLISHING PRACTICES

In the following section, we showcase different examples of the implementation of the technical standards/requirements discussed: two diamond OA journals, and three publishing platforms run by members of the CRAFT-OA consortium. Each of these examples emphasises different standards/requirements (DOAJ, Plan S, FAIR principles.). The examples are meant as good practice case studies to showcase how the discussed technical standards are implemented in practice. For example, one can observe that the use of journal management systems, such as OJS or Lodel, support the implementation of many standards very well. A second purpose of this section is to present different ways of approaching the topic of technical standards: For the purpose of getting indexed in the DOAJ, as guiding principles when setting up a new service, as review of the own performance and as self-assessment in the light of the FAIRifcation of own processes.

6.1 Open Journal of Mathematical Optimization

OJMO[30] is an online journal of computer science and mathematics owned by its Editorial Board (scholar-led). It publishes high-quality articles in open access free of charge, meaning that neither the authors nor the readers have to pay to publish, or to access the content of the published papers. OJMO has recently been indexed in DOAJ. It also meets the extra quality requirements recommended by DOAJ to achieve the DOAJ Seal.

A summary table for *Open Journal of Mathematical Optimization* is provided in Table 6.

Digital preservation The journal content must be continuously deposited in one of these archives: ● any archiving agency included in Keepers Registry ● Internet Archive ● PubMed Central	https://portal.issn.org/resource/ISSN/2777-5860
Persistent article IDs Articles must use persistent article IDs. DOI, ARK or Handle are the most commonly used. All persistent links must resolve correct	This journal uses DOIs: https://ojmo.centre-mersenne.org/articles/10.5802/ojmo.1/

[30] https://ojmo.centre-mersenne.org/

Metadata supply to DOAJ Article metadata must be uploaded to DOAJ regularly.	https://doaj.org/toc/2777-5860
Licence type The journal must permit the use of a Creative Commons (CC) licence that allows the creation of derivative products. ● CC BY ● CC Attribution-ShareAlike License (CC BY-SA) ● CC Attribution-NonCommercial License (CC BY-NC) ● CC Attribution-NonCommercial-ShareAlike License (CC BY-NC-SA)	All content is licensed under the Creative Commons Attribution 4.0 International License so that interested researchers are free to remix, transform, and build upon the material for any purpose. https://ojmo.centre-mersenne.org/page/policy_en/
Licence information in articles CC licensing information must be displayed in all full text article formats.	https://ojmo.centre-mersenne.org/articles/10.5802/ojmo.22/
Copyright and publishing rights Authors must retain unrestricted copyright and all publishing rights when publishing under any licence permitted by the journal.	The author retains unrestricted copyrights and publishing rights. https://ojmo.centre-mersenne.org/page/policy_en/
Self-archiving policy Authors must be permitted to deposit all versions of their paper in an institutional or subject repository. ● Preprint ● Author's Accepted Manuscript ● Published article (Version of Record) ● An embargo may not be applied.	The authors of a published paper are free to deposit a copy of all versions of their paper in an institutional or other repository of their choice, including the submitted version, the accepted version, and the published version. https://ojmo.centre-mersenne.org/page/policy_en/

Table 6: Summary table for Open Journal of Mathematical Optimization

6.2 ARHEOLOŠKI VESTNIK

Arheološki vestnik [Archaeological Journal][31] was founded in 1950 as a Slovenian archaeological journal with an interdisciplinary and international character. It is published by Založba ZRC[32] (and co-published by the Slovenian Academy of Sciences and Arts). The journal uses the open-source software Open Journal System (OJS) and is indexed in DOAJ. All articles are published as diamond open access.

A summary table for *Arheološki vestnik* is provided in Table 7.

Digital preservation The journal content is continuously deposited in:	National and university library (dLib)
Persistent article IDs Articles must use persistent article IDs. All persistent links must resolve correct	This journal uses DOIs and ORCiD IDs. https://ojs.zrc-sazu.si/av/article/view/10937
Metadata supply to DOAJ	Article metadata are uploaded to DOAJ regularly.
Licence type The journal must permit the use of a CC licence that allows the creation of derivative products.	This journal uses a licence CC BY-NC-SA. https://ojs.zrc-sazu.si/av/prispevki
Licence information in articles CC licensing information must be displayed in all full text article formats.	https://ojs.zrc-sazu.si/av/article/view/10937 and in PDF: https://ojs.zrc-sazu.si/av/article/view/10937/10119
Copyright and publishing rights Authors must retain unrestricted copyright and all publishing rights when publishing under any licence permitted by the journal.	The author retains unrestricted copyrights and publishing rights. https://ojs.zrc-sazu.si/av/prispevki

[31] https://ojs.zrc-sazu.si/av/index
[32] https://zalozba.zrc-sazu.si/en/predstavitev

Self-archiving policy Authors must be permitted to deposit their paper in an institutional or subject repository. ● Published article (Version of Record) ● An embargo is not applied.	The authors of a published paper are free to deposit a copy of their paper in an institutional or other repository of their choice (the published version). https://ojs.zrc-sazu.si/av/prispevki

Table 7: Summary table for Arheološki vestnik

6.3 TIB Open Publishing

TIB Open Publishing[33] is a publishing platform for diamond open access journals and conference proceedings based at TIB, and is open to all institutions/editors and academic disciplines. It uses the open-source software Open Journal Systems (OJS) to support the publishing workflow from submission to publication. While editorial control over the publications is with the editors of each publication (scholar-led approach), TIB Open Publishing takes care that editorial best practices are implemented, publishing standards are met, and technical workflows are standardised. When setting up the service, TIB reviewed various guidelines and recommendations (namely the DOAJ (Seal) criteria, the OASPA[34] membership criteria, the Plan S technical implementation guidelines, as well as the Commitee on Publication Ethics guidelines (COPE[35]) and used them as guiding principles.

A summary for TIB Open Publishing is provided in Table 8.

Requirement/recommendation	DOAJ	Plan S	TIB Open Publishing example
Open Access statement	x		Publication level statement, e.g. https://www.tib-op.org/ojs/index.php/scp/about
Open licence	x	x	All articles are published under CC BY (3.0 DE or 4.0 International)
ISSN	x		All publications have an ISSN (online)
Embedded licences	x	x	Licences are embedded in all article formats (landing page, PDF, XML/HTML), see e.g.

[33] https://www.tib-op.org/ojs/index.php/index/index
[34] https://oaspa.org/
[35] https://publicationethics.org/

			https://www.tib-op.org/ojs/index.php/bis/article/view/52 and machine-readable in the metadata
Copyright	x		Authors retain copyrights and full publishing rights, see e.g. https://www.tib-op.org/ojs/index.php/th-wildau-ensp/about
Plagiarism	x		All manuscripts submitted to a publications on TIB Open Publishing are scanned for plagiarism (Similarity Check), https://www.tib-op.org/ojs/index.php/index/tibop/infrastructureandtechnology
Unique IDs & structured data	x	x	All articles receive a DOI and provide ORCID iDs in their metadata, e.g. https://www.tib-op.org/ojs/index.php/ocp/article/view/78. In addition RORs are provided.
Initiative for Open Citations (I4OC) standards	x	x	TIB Open Publishing delivers reference lists to Crossref
Long-term archiving	x	x	All contents are archived by the German National Library and the TIB archiving system/or Portico
Metadata in CC0		x	Metadata is provided in a non-proprietary format at an OAI-PMH interface, e.g. https://www.tib-op.org/ojs/index.php/ocp/oai
Machine-readable full text		x	TIB Open Publishing (partly) provides full text XML (rendered as HTML)

OpenAIRE compatible metadata		x	DC metadata is provided at OAI-PMH interface, e.g. https://www.tib-op.org/ojs/index.php/ocp/oai

Table 8: Summary table for TIB Open Publishing

6.4 HRČAK - Portal of Croatian scientific and professional journals

HRČAK[36] is a central portal of more than 530 Croatian scientific and professional journals that publish in open access. Besides its main function as an OA journal publishing portal, HRČAK also provides journal editors with the ability to use a maintained instance of Open Journal Systems (OJS). The technical maintenance of HRČAK is carried out by the University of Zagreb University Computing Centre (SRCE). The editors of the individual journals hold the responsibility for the content within their respective journals. HRČAK was established and continues to develop in collaboration with several institutions and experts in Croatia. It aims to meet the needs of the community while also keeping up with the trends of modern scientific publishing. The technical standards implemented on HRČAK are listed in the table below and can be observed in the example of the journal Food Technology and Biotechnology.[37]

Table 9 provides a summary of HRČAK and the functionality that supports the FAIR principles.

Recommendation/requirement	Description	FAIR
PID	The use of DOI for objects and linking authors with their ORCID IDs is supported and strongly recommended.	F
SEO	Highwire Press HTML meta-tags are integrated into each object for the purpose of indexing in Google Scholar.	F
Unique URL for landing pages	Each article in HRČAK has its own, unique URI that serves as its landing page.	F
Link between metadata and research data	There is an option to link the article and associated research data published in a different location.	F

[36] https://hrcak.srce.hr/en
[37] https://hrcak.srce.hr/en/ftb

Standardised protocol for metadata harvesting	Metadata can be accessed and harvested via OAI-PMH protocol.[38] The list of HRČAK journals is available in KBART compliant format.[39]	A
Accessibility guidelines	HRČAK has ensured the accessibility of its content in accordance with the Digital Accessibility Guidelines - Croatian Academic and Research Network (CARNET)[40] that, among other resources, refer to the WCAG Guidelines.	A
OpenAIRE Guidelines for Literature Repositories v3 compatibility	Metadata are made available through the OAI-PMH interface using the DC metadata schema.	I
Metadata exchange protocols	HRČAK has implemented the OAI-PMH interface that supports two metadata schemas: DC and more detailed Metadata Object Description Scheme (MODS). Journals can transmit their metadata to the DOAJ through the DOAJ API.	I
Information about the reuse licence	The licence must be specified in the imprint, author guidelines and each article (PDF and/or JATS XML version). HRČAK strongly encourages the usage of CC licences.	R
Machine-readable full text	The support for publishing full texts in the JATS XML format is provided.	R

Table 9: Summary table for HRČAK

[38] https://hrcak.srce.hr/oai/?verb=ListRecords&metadataPrefix=oai_dc&set=journal:185
[39] https://hrcak.srce.hr/en/interoperabilnost
[40] https://www.carnet.hr/en/accessibility/

6.5 OpenEdition's journals FAIR assessment

OpenEdition is a French research infrastructure maintaining four publishing platforms for the Social Sciences and Humanities. The OpenEdition Journals platform hosts more than 600 diamond journals, providing open access to at least the HTML version of the contents. The application criteria for journals ensure both the scientific and publishing quality of the publications and their compliance with open access principles. Through the use of its publishing CMS, Lodel, the infrastructure can generate standard metadata in various formats that ensure referencing of the publications on major aggregators and catalogues.

More specifically, the OpenEdition team assessed the FAIR compliance of all its data, considering the published contents and their metadata, but also the controlled vocabularies and TEI corpora handled by the infrastructure. Regarding the OpenEdition Journals platform, the objective was both to evaluate and to improve the compliance level of the publications with the FAIR principles. Using the FAIR principles as an analytical grid helped to confirm the technical quality of journals' content management and, more importantly, to identify the areas of potential improvement.

The work conducted by OpenEdition's[41] team helped to define a methodology for FAIR assessment that could be applied by other publishing platforms.

Table 10 provides a summary of OpenEdition and the functionality that supports the FAIR principles.

FAIR review of OpenEdition Journals	
Data summary	
Data sources	Data produced by publishers and users Can be updated (not fully controlled by the organisation) Documentary units' distinct levels: text, issue and collection levels
Data expressions	Raw data: database (used for the HTML version) Other expressions: TEI OpenEdition, PDF, ePub Metadata

[41] The work and the methodology are described in: Avanço, K. & Gingold, A., (2022) "FAIRifying a scholarly publishing service: Methodology based on the OpenEdition's internal FAIR audit", The Journal of Electronic Publishing 25(2). doi: https://doi.org/10.3998/jep.1540.

Commentary	Different properties depending on the type: - Volume contains: Publications (issues, columns, annual columns); Documentary unit contains texts, - Different types of texts (article, column, editorial, review, ...), - Attached file types can contain data (Excel Spreadsheet (XLS), CSV, sound, image, video files), Not all the different types are available in all the different expressions (TEI, PDF, ePub) Question: Should the review consider the types that don't correspond to specific content (subpart, section, site, directory, etc.)?		
Findable	**FAIR components**	**FAIR enabling**	**FAIR prospects**
F1. Metadata are assigned a globally unique and PID	Objects: - DOI: available only for some data (depending on the types and publishers' wishes) - Issue for resources with multiple DOIs - Handles generated by Isidore harvesting platform (not retrieved by OE) Persons: ORCiD Organisations: a few IDs from Crossref Funding registry.	- OAI IDs exist for all documentary units but are not PIDs - All documentary units are identified in the information system through the concatenation: Platform+SiteName+ID	Objects: - Some data without any PID - DOIs may exist for data published on another platform that we do not retrieve. Persons: Authors and contributors connected to authority files
F2. Data are described with rich metadata (defined by R1 below)	- Metadata available in the OAI-PMH repository - Formats: DC, DC Terms, Metadata Encoding and Transmission Standard (METS)	Rich metadata is available; could be extensively integrated into the OAI repository. In OAI, metadata is available only for certain	

		types (subpart, heading, and news are missing)	
F3. Metadata clearly and explicitly includes the ID of the data they describe	In the OAI repository: - ID from OAI - DOI when available		Some data without any PID (see F1)
F4. Metadata are registered or indexed in a searchable resource	OpenEdition Search interface (search.openedition.org): - only a selection of data is available (some types are excluded) - metadata are not complete on the website. Metadata is also searchable in other directories (e.g. Isidore harvests OE's OAI repository)	No public API is available yet, but all the information is available through the search software (SolR)	
Accessible	**FAIR components**	**FAIR enabling**	**FAIR prospects**
A1. Metadata are retrievable by their ID using a standardised communications protocol	HTML: accessible via the DOI Metadata: accessible via the OAI ID		Some data without any PID (see F1)
A1.1 The protocol is open, free, and universally implementable	Hypertext Transfer Protocol (HTTP) for the data OAI-PMH for the metadata		

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary	All protocols are open, but not all allow for authentication. Protocol used for restricted access contents: - Transmission Control Protocol/Internet Protocol (TCP/IP) for contents requiring authentication (TEI version's case) Other protocols used where authentication is not required: - HTTP for open access contents - OAI-PMH for the metadata		Lack of a tool dedicated to the management of authentication and authorization processes.
A2. Metadata are accessible, even when the data are no longer available	No		No records for the deleted data.
Interoperable	**FAIR components**	**FAIR enabling**	**FAIR prospects**
I1. Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.	TEI, DC, METS		No semantic layer is implemented.
I2. Metadata use vocabularies that follow FAIR principles	In some journals, use of disciplinary controlled vocabularies (e.g. French Pactols).	Some disciplinary controlled vocabularies (Journal of Economic Literature Classification System (JEL),	For most of the journals, no controlled vocabulary is used.

		GeographieUN) could be integrated with thesaurus management tools	
I3. Metadata include qualified references to other metadata	In OAI repository: - is part of - relation with OpenAIRE access right field Some links with translations	- Citation and Cited-by available but not disseminated - Link with translations not recorded in the OAI repository - ongoing project: OE Review of Books	
Reusable	**FAIR components**	**FAIR enabling**	**FAIR prospects**
R1.1. Metadata are released with a clear and accessible data usage licence	All journals have a CC licence. By default licence: CC BY-SA		No clear provision to allow for the text and data mining exception (acknowledged by French law “Loi pour une république numérique”)
R1.2. Metadata are associated with detailed provenance	Internal creation process not described (can be created through Lodel, outsourced digitisation, etc.)		
R1.3. Metadata meet domain-relevant community standards	I1: metadata meet community standards for textual contents, including TEI. I2: Fewer metadata meet disciplinary		

	communities standards (SSH disciplines' vocabularies).		

Table 10: Summary table for OpenEdition Conclusions

This report has provided an overview of the current state of scholarly publishing practices and technical requirements in the context of FAIR principles. The report highlights the importance of interoperability to enable discoverability, reuse, and reproducibility of research outputs.
In addition to creating an initial connection between scholarly publishing practices and the technical requirements of the FAIR principles, this is (as far as we know) the first attempt to systematically collect and compare the different requirements set by the selected policies and services with each other. From the perspective of a publisher, it would be desirable for the requirements set by different actors to be aligned (so as not to be incompatible with each other), and offer some degree of progression in compliance and implementation so that it is not a matter of all or nothing. This is particularly relevant for the requirements set by DOAJ and cOAlition S, which are essential for most OA journals to fulfil. The requirements criteria set by both of these organisations include both basic and recommended levels. Based on our review, we found that they are well-aligned. If a journal fulfils the requirements of one, it will fulfil a number of requirements of the other.

We identified and showcased five examples of journals and platforms that comply with several of the compiled technical standards described in the report. The journals and platforms presented are only examples of how technical standards can be implemented, and many more journals and platforms exist that adhere to the various standards. The case studies we presented applied different benchmarking approaches, so we could shed light from different angles on the implementation of technical standards: Both journals presented achieved DOAJ indexing. Thus they can serve as "role models for journals wanting to be indexed in DOAJ. For the three platforms showcased, we focussed on different aspects of the analysed standards: compliance with DOAJ and Plan S (TIB Open Publishing), technical standards to support overall interoperability and the FAIR principles (HRČAK) and the FAIR review of OpenEdition.

However, we are aware that there are journals and platforms which struggle to fulfil such requirements, so the next tasks will try to identify them *(T.3.2. Challenge and gap identification for OA journals and platforms to comply with standards)* and provide support *(T.3.3 Training and education to enable the adoption and implementation of technical specifications)*. It is important to raise awareness that technical standards and their implementation are needed to fulfil crucial functionalities of open access (in terms of reusability, visibility, availability, etc.); and WP3 will provide information about existing

technical solutions on journal platforms that enable journals to implement technical specifications.

As an outcome of the work conducted for this report, we suggest that publishers conduct regular audits to assess their compliance with the technical standards we presented along the FAIR principles and identify areas for improvement. The digital environment continues to evolve and mature, and publishers need to periodically reflect on their current position in relation to current possibilities and external requirements. Looking at the path ahead, the report highlights the need for collaboration among stakeholders, including publishers, to continue to create, refine, and align technical and process requirements so that individual articles, as well as large publisher portfolios of journals, are equally accessible and discoverable. Collaboration can also solve some friction generated by continuously changing requirements that might be seen as frustrating. There should not be requirements set that cannot be implemented with reasonable effort by publishers regardless of size, so open source software development of publishing software (e.g. OJS) and work that goes into formulating and modifying requirements should preferably be done in dialogue.

This report provides valuable guidance for publishers wishing to align their practices with FAIR principles and contribute to a more open, transparent, and trustworthy scholarly communication ecosystem. By adopting these standards and best practices, publishers can help ensure that research outputs are findable by peers and the interested public, accessible to all stakeholders, interoperable with various services, and reusable in new research endeavours or policy decisions.

7 REFERENCES

All references and websites mentioned in the document were last checked for availability on 30.06.2023.

Ancion, Zoé, Borrell-Damián, Lidia, Mounier, Pierre, Rooryck, Johan, & Saenen, Bregt. (2022). *Action Plan for Diamond Open Access*. Zenodo. https://doi.org/10.5281/zenodo.6282403

Armengou, C., Redhead, C. & Rooryck, J. (2023). *Extensible Quality Standards in Institutional Publishing V.1.0 (EQSIP V1.0)*. https://zenodo.org/record/7923916#.ZFy866DP02x

BASE (n.d.)a.; https://doi.org/10.5281/zenodo.6282403

BASE (n.d.)b.; https://doi.org/10.5281/zenodo.6282403

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

cOAlition S (2019). https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/

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 Vries (2005). IT Standards Typology. In: Jakobs K. (Ed.) *Advanced Topics in Information Technology Standards and Standardization Research Volume 1*, Hershey, PA, USA, Idea Group Publishing, 11-36.

Council of the European Union (2023). *Conclusions on high-quality, transparent, open, trustworthy and equitable scholarly publishing*. 23rd of May 2023. https://data.consilium.europa.eu/doc/document/ST-9616-2023-INIT/en/pdf

FAIR Data Maturity Model Working Group. (2020). *FAIR Data Maturity Model. Specification and Guidelines (1.0)*. https://doi.org/10.15497/rda00050

DOAJ (2023) https://doaj.org/apply/guide/#basic-criteria-for-inclusion

DOAJ (n. d.) https://doaj.org/apply/guide/

EOSC Future (n.d.)a. EOSC Interoperability Framework Governance (EOSC IF), (Accessed 4th of June 2023). https://wiki.eoscfuture.eu/pages/viewpage.action?pageId=24515805

EOSC Future (n.d.)b *EOSC IF Guidelines overview*, (Accessed 14th of June 2023). https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+IF+Guidelines+overview

Go-fair.org (n.d.) FAIR Principles. https://www.go-fair.org/fair-principles/

Google Scholar (n.d.) https://scholar.google.com/intl/de/scholar/inclusion.html#indexing

Jardim, S. V. B. & Martins, A. C. (2016). *An Overview and a Future Perspective in Health Information Systems in Portugal*. Encyclopedia of E-Health and Telemedicine, 987–997. https://doi.org/10.4018/978-1-4666-9978-6.ch07

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

NIFO (n.d) NIFO - National Interoperability Framework Observatory. Glossary. https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/glossary/term/interoperability#:~:text=Interoperability%20is%20a%20key%20factor,are%20understood%20by%20all%20parties.

NISO (2004). *Understanding Metadata*. Bethesda, MD: NISO Press, National Information Standards Organization.

OpenAIRE (2018). *OpenAIRE Guidelines for Literature Repository Managers v4*. https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/

8 List of Tables