



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

OPERAS

White Paper

Common Standards and FAIR Principles



This is the White Paper of the “**Common Standards and FAIR Principles**” Special Interest Group.

Starting date: July 2019

Delivered: June 2021

Participants:

CEON: Milica Ševkušić, Nikola Stanić

DH+ SIB: Claire Clivaz

EKT: Paraskevas Kamatsos, Marina Angelaki
(SIG Coordinators)

IBL PAN: Marta Błaszczewska, Cezary
Rosinski

Open Edition: Arnaud Gingold, Pierre

Mounier, Karla Avanzo, Chloé Lebon,

Panayota Polydoratos, Paulin Ribbe

PKP: Kevin Stranac

University of Bielefeld: Jochen Schirrwagen

University of Milano Bicocca: Federico

Pianzola

University of Turin: Elena Giglia

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



<https://doi.org/10.5281/zenodo.5653414>

Attribution 4.0 International



CONTENT

EXECUTIVE SUMMARY	6
1. INTRODUCTION	6
2. FRAMEWORK AND SCOPE	7
2.1 OPEN SCIENCE	7
2.2 MANDATES AND PRINCIPLES TO SUPPORT OPEN SCIENCE AT EU LEVEL	8
2.3 THE EUROPEAN OPEN SCIENCE CLOUD (EOSC)	10
2.4 COMMUNITIES ENGAGED IN SCHOLARLY COMMUNICATION	10
3. STATE OF THE ART	12
3.1 EMERGING TRENDS	13
3.2 INITIATIVES FOR THE IMPLEMENTATION OF STANDARDS	17
3.3 OPERATIONAL AND TECHNICAL STANDARDS FOR E-INFRASTRUCTURES	19
3.3.1. Suggested framework for the implementation of standards	19
3.3.2. Content quality	20
3.3.3 Findability and accessibility	21
3.3.4 Interoperability	22
3.3.4.1 Technical interoperability	22
3.3.4.2 Semantic interoperability	22
3.3.4.3 Organizational interoperability	23
3.3.4.4 Legal interoperability	23
3.3.5 Reusability	23
3.3.6. Processability	23
3.3.7. Impact assessment	24
4. FAIR PRINCIPLES	24
4.1 OVERVIEW	24
4.2 CONTEXT AND PROSPECTS	25
5. OPERAS CONTRIBUTION IN PROMOTING COMMON STANDARDS AND ORIENTATIONS FOR FUTURE WORK	27
5.1 PROJECTS AND INITIATIVES	27

5.1.1 HIRMEOS	27
5.1.2 TRIPLE	28
5.1.3 Diamond OA Study	29
5.2 FUTURE ORIENTATIONS FOR WORK	30
REFERENCES	33

EXECUTIVE SUMMARY

The White Paper “Common Standards and FAIR Principles” aims at exploring the workflows, mediums and technical standards that have emerged because of the changes brought about by the transition to Open Science. In the context of the work undertaken by the SIG, common standards and FAIR (findable, accessible, interoperable, and reusable) principles are examined as key operational and technical aspects that ensure content quality and interoperability for scholarly output in the social science and humanities (SSH) and beyond. A key activity during the relaunching of the SIG has focused on updating the 2018 White Paper, while identifying future actions/ tasks that SIG members (and other OPERAS members) could develop and support and exploring ways in which the SIG could feed-in the work of other OPERAS activities. To do so, we have taken into consideration developments since then and placed emphasis on the operational and technical aspects to be addressed by digital research infrastructures and service providers, as well as FAIR principles and the various projects that OPERAS members have been involved in.

The White Paper presents recent developments in realizing the transition towards the Open Science paradigm and the complementary roles of key stakeholders in this process like research funders, research institutions, infrastructure providers and researchers, while highlighting the role of standards. It then goes on to focus on technical and operational standards for digital research infrastructures and service providers by identifying four areas of importance: content quality and impact, interoperability, availability and processability. The FAIR principles are presented in the following section of the White Paper, while the final section presents some recent contributions of OPERAS in promoting common standards and proposes some orientations for future work for the SIG.

1. INTRODUCTION

The White Paper “Common Standards and FAIR Principles” builds on the work done by the Special Interest Group (SIG) “Common Standards” (renamed in 2020 “SIG Common Standards and FAIR Principles”). The SIG was launched in June 2017 under the name “Working Group” in the context of OPERAS.¹ The first output of the OPERAS (seven) Working Groups took the form of White Papers that were presented at the (first) OPERAS Conference in Athens in 2018. Following the creation of the OPERAS legal entity in March 2020, the Working Groups have been relaunched as “Special Interest Groups” (SIGs).² The SIGs constitute the place where discussions and new initiatives take place and emerge through collaborative work and the exchange of ideas among its members. Their work will at the same time constitute the pillars where the infrastructure will build its strategy and services.

The SIG “Common Standards and FAIR Principles” aims in particular at exploring the workflows, mediums and technical standards that have recently emerged as a result of the changes brought about by the transition to Open Science. In the context of the work undertaken by the SIG, common standards and FAIR (findable, accessible, interoperable, and reusable) principles, are examined as key operational and technical aspects that ensure content quality and interoperability for scholarly output in the social science

¹ OPERAS (<https://operas.hypotheses.org/>) is a European research infrastructure for the development of open scholarly communication, particularly in the Social Sciences and Humanities.

² For further information on OPERAS SIGs <https://www.operas-eu.org/special-interest-groups/>

and humanities (SSH) and beyond. A key activity during the relaunching of the SIG has focused on updating the 2018 White Paper, while identifying future actions/ tasks that SIG members (and other OPERAS members) could develop and support and exploring ways in which the SIG could feed-in the work of other OPERAS activities. To do so, we have taken into consideration developments since then and placed emphasis on the operational and technical aspects to be addressed by digital research infrastructures and service providers, as well as FAIR principles and the various projects that OPERAS members have been involved in.

The structure of the White Paper is as follows: the next section presents recent developments in realizing the transition towards the Open Science paradigm and the complementary roles of key stakeholders in this process like research funders, research institutions, infrastructure providers and researchers, while highlighting the role of standards. It then goes on to focus on technical and operational standards for digital research infrastructures and service providers by identifying four areas of importance: content quality and impact, interoperability, availability and processability. The FAIR principles are presented in the fourth section of the White Paper, while the final section presents some recent contributions of OPERAS in promoting common standards and proposes some orientations for future work for the SIG.

2. FRAMEWORK AND SCOPE

This section describes the recent developments in research and the complementary roles of researchers, funders, research institutions, infrastructure providers and the EU in realising the Open Science paradigm. In addition, it identifies existing and emerging challenges that stipulate the central role of e-infrastructures and the importance of standards in shaping a global communication framework for all communities engaged in research.

2.1 OPEN SCIENCE

Open Science represents a new approach to the scientific process that seeks to ensure that access to the entire life-cycle of research remains fundamentally open and replicable. This approach shifts the emphasis from the standard practices of publishing research results in the form of scientific publications towards sharing and using knowledge. At a more practical level, this new paradigm entails important and on-going transitions in the way research is performed, researchers collaborate, knowledge is shared and science is organized. The shift from open access (to publications and research data) to Open Science allows for a more encompassing approach that includes elements like open peer review, open methodologies, open educational resources, and other participatory processes like citizen science and brings to the forefront issues around the development of appropriate policies, infrastructures and standards for supporting this transition.

Within the EU, Open Science is one of the three goals for EU research and innovation policy summarized as “Open Innovation, Open Science and Open to the World”. The EU’s interest in supporting Open Science has been confirmed in Council Conclusions on the transition towards an Open Science system adopted on 27 May 2016. The Council acknowledged “that open science has the potential to increase the quality, impact and benefits of science and to accelerate advancement of knowledge” and called on the Commission, the Member States and the stakeholders to “take the necessary actions needed to making open science a reality and to advocate the need for concerted actions” (Council EU, 2016).

To support further the development of Open Science policy the Directorate General for Research and Innovation (DG RTD) set up in 2016 an Open Science Policy Platform (OSPP) a high level advisory group comprising 25 of the most important relevant stakeholders including *inter alia* research funding and research performing organisations, libraries, and scientific publication associations. The platform provided a forum for a structured discussion and gave advice to the Commission on the basis of the European Open Science agenda. The latter was structured around the following themes: 1) fostering and creating incentives for Open Science, 2) removing barriers for Open Science, 3) mainstreaming and further promoting open access policies, 4) developing research infrastructures for Open Science and 5) embedding Open Science in society as a socio-economic driver.

The actions were in turn translated into eight topics of policy concern, namely: rewards, altmetrics, Open Science Cloud, changing business models for publishing, research integrity, citizen science, open education and skills and FAIR open data. The OSPP final report was published in April 2020 and called on the member states to help co-create and maintain a “Research system based on shared knowledge by 2030” and identified five priorities³ (European Commission, 2020). The work of the OSPP has been further supported through the Open Science Monitor aimed at providing data and insight for the implementation of related policies. The Monitor’s final report was published in 2019 and includes policy recommendations for the future of open science and the monitor (OSM, 2019).

Late in 2019, UNESCO launched a process towards defining a global open science Recommendation as a legal instrument (to be adopted in November 2021). As a result of inclusive and transparent global consultations, a draft recommendation was presented in September 2020.⁴ The draft document highlights the importance of Open Science in *reducing the digital, technological, gender and knowledge divides* and calls for the transformation of the entire scientific process in order to *make science more open, accessible, efficient, democratic, and transparent*. The Recommendation is expected to define shared values and principles for Open Science, and identify concrete measures on Open Access and Open Data.

2.2 MANDATES AND PRINCIPLES TO SUPPORT OPEN SCIENCE AT EU LEVEL

The European Commission has been an active supporter of open access based on the notion that “there should be no need to pay for information funded from the public purse each time it is accessed or used”. Open access is expected to contribute to generating growth through greater efficiency, faster progress and improved transparency of the scientific process through the involvement of citizens and society. The benefits for researchers are associated with the positive impact on the visibility of research outputs and on the increase in usage and impact.

The support provided by the EU to open access has been affirmed in the 2012 Recommendation of the European Commission “on access to and preservation of scientific information” and its 2018 update and reaffirmed through the Council of the European Union Conclusions of May 2016. The Council recognized that “the exponential growth of data, the increasingly powerful digital technologies, together with the

³ The five priorities are the following: 1) an academic career structure that fosters outputs, practices and behaviours to maximise contributions to a shared research knowledge system, 2) a research system that is reliable, transparent and trustworthy, 3) a research system that enables innovation, 4) a research culture that facilitates diversity and equity of opportunity, 5) a research system that is built on evidenced-based policy and practice

⁴ The draft Recommendation can be accessed here:

<https://en.unesco.org/news/milestone-unescos-development-global-recommendation-open-science>

globalization of the scientific community and the increasing demand for addressing the societal challenges contribute to the ongoing transformation and the opening up of science and research which is referred to as “open science” (Council EU, 2016). It called on Member States, the Commission and stakeholders to remove financial and legal barriers and agreed to promote the mainstreaming of open access to publications by continuing to support a transition to immediate open access as the default by 2020. In 2019, a Directive was adopted on open data and the re-use of public sector information.⁵

Within Horizon 2020 Open Access was required (mandatory) for all peer-reviewed publications resulting from projects funded under the programme. This decision followed the pilot action on Open Access, which was implemented in FP7 for part of the funding period. Following the pilot action on Open Access to research data generated in Horizon 2020, the Commission extended the pilot to all thematic areas as stated in the 2017 Work Programme. Acknowledging that not all data can be open, the possibility of opting out (at any stage before or after signing the Grant) was provided. The Commission’s approach is therefore best described as “as open as possible as closed as necessary”. The open access mandate is translated into specific requirements in the Model Grant Agreement (articles 29.2 and 29.3) and in the H2020 work programme. In relation to research data, the European Commission also produced a set of guidelines on FAIR data management in Horizon 2020 to help beneficiaries make their research data findable, accessible, interoperable and reusable (FAIR) (European Commission, 2016). The Commission stresses the importance of Data Management Plans (DMPs) as key components of good data management and as such provides guidance to support researchers in developing their DMPs.

The open access mandate has been reaffirmed and strengthened further under Horizon Europe, the EU’s research and innovation framework programme running from 2021-2027. The draft Model Grant Agreement released in February 2021 stipulates immediate open access via repository under a CC BY license or equivalent of the author accepted manuscript (AAM) of version of record (VoR). Metadata of deposited publications must be under a CC 0 license or equivalent in line with FAIR principles. Beneficiaries must also manage their research data in line with the FAIR principles. To facilitate further researchers in complying with the open access mandate, the European Commission launched the Open Research Europe (ORE) platform (‘Open Research Europe: Open Access Publishing Platform. Beyond a Research Journal’ 2020).

Open access is also a priority of the European Research Area (ERA), namely Priority 5b: “Open access to publications and data in an open science context” (Priority 5 “Optimal circulation, access to and transfer of scientific knowledge”) and headline indicator 5b- “Open Access”. The 2018 ERA Progress Report, while acknowledging the progress made, stresses the need for further coordination and harmonisation across countries and points to the financial and technical challenges that need to be addressed as they are preventing the full transition to an open science environment (European Commission, 2019).

In 2018 a group of research funders announced the launch of an initiative (cOAlition S) to make immediate and full open access to research publications a reality through the implementation of a set of 10 principles of Plan S. Since then, the cOAlition has announced its Rights Retention Strategy and the Journal Checker Tool to support researchers in complying with the policy. In its Technical Guidance and Requirements, guidance is provided in terms of the mandatory technical requirements of publications venues (such as the use of persistent identifiers, high-quality article level metadata). A recent study on OA

⁵ Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information

<https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024>

Diamond journals showed that these are not yet fully compliant with the technical requirements set by Plan S and provided a set of recommendations and an action plan that would help in overcoming the challenges identified (Becerril et al., 2021).

2.3 THE EUROPEAN OPEN SCIENCE CLOUD (EOSC)

The European Open Science Cloud (EOSC) is a vision of the European Commission to provide an infrastructure to support open science and open innovation through the creation of a virtual environment with open and seamless service that will allow researchers to store, manage, analyse and reuse research digital objects, following the FAIR principles. Through EOSC, Europe wants to ensure that its researchers reap the benefits of data-driven science. Its use will not be limited to researchers, as it is also expected to serve education and training purposes and to be used by governments and the business sector. Overall, EOSC is expected to leverage other related EU initiatives and actions under the Open Science agenda.

Within this context, the EOSC pilot project looks into the technical, scientific and cultural challenges that need to be addressed in the deployment of EOSC. To achieve this, EOSC pilot will propose and trial a governance framework, develop a number of demonstrators, engage with a broad number of stakeholders to build the trust and skills required.

The EOSC Strategic Research and Innovation Agenda (SRIA) released in early 2021 sets the roadmap for the next years for achieving the EOSC vision. The Report identifies a series of technical challenges that need to be addressed for implementing the EOSC ecosystem which include: identifiers, metadata and ontologies, FAIR metrics and certification, authentication and authorisation infrastructure, user environments, resources provider environments, and the EOSC interoperability framework (EOSC, 2021). Technical interoperability is also discussed in the EOSC Interoperability Framework, along with other three layers (semantic, organisational and legal) and related recommendations (European Commission 2021). The EOSC Working Groups⁶ provide further guidance towards the achievement of the EOSC vision.

2.4 COMMUNITIES ENGAGED IN SCHOLARLY COMMUNICATION

The changes brought about by this new approach to conducting and communicating research result in an increased diversity of practices. They also require the involvement of a variety of stakeholders with different roles and needs, which are briefly presented below:

Researchers: The fundamental research practices of collecting, organising, processing and disseminating scientific information are highly related to the availability and discoverability of primary resources. Thus, for research to be effective and fruitful, scientific content has to be widely disseminated and effortlessly accessed -a condition that could potentially be met within the digital academic ecosystem, where scholarly communication is performed across a variety of channels and venues: developments such as the advent of Web 2.0 functionalities open new pathways in scholarly communication and significantly increase researchers' capacity to discover and exchange resources and information; moreover, an increasing number of dedicated tools and mediums underpin researchers' capacity to process and enrich a variety of sources available in different formats (e.g. texts, images, datasets).

⁶ For the period 2019-2020 the following WG were set up: Landscape, FAIR, Architecture, Rules of Participation, Skills and Training, Sustainability. Further information: <https://www.eoscsecretariat.eu/eosc-working-groups>

In this constantly evolving context, emphasis has to be placed on the implications of researchers' enhanced digital skills and the importance of the recently adopted processes and research methods. The advent of Digital Humanities raises issues related to the sufficient support of prevailing scholarly activities, which now involve a wide spectrum of user-driven innovative practices that entail providers' commitment in designing long-term strategies and tools for managing and preserving resources, enabling collaborative work, and disseminating research outputs. As researchers ask for inclusive publishing venues that can accommodate new types of research outputs (such as media), link research data to publications, and allow users' intervention (commenting, annotating), the quantity and quality of user-generated content becomes a question of crucial importance.

Publishers: Academic publishing has evolved into a diverse enterprise, involving small-, medium-, and large-scale independent or commercial publishers, different business models, practices and dissemination venues. As digital publishing becomes a norm, the existing variety of actors and models often results in wide discrepancies in terms of operational and technical standards.

Moreover, recent developments challenge the perceived role of academic publishers, who need to maintain their central place within the scholarly communication landscape, while being asked to respond to an increasing diversification of publishing practices and mediums. Nowadays, delivered value is equally generated by scientific quality and a variety of digital content-related attributes and features, such as availability and processability. Publishers are asked to develop new tools and services for researchers, and engage in incentives towards the optimization of digital workflows and content.

As the OPERAS survey (OPERAS-D, 2018) indicates, publishers have largely conceived this necessity, and share common interest in developing integrated services as well as standardized publishing and dissemination practices. To this end, the establishment of common standards enables more systematic collaboration among existing publishing initiatives and facilitates the deployment of innovative publication models and tools that help researchers to discover resources, communicate effectively, and assess the impact of their work.

Research funders: Public and private research funding bodies have been widely acknowledged as drivers of Open Science. Research funders provide a range of grant schemes to facilitate innovative research, and have adopted policies to make research outputs available in open access.

As funders need to assess grantees' compliance with their open access requirements, it is essential to introduce standards and services that link funded research and researchers with all relevant published content. On the other hand, provisions need to be made for proper assessment of the impact of funding mechanisms and open access policies: data for funders and authors has to be registered with published content, allowing proper identification and interlinking of all agents involved in a funded research project.

Service and infrastructure providers: As the OPERAS landscape study (OPERAS-D, 2017) indicates, fragmentation (both in terms of the size and nature of publishers and of their business models) is a key characteristic of the academic publishing landscape. In this context, the main challenge in designing sustainable open access publishing models is to identify current needs and limitations that permeate the scholarly communication framework.

A successful publishing service should deploy infrastructures designed to interoperate with a multitude of systems designed for the management and provisioning of digital content. Thus, platform providers cope with a series of administrative and technical issues related to the potential for content reuse, such as the

need for effective integration with repositories and/ or search engines; the incorporation of procedures that would ensure the long-term preservation and utilization of the content; and the development of tools to enable identification, authentication, metadata enrichment and discovery.

The introduction of common standards tackles the main obstacles towards the full interoperability of publishing infrastructures and paves the way for innovative services at inter-platform level by providing additional data, links and interactions to published material. It also allows a wide adoption of the fast technological developments that occur in the fields of open public data and of open digital content and enables broad reuse and organization of published content.

Libraries: The main role of libraries is to collect, preserve and provide access to scholarly resources. Due to their active participation in the research cycle, libraries face a number of important challenges stemming from the increasing volume of digital content, the predominance of digital dissemination mediums that scholars choose to make their work publicly available, as well as the varying types of material to be curated. Libraries are required to handle digital versions of printed content and, at the same time, make provisions for the preservation of digital resources.

Moreover, academic institutions develop publishing models in the context of which libraries assume various and combined roles in regard to content management. The realization of libraries as publishers and curators implies that their technical infrastructure and operational principles are compatible to the wider context of the digital research ecosystem, and entails challenges related to the introduction of additional workflows and outputs (publications, datasets, multimedia etc).

The introduction of shared and collectively applicable standards will enhance libraries' capacity to serve researchers in their binary status as producers and consumers of scientific content.

The table below summarises the importance of common standards for each stakeholder category:

<ul style="list-style-type: none"> •Researchers: inclusive publishing venues to accommodate new types of research outputs; link research data to publications; support of collaborative work •Publishers: new tools and services for researchers; optimisation of digital workflows; innovative publication models; content delivered in multiple formats •Funders: identification and interlinking of all agents and outputs of a funded research project •Infrastructure providers: long-term preservation and utilisation of the content; tools to enable identification, authentication, metadata enrichment and discovery •Libraries: assume new roles as publishers and curators; handle digital versions of printed content; long-term preservation of digital resources

3. STATE OF THE ART

The section focuses on technical and operational standards for digital research infrastructures and service providers, and highlights their importance in providing a sustainable framework for open scholarly communication. Six complementary areas of assessment have been identified: 1. Content quality and impact, 2. Findability and accessibility, 3. Interoperability, 4. Reusability, 5. Processability, and 6. Impact assessment.

3.1 EMERGING TRENDS

Research has evolved into a multifaceted activity that encompasses complex methodologies and workflows:⁷ text has ceased to be the exclusive resource for researchers, as the use of new applications enables scholars to discover, process and reproduce a wide range of digital-born or digitized sources (such as image sets, corpora, data sets, visualisations), and introduces techniques that allow collective contributions and metadata production. Unhindered flow of information gradually becomes a precondition for the incorporation of Humanities research into the digital ecosystem, as scholarly communication encompasses innovative practices such as information commenting, data extraction and metadata harvesting.

In this context, research practices in the Humanities increasingly relate to the systematic use of digital resources and tools. Digital Humanities (DH) has recently emerged as an innovative scholarly activity that successfully deploys digital workflows and introduces new methodologies based on collaborative and interdisciplinary work;⁸ the above developments reflect the ways in which research practices progress and science is taught, performed and communicated within the digital ecosystem.

This, in turn, suggests the implementation of new principles and standards that ensure openness, interoperability and processability for all scientific information ([Warren 2015](#)). A significantly increasing proportion of published material in the Humanities is available in open access and new venues of communicating research are emerging ([Eyman and Ball 2015](#)): in addition to publishing in conventional forms (e.g. journal articles and monographs in print or digital formats), researchers publish pre-prints, compile and deposit datasets and post their work in scientific blogs and other alternative dissemination venues.

Despite these developments and growing technical possibilities, the diversity of publication formats is still limited. Journal articles and the English language are still prevalent in scholarly communication, largely due to the market position of major commercial publishers and the focus of most national and international research evaluation frameworks on quantitative indicators and, especially, on the journal impact factor. The growing awareness that diversity in scholarly communication must be preserved and fostered is expressed in calls for bibliodiversity ([‘Jussieu Call’ 2017](#); [Shearer et al. 2020](#)). This overarching concept puts forward the idea of scholarly communication as an ecosystem which should rest on multilingualism, open and shared infrastructures and services, the diversity of business models in OA publishing, and quality-based research assessment. The diversity of business models has already been widely recognized and discussed in the OPERAS community, both in the context of monographs and journals. Notable results include the OPERAS Open Access Business Models White Paper ([Speicher et al. 2018](#)) and the Diamond OA Study ([Bosman, Jeroen et al. 2021](#)). The latter document reveals a correlation between the non-profit no-fee model and a lower capacity to apply technical standards, highlighting the need for coordination and support.

⁷ Harley, Acord and Earl-Novell’s “Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines” provides a thorough review of the current and emerging scholarly communication practices. ([Harley et al. 2010](#))

⁸ Programming Historian (<https://programminghistorian.org/>) is an indicative example of collaborative initiatives by the DH research community.

Openness emerges as a cultural value that permeates research; in addition to removing access barriers, openness also becomes a norm in the processes of reviewing and assessing research outputs. The concept of open peer review (OPR) encompasses a wide spectrum of practices ([Ross-Hellauer 2017](#)), ranging from revealing authors' and reviewers' names, to collective commenting by "non-experts" and post publication peer-review platforms (e.g. PubPeer, Winnower). OPR is an essential component of Open Science and is closely intertwined with digital publishing, as it is performed with the use of specific features (e.g. annotation technologies) and generates new outputs (e.g. conversation threads around published content) that diverge from conventional publication forms. Peer-review reports are increasingly often published along with articles and assigned PIDs. The idea of open peer-review encourages infrastructure development: it has become important not only to make peer-review reports easily identifiable and citable but to provide infrastructural support to new workflows associated with open-peer review. Some recent efforts focus on developing tools that would visually represent decoupled information about peer-review – e.g. PiePlate ([Boston 2020](#)).

The recent proliferation of disciplinary⁹ and regional¹⁰ preprint servers is closely tied to the emergent publishing model known as the overlay publishing model, where peer-review and publishing workflows build upon a distributed, interoperable, and sustainable network of digital archives, repositories, and preprint servers.¹¹ A notable initiative in this direction is Pubfair, which draws on the Next Generation Repositories vision, presented by COAR ([Ross-Hellauer et al. 2019](#)). The shift towards publishing platforms enabling an integrated and modular workflow (from submission, through peer review, to publishing) is endorsed by the EU initiative to establish Open Research Europe.

Integrating/linking publications and research data (esp. research data tracking) is another challenge. Despite the development of data repositories, (open) research data mandates and a growing awareness of the importance of FAIR data management, results are limited and significant progress in this area is yet to be made. Infrastructure and common standards are required to achieve interoperability and machine-readability.

Two major challenges and areas of action include:

- Automated or semi-automated discovery of research data references in publications ([Ikoma and Matsubara 2020](#); [Ritze and Boland 2013](#)),
- Indicating links in both human and machine-readable way (e.g. using badges).¹²

An important incentive in this process is provided by the policies intended to facilitate the transition to Open Science, such as OA publishing policies and open data policies. The awareness of transparent publishing policies has been growing, especially since DOAJ introduced new indexing requirements, which provide a compliance framework for the publishing policies of OA journals ([Penicina 2017](#)). There have been attempts to indicate visually the policies' completeness and compliance with standards using badges, e.g. Journal Publishing Practices and Standards (JPPS) badges,¹³ used by African Journals Online (AJOL) and the DOAJ Seal.¹⁴ At the same time, Center for Open Sciences' TOP (Transparency and

⁹ E.g. [SocArXiv](#) and [PsyArXiv](#) in 2016, [EdArXiv](#) in 2019.

¹⁰ [INArxiv](#) in 2017, [AfricArxiv](#) and [ArabiXiv](#) in 2018, [IndiaRxiv](#) and [Preprints.ru](#) in 2019.

¹¹ Cf. the experimental workflow tested on the platform Episciences.org ([Berthaud et al. 2014](#)).

¹² Along with giving credit to authors who share data and other materials related to publications, Open Science Badges, designed by Center for Open Science, establish a machine-readable link between publications and data ([Kidwell et al. 2016](#)).

¹³ <https://www.journalquality.info/en/>

¹⁴ <https://doaj.org/apply/seal/>

Openness Promotion) Factor is an attempt to express quantitatively, as a metric, the compliance with eight transparency standards in publishing policies, mostly focusing on the availability of data, code, preregistration, links between data and publications, etc. (Kepes, Banks, and Keener 2020). Further developments in this area may be expected in the years to come. In order to ensure wider implementation, machine readability, and full interoperability, it is necessary to standardize both policy compliance frameworks and technical specifications for badges.

Within this composite context, copyright issues and proper licensing of publicly available material has become a question of critical importance. As the pool of resources that can be accessed, distributed and reused is growing, it is essential for researchers to encourage access in a standardized way that allows others to share and build upon existing content. The need of all subsequent versions of the originally published work to be granted appropriate permissions leads to the emergence of flexible licensing processes, whereupon non-exclusive distribution of the originally published version of the work, or even prior to and during the review process, is enabled, as this can lead to productive exchanges, as well as earlier and greater citation.

Public access to scientific content (prior to or after publication) results in innovative collaborative workflows, successful scientific ventures, increased impact and widespread dissemination of researchers' work. On the other hand, it requires the global adoption of common operational and technical standards, to facilitate the dissemination of content in an organized manner that regulates copyright and access issues (Hutchens 2013; Browning, Guedon, and Kaplan 2014), while stimulating the activities of institutional stakeholders engaged in the scholarly communication cycle. The adoption of standardized licensing options suitable for OA content (usually creative Commons Licenses) has been pushed forward by funder policies¹⁵ and DOAJ application requirements.¹⁶ However, ensuring machine-readability of the license information remains a challenge because many publishing platforms and repositories still fail to provide licensing information in the metadata and expose it through protocols accessible to machines. The FAIR principles, where R (reusable) refers to licensing, put forward the requirement that the "conditions under which the data can be used should be clear to machines and humans".¹⁷ Plan S also addresses machine-readability of the license information by requiring that publishers and repositories embed metadata describing the OA status and the licence within the article itself or, in case of small publishers who cannot facilitate this, the OA status of the work and the license assigned to must be included in the description metadata.¹⁸

Free/Open/Libre Source Software¹⁹ has been playing an important role in the development of institutional and community-owned infrastructures (publishing platforms, repositories, tools, etc.), some of which would not be possible without it. OPERAS community members are already using a wide range of open-source solutions (OJS, Janeway, Hypotheses.is, DSpace, etc.), some of which are discussed in the OPERAS Tools Research and Development White Paper (Di Donato et al., 2018). Open source software

¹⁵ The European Commission endorses Creative Commons licenses ('The Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020' 2017).

¹⁶ <https://doaj.org/apply/guide/>

¹⁷ Principle R 1.1: "(Meta)data are released with a clear and accessible data usage license." <https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/>

¹⁸ <https://www.coalition-s.org/faq-theme/technical-requirements/>

¹⁹ According to a widely accepted definition, Free/Open/Libre Source Software (usually referred to as open source software, OSS, FOSS, FLOSS) is software with an open source code available online, and explicitly licensed as such. It is marked by collaborative code development.

is also expected to facilitate the transformation of scholarly publishing by developing modular technologies and workflows (McGonagle-O'Connell and Ratan 2019). The idea of open community-owned infrastructures has particularly gained prominence with the rising threat of vendor lock-in and infrastructure acquisition by great commercial publishers. The most comprehensive report on the state-of-the-art in open source publishing tools and platforms, *Mind the Gap* (Maxwell et al. 2019), found a wide variety of tools covering the entire publishing workflow, but it also highlighted overlaps and duplication of effort, as well as a lack of coordination, collaboration and interoperability. The result is a fragmented landscape, with many tools relying on small developer communities, unable to ensure funding and sustainability. To overcome these problems, it is necessary to adopt and implement common standards that would serve as a framework for collaboration, convergence and integration towards enabling interoperability among different platforms and an easy transition between platforms. Standardization of governance models in this area is also required in order to ensure long-term sustainability.

The development of digital tools inevitably opens up many new possibilities in the area of usage and impact tracking, contributing to the discussion on inclusive, responsible and sustainable metrics and research assessment (Waltman 2018; Wilsdon et al. 2017). The DORA Roadmap²⁰ (June 2018) seeks to operationalize the principles embodied in the San Francisco Declaration on Research Assessment (DORA, 2013)²¹ and the Leiden Manifesto (2015)²² by increasing “awareness of the need to develop credible alternatives to the inappropriate uses of metrics in research assessment”, research and promotion of “tools and processes that facilitate best practice in research assessment” and extending the “reach and impact of DORA’s work across scholarly disciplines and in new areas of the world”. The DORA Roadmap was followed by the decision to add signing DORA as a mandatory criterion for indexing in Redalyc.²³ In 2019, Plan S principles were updated to reflect a commitment made by the funders to “revise methods of research assessment along the lines of the San Francisco Declaration on Research Assessment (DORA)”.²⁴ Nevertheless, while pushing research assessment away from inappropriate applications of journal-based metrics, DORA principles do not provide specific solutions to many open issues, especially in SSH. Two major challenges in this area are defining assessment indicators so as to strike a balance between quantitative and qualitative approaches, and tracking the indicators. The latter is closely related to the development of infrastructure and tools.

Initiatives promoting the openness of bibliographic metadata may encourage the creation of more inclusive and cumulative metric-tracking services relying on open data and covering a wider range of publications and small publishers from all over the world. Although the first such initiative had emerged already in 2010 in the context of Semantic Web, it was only with the rise of the Initiative for Open Citations (I4OC)²⁵ in 2017 that a significant number of publishers started opening citation data. Notable examples of services using open citation data include the COCI, the OpenCitations Index of Crossref open

²⁰

<https://sfdora.org/2018/06/27/dora-roadmap-a-two-year-strategic-plan-for-advancing-global-research-assessment-reform-at-the-institutional-national-and-funder-level/>

²¹ <https://sfdora.org/read/>

²² <http://www.leidenmanifesto.org/>

²³

<https://sfdora.org/2018/10/09/signing-dora-is-a-mandatory-requirement-for-the-journals-indexed-by-redalyc/>

²⁴ <https://www.coalition-s.org/revised-implementation-guidance/>

²⁵ <https://i4oc.org/>. In 2020, a sister initiative, the Initiative for Open Abstracts (I4OA) was launched: <https://i4oa.org/>.

DOI-to-DOI citations²⁶ and the Open Ukrainian Citation Index.²⁷ There were also attempts to use open citation data in national research assessment systems (Di Iorio, Peroni, and Poggi 2019). Within the framework of Plan S, openly accessible data on citations according to the I4OC standards are one of the strongly recommended technical requirements for publishing platforms.²⁸

3.2 INITIATIVES FOR THE IMPLEMENTATION OF STANDARDS

Due to the multiplicity of workflows, object types and content carriers/mediums, the introduction of common standards into the scholarly communication digital landscape becomes a subject of crucial importance. To this end, an increasing number of international organisations have been working collaboratively towards the effective regulation of all issues related to knowledge representation and content dissemination and reuse, by developing protocols for digital content and online communication processes.

The **Dublin Core Metadata Initiative** (DCMI)²⁹ has long experience in the field of metadata standardization, and emerged as one of the main agents involved in monitoring, maintaining, and promoting standards. DCMI has adopted a federated structure, which comprises several communities and specialized Task Groups, committed to maintaining and updating metadata vocabularies. This collective effort leads to specific deliverables, updates of existing guidelines, and the adoption of additional recommendations and suggested terminology refinements for the **Dublin Core Schema (DC)**,³⁰ a core set of vocabulary terms used in the identification of digital objects (images, texts, web pages, etc.). DC consists of 15 elements describing the content, carrying medium, licensing and other properties of digital objects, and has been recently supplemented by additional metadata elements as well as a set of controlled vocabularies for the interpretation of element values.

The **World Wide Web Consortium** (W3C)³¹ is an extensive network of collaborating communities for the development of Web standards. Among the different working groups operating under the supervision of the W3 consortium, the Publishing Business Group (“Publishing BG”) and the Publishing Working Group (“Publishing WG”) are dedicated to the development of technologies and workflows that render the Web into a suitable ecosystem for publishing. The joint mission of the two Groups is to enhance publication accessibility, usability, distribution, archiving, as well as achieve reliable cross-referencing.

W3C has developed standards and specifications for a spectrum of web-oriented processes, technologies and tools, including default standards for TCP/IP communication protocols. It also provides recommendations for a variety of web-based languages used for knowledge representation (**OWL** - Web Ontology Language),³² text (**XML** - Extensible Markup Language)³³ and hyper text (**HTML** - HyperText Markup Language)³⁴ markup. A main contribution of W3C comes in the form of the Resource Description Framework (**RDF**),³⁵ a set of specifications that has evolved into a framework for information modeling.

²⁶ <https://opencitations.net/index/coci>

²⁷ <https://ouci.dntb.gov.ua/>

²⁸ https://www.coalition-s.org/technical-guidance_and_requirements/

²⁹ <http://dublincore.org/>

³⁰ <http://dublincore.org/documents/dces/>

³¹ <https://www.w3.org/>

³² <https://www.w3.org/OWL/>

³³ <https://www.w3.org/XML/>

³⁴ <https://www.w3.org/html/>

³⁵ <https://www.w3.org/RDF/>

Another body whose work relates to the implementation of electronic publishing standards is the **International Digital Publishing Forum (IDPF)**.³⁶ Its specific goal is to encourage the adoption of standards by identifying, evaluating and maintaining specifications for publishing workflows and technologies. IDBF also specializes in the development of applications and formats, such as the **EPUB**³⁷ content publication standard that enables the creation and dissemination of various content types as digital publications.

EDItEUR,³⁸ an international group for the implementation of standards designed to support e-commerce activities in the publishing sector, provides recommendations covering such diverse areas as e-infrastructures, bibliographic information and licensing. EDItEUR has developed a family of machine- and human- readable XML formats for the transmission of publication metadata records. **ONIX for Books Product Information Message**³⁹ and the set of Onix for Subscription Products formats provide a consistent way of incorporating metadata into electronic resource management systems and provide to end users links to these resources, as well as information on their licensing and terms of use.

The **Research Data Alliance (RDA)**⁴⁰ establishes standards to overcome current fragmentation within the research data landscape and facilitates the implementation of the FAIR data principles. Like many other organisations involved in the field of standards, RDA Comprises several working groups, dedicated to the establishment of a common framework for data production and reuse in a variety of SSH and STEM disciplines. RDA regularly issues recommendations and guidelines, introduces best practices and updates in issues of data curation, exchange and dissemination. It also assists research communities in understanding and following optimal data publishing workflows and increases researchers' awareness of emerging standards and best practices.

The **Text Encoding Initiative (TEI)**⁴¹ is a long standing community of practice, composed of institutions and researchers committed to developing and updating standards for the annotation of digital texts, with a special focus on the Humanities and Linguistics. The TEI consortium provides guidelines and other resources (trainings, bibliography and TEI-adopted software) that have been widely used by cultural and academic institutions for the digital representation of texts.

Standards apply not only to content and metadata, but also to information integrity and publishing workflows. The **Committee on Publication Ethics (COPE)**⁴² has released a series of core codes of conduct, with an aim to introduce documented practices, publication ethics guidelines and recommendations for editors and publishers. To support editorial teams in their effort to ensure integrity and transparency, COPE releases mandates addressing important aspects of the editorial and publishing processes, such as content reproducibility, licensing and issues of intellectual property, peer review and journal management.

3.3 OPERATIONAL AND TECHNICAL STANDARDS FOR E-INFRASTRUCTURES

This section focuses on technical and operational standards for e-infrastructures, highlighting their importance in providing a digital scholarly communication framework that fosters content reuse and

³⁶ <http://idpf.org/>

³⁷ <http://idpf.org/epub>

³⁸ <http://www.editeur.org/>

³⁹ <http://www.editeur.org/83/Overview/>

⁴⁰ <https://www.rd-alliance.org/>

⁴¹ <http://www.tei-c.org/>

⁴² <https://publicationethics.org/>

improved user experience.⁴³ Integrated e-infrastructures perform a series of basic functions related to content and user management, metadata indexing, identification and interlinking of resources and various entities, such as contributors, institutions, etc. to create complex knowledge graphs.

3.3.1. Suggested framework for the implementation of standards

The current diversity of workflows and operational models underpins the necessity for a global introduction of standards that will serve as a framework for shaping an integrated scholarly communication landscape. This may prove a rather difficult venture, and any sustainable approach should make provisions for these standards' scalable implementation and adjustment to the different infrastructure types and editorial workflows. The suggested framework for the adoption of standards across infrastructures identifies two different levels/layers for the introduction technical and operational improvements:

Platform/system level: the standard functionalities of publishing platforms should be deployed in a manner that allows basic functionalities such as a) content/metadata retrieval and disposal to third-party applications b) online browsing and retrieval of content c) access to metadata related to intellectual property issues and d) meta-search and access to content with persistent identifiers. To increase the potential of content reuse and effectively correspond to the needs of the research and publishing communities, e-infrastructures for scholarly communication should also support long-term preservation schemes and generate content usage/access statistics.

Inter-platform/semantic level: in the digital scholarly communication landscape, various levels of interoperability become an element of crucial importance, as they enable the design of advanced content identification, delivery and processing services. Communication across research infrastructures requires a) the provision of "meaningful", (i.e. machine-readable) metadata b) the use of standardised ontologies and controlled vocabularies c) the use of widely adopted knowledge representation languages d) the compliance of metadata with a specific encoding and e) the introduction of a common set of principles for data interlinking in the Semantic Web.

At this introductory stage, the main goal of the SIG is to trace the standards at platform and inter-platform level, and identify key areas for their implementation, as an essential step to ensure content quality, as well as findability and accessibility, interoperability, and reusability, i.e. compliance with the FAIR principles.⁴⁴

3.3.2. Content quality

In the context of this section, content refers to a) publicly available information about digital scholarly editions (e.g. journals) and b) published scholarly content (e.g. monographs, journal articles) available via digital infrastructures.

⁴³ See also Stathopoulos & Houssos. "Specifications and interoperability features for open digital content" <http://helios-eie.ekt.gr/EIE/handle/10442/8887>

⁴⁴ A detailed list of tools and services for e-publishing infrastructures may be found in the relevant OPERAS white paper, available at: <https://operas.hypotheses.org/aboutoperas/working-groups/platforms-and-services-wg>

In a recently published report,⁴⁵ the Committee on Publication Ethics (COPE),⁴⁶ the DOAJ,⁴⁷ the Open Access Scholarly Publishers Association (OASPA),⁴⁸ and the World Association of Medical Editors (WAME)⁴⁹ defined a number of core principles of transparency and provided recommendations for a range of managerial and editorial practices for scientific publications. The recommendations define the necessary information that should be provided on journal websites: journal's identity, focus and scope, names and affiliations of the editorial staff, roles and responsibilities, peer review information, a description of editorial processes and author guidelines. In order to be indexed in DOAJ, journals must also have a transparent Open Access and licensing policy. Furthermore, the same set of requirements are mandated by Plan S. The OPERAS certification service for book publishers⁵⁰ and platforms, which builds upon the DOAB requirements for participation by publishers,⁵¹ puts forward similar requirements: publishing practices, the peer review procedure and the licensing policy should be clearly defined on the publisher's website. These recommendations should be adopted as business standards, and be applied during the design and development of web-based interfaces for academic editions.

Regarding the quality of published scientific content, a key element in ensuring academic transparency and research integrity is peer review. As peer review methods maintain quality standards and provide credibility to scientific editions, editorial teams should encourage the engagement of reviewers and assist them in conducting and communicating their review. Whenever possible, digital tools (e.g. for suggesting reviewers, assessing review reports by the editorial staff and authors) should be made available to facilitate the review process. Current and emerging peer review practices entail certain challenges for scholarly communication e-infrastructures, which should support all different types of peer review, keep track of and compile records of exchanged communication between engaged parts, store and provide access upon demand to all different versions of submitted manuscripts. With the emergence of open peer review as common practice, provisions should also be made for future introduction of web 2.0 functionalities in publishing platforms and the implementation of PIDs for peer-review reports.

In sum, from an operational point of view, high content quality implies – at minimum – the implementation of complementary quality standards and editorial workflows that underpin the potential of digital infrastructures to serve as a venue of scholarly communication, support researchers' enhanced digital skills and encourage users' increased involvement during the peer review procedures. This could be accomplished by designing and implementing e-infrastructures based on the "software as a service" (SaaS) concept, which would impose common standards and compliance with the FAIR principles at the platform level by introducing validation criteria for content conformance with adopted quality standards, designing user friendly interfaces, enriching software functionalities with detailed guidelines (e.g. knowledgebase) to proactively support users, and introducing best practices for producing, reviewing and publishing scientific content, along with the above-mentioned platform/system level functionalities.

⁴⁵ "Principles of Transparency and Best Practice in Scholarly Publishing". Available at: https://publicationethics.org/files/Principles_of_Transparency_and_Best_Practice_in_Scholarly_Publishingv3.pdf

⁴⁶ <https://publicationethics.org/>

⁴⁷ <https://doaj.org/>

⁴⁸ <https://oaspa.org>

⁴⁹ <http://www.wame.org/>

⁵⁰ <https://www.doabooks.org/en/publishers/certification-service>

⁵¹ <https://www.doabooks.org/en/publishers/join-doab>

3.3.3 Findability and accessibility

A sustainable approach for e-infrastructures should be based on a framework that enhances metadata availability and interlinking ([Day 2005](#)), while respecting restrictions deriving from scholarly communication regulations; it should, furthermore, comply with researchers' needs of discovering and accessing a) files and metadata b) resources, identifiers for resources and contributors and c) information on dissemination and reuse rights.

A combination of technical and operational specifications should allow for research-oriented added value services:

Resource and metadata: in addition to online browsing and content downloading, electronic infrastructures should also support advanced search options and combined content retrieval features. Rich metadata improve content discoverability. Publishers should enrich their CrossRef metadata records with open abstracts and citations.

Identification: the use of persistent identifiers for content, contributors, funding agents or institutions is essential, as it facilitates a series of meta-services based on proper element interlinking. In addition to providing and/or displaying as relevant metadata unique persistent identifiers for persons, organisations and digital objects, identification also involves long-term commitment to resolving digital resources.

Preservation: content preservation is an essential part of sustainable planning for research infrastructures. A feasible preservation mechanism should be based on provisions for at least one remote copy of digital objects and relevant metadata entries, as well as automated processes for remote backup of digital content. It should also be designed upon commonly applied preservation schemes (OAIS) and eventually incorporate future changes in technologies and data formats.

3.3.4 Interoperability

The concept of interoperability is widely discussed in the context of EOSC and FAIR principles. The report of the EOSC Executive Board Working Groups FAIR and Architecture *EOSC Interoperability Framework* ([European Commission 2021](#)) identifies a number of challenges and presents recommendations, some of which indicate that the necessary tools and services enabling an optimal interoperability are yet to be created. The framework revolves around the concept of the FAIR Digital Object ([European Commission 2021, 29–30](#)) and identifies four levels of interoperability.

3.3.4.1 Technical interoperability

In technical terms, interoperability refers to the capacity of digital infrastructures and software to communicate in an automated manner and exchange reciprocally stored information and files. Technical interoperability is achieved with the implementation of common metadata standards across systems, and supported by the introduction of open APIs and Web Services that enable data transfers under globally applied communication protocols.

Limited interoperability has important implications in research and disturbs scholarly communication processes, as it significantly confines researchers' ability to exploit the full potential of web applications. According to the EOSC Interoperability Framework, major challenges in this area include authentication and authorization issues, diversified PID landscape and policies, a great variety of data formats and multiple levels of data granularity. Possible limitations in discovery, retrieval and dissemination of

scientific resources and information need to be taken into account when designing of e-infrastructures, and be addressed in combination with the prospective developments in research methods and information technologies.

As Almeida, Oliveira, and Cruz (2010) suggest, Open Source and Open Standards play a crucial role in interoperability issues. Open specifications are particularly highlighted in the context of EOSC (European Commission 2021). A basic set of recommendations for the implementation of interoperability standards and e-infrastructure functionalities could be as follows:

Harvesting and aggregating features: e-infrastructures must be able to provide data to third party applications, through APIs (Application Programming Interfaces) that conform to appropriate protocols. Data should also be deliverable partially in clusters and/or metadata form, thus allowing combined harvesting with the implementation of certain criteria. To ensure an optimal use of harvesting and aggregating features, the registration of e-infrastructures with relevant registries and aggregators should be included in the standard workflow. Monitoring mechanisms offered by aggregators should be used to optimize harvesting efficiency.

Data exchange: the main challenges that have to be addressed relate to the designing of a common communication framework that allows systems not only to exchange, but also identify data. This implies the use of common (and appropriate for each information type) metadata schemes, the availability of individual metadata records in a structured way (XML), and their compliance with several predefined formats to enable their incorporation into collective schemata.

3.3.4.2 Semantic interoperability

According to the EOSC interoperability framework, semantic interoperability relies on “precise and publicly available definitions of all concepts, metadata and data schemas” (European Commission 2021, 4). To ensure semantic interoperability across communities, it is necessary to define a minimum metadata model and crosswalks (mapping into other models). It is also important for infrastructure providers to adopt appropriate knowledge representation languages, and established ontologies for the documentation of their digital resources, in compliance with the principles of Linked Open Data (Yu 2014; Alexiou et al. 2016). Moreover, semantic interoperability assumes the interlinking of each metadata element with a suitable equivalent in a semantic artefact, i.e. a predefined list of values (e.g. vocabulary, list of standard terms, ontology, a thesaurus). However, the lack of common registries of semantic artefacts is a major obstacle in achieving semantic interoperability.

3.3.4.3 Organizational interoperability

In order to achieve organizational interoperability, services should be available, easily identifiable, accessible and user-focused. This level of interoperability focuses on the rules of participation, documentation, integration or alignment of the processes in different organisations and responsibilities for providing and maintaining common services, such as service catalogues, registers and common PID services (European Commission 2021).

3.3.4.4 Legal interoperability

Legal interoperability involves the ability to combine data from multiple sources without conflicts among restrictions imposed by respective licenses, taking into account relevant legal restrictions (intellectual property law, national security, privacy regulations, etc.). It rests upon various types of legal documents that are needed to ensure information exchange and data reusability among individuals and organizations and across jurisdictions (European Commission 2021).

This in practice means that legal use conditions for individual data objects should be clearly defined and determinable through standardised human and machine-readable licenses.

3.3.5 Reusability

Proper licensing is a key element for scholarly communication. Combined with the appropriate technical workflows, it enables optimal data flow across interlinked infrastructures and wide dissemination of research outputs and primary data. The use of Creative Commons licenses for open access content prevents copyright infringements and allows authors to define the terms of reuse and distribution of their work. Licensing information should be clearly indicated in all published formats. By making metadata available under the CC0 license publishers facilitate harvesting and reuse and improve cross-platform discoverability.

3.3.6. Processability

With the advent of digital methods and tools, research in the humanities shifts towards large-scale projects, often undertaken by multi-disciplinary, multi-institutional networks. This distributed production of knowledge drives digital workflows away from the basic functionalities of content uploading/downloading, and promotes online collaborative work in a cloud-based environment. It also introduces innovative research methods based on content mining and aggregation, text annotation and markup etc.

Digital research often produces deliverables in multiple formats, which should be supported by enriched workflows applied across publishing platforms. To effectively address issues stemming from the emergence of augmented and dynamic texts as communication medium, publication e-infrastructures should develop tools and workflows to support online/native authoring (e.g. Hyde 2015), submission and peer review, as well as the conversion of semantic-based inputs into other file types available to end users.

3.3.7. Impact assessment

As for impact assessment, web-based publication and interlinking of scholarly resources opens new pathways in measuring the outreach of research outputs. While citations remain the most widely acknowledged medium for evaluating research, there is a growing trend to assess impact by documenting users' engagement with published content and scientific data. *Altmetrics* refers to a family of relevant indicators, such as the number of actions and user responses to published content (views, discussions, downloads), references and citations in external resources, even shares in social media platforms.⁵² Thus, a comprehensive approach to digital publishing platforms should include measures to define altmetrics standards, and increase their technical capacity to provide usage-related statistics.

⁵² <https://www.altmetric.com/>

When measuring the outreach of research outputs, one should bear in mind that harvested content is available across platforms. Therefore, a true measure of impact should also take into account usage metrics from the harvesting platforms. This is not always possible for various organizational and technical reasons but platforms should collaborate towards providing aggregated usage statistics.

By contributing to initiatives for open bibliographic data, and especially I4OC and I4OA, publishers support the development of innovative and inclusive citation indexes and research graphs that stir the discussion about sustainable impact indicators and may eventually lead to the transformation of the research evaluation landscape.

4. FAIR PRINCIPLES

4.1 OVERVIEW

The FAIR principles provide generically applicable recommendations regarding data management and dissemination to facilitate its interoperability and reusability. The acronym stands for Findable, Accessible, Interoperable, Reusable.

They are mainly technical principles, considering data and metadata. The FAIR principles summarize in 15 recommendations various best practices identified over the years in the field of open science. They can be seen as a minimum set of requirements able to provide a technical building block for open science. As principles, they are generic and allow for various implementations, according to priority, feasibility, and existing practices criteria. Theoretically, although the first and main focus of the principles is STM big volumes of machine-generated data, they can be applied to any digital object, either native or derived.

A potential tool for building the technical environment of open science, the FAIR principles, however, are not focused on openness, but rather on accessibility. The Accessibility principle only recommends that there is a clear and standardized potential access to the (meta)data. Accessibility can be reached with fully closed data, and in fact the FAIR principles are also being adopted by private companies to handle their proprietary data. Authentication and authorization processes are not in contradiction with Accessibility as long as the access to the data relies on clear, standardized and free protocols.

The main focus of the FAIR principles are the common standards. As principles, they do not recommend any standard in particular. They recommend instead the use, for every data and metadata, any existing and widely used technical or community standard. The final objective of the recommendations is to increase (meta)data interconnection and reuse.

It is important to note that the most important quality of the standards and technologies underpinning FAIRification is machine-readability. Machine-readability is indeed the main criteria of a FAIR implementation. Machine-readable standard identifiers, descriptive metadata, access protocols, controlled vocabularies, licenses make a machine able to find, understand, analyse, combine and process the (meta)data. Such machine-readable standards are fully FAIR. In that sense, the FAIR perspective converges with the AI perspective.

Concretely, the **Findability** principle relies mainly on the use of persistent identifiers, rich descriptive metadata, and storage and indexing in accessible repositories. A counterexample is a corpus stored on a USB device with descriptive information only in the file name.

The **Accessibility** principle relies on the use of open, free and documented protocols, such as HTTP, OAI-PMH, FTP, even if they are combined with AAI processes. A counterexample is the data exchanged through emails on individual request.

The **Interoperability** principle relies on the use of standard representation of the data, like DublinCore, TEI, RDF, documented controlled vocabularies shared within a broad community, and machine-readable information on links with other (meta)data. A counterexample is a dataset described according to an individual and not documented vocabulary.

The **Reusability** principle relies on clear licensing information, as liberal as possible, made machine-readable, and recorded in the metadata. It should be completed with clear provenance information, which allows to better assess the reusability possibilities. A counterexample, without mentioning the lack of any license, is a license containing unclear provisions in a PDF file.

The various principles, on one hand, may overlap. Interoperable controlled vocabularies increase Findability, machine-readable standard licenses ensure Interoperability and Reusability, etc. On the other hand, some major aspects of good data management are not explicit in the FAIR principles, but can be considered as a consistent interpretation of the FAIR principles. This is especially the case of the long-term policies, which clearly improves (meta)data Findability and Reusability.

4.2 CONTEXT AND PROSPECTS

In concrete terms, the FAIRification guidance and implementation is characterized by a multiplicity of initiatives and tools.

At the **organization level**, beyond the various recommendations coming from funders and policy makers, it is possible to mention the GOFAIR initiative, funded by some European Union State members (Netherlands, Germany, France).⁵³ GOFAIR relies on Implementation Networks (INs) which represent different research communities leading the FAIRification in their own area. The communities can be dedicated to a discipline, an object, a project, etc. CO-OPERAS is the IN for the FAIRification of SSH data and publications.

The EOSC building will also rely on the FAIR principles and has set up a working group dedicated to the FAIR principles. Together with the EOSC Architecture WG, the FAIR WG has established the Interoperability framework (see section 2 of the current document on EOSC), which is strongly connected to the FAIR principles, especially regarding the technical and semantic interoperability layers.

Broader discussions also take place within the Research Data Alliance (RDA) as various RDA groups tackle the FAIRification of research data worldwide. Notably, this is the case of the FAIR Data Maturity Model⁵⁴ and the FAIR digital object⁵⁵ (former Data fabric) groups.

Finally, a certain number of organizations, including SSH RIs, are involved in the FAIRsFAIR H2020 project. The FAIRsFAIR project⁵⁶ produces studies and reports to foster FAIR principles adoption in the prospect of the EOSC. The project also organizes workshops and webinars aimed at gathering feedback from specific communities.

⁵³ GOFAIR was meant to last until 2021; the potential evolution of the initiative is currently under discussion.

⁵⁴ <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>

⁵⁵ <https://www.rd-alliance.org/group/data-fabric-ig.html>

⁵⁶ <https://www.fairsfair.eu/>

At the **implementation level**, there is an increasing number of tools for FAIRification, as well as an increasing number of objects to which the principles are applied (e.g. software). If we consider also the training material as a tool for FAIRification, as these tools often consist of a mix between self-assessment questionnaire and documentation, the FAIR implementation does not lack guidance. However, such tools are often directed towards individual researchers, with a focus on research datasets. For a wider range of data and actors, the existing FAIR-enabling tools often do not match the needs. In fact, the FAIRification process as a whole does not involve only the researchers, but the entire chain of data production, standardization, and dissemination actors, that is: data stewards, librarians, data managers, publishers, etc.

In line with this wider perspective, we can mention three main conceptual tools, which will probably shape the future FAIRification at the level of the infrastructures. The first one, the Implementation Profiles (IPs), has been provided by the GOFAIR team. The IP grid presents the full list of detailed FAIR principles as a questionnaire allowing to specify the existing FAIR implementations both for data and for metadata. The objective is to obtain a list of resources (software, standards, protocols, etc.) enabling FAIR and already in use in a specific community. The comparison between the various IPs makes it then possible to identify existing similarities and potential convergences between communities.

In the context of both GOFAIR and RDA, another FAIR-enabling process is being developed: the FAIR metrics, which should allow for an automated checking of the FAIR principles implementation. Simple HTTP queries can indeed verify for a given repository the existence of PIDs, the use of metadata standard, etc. The FAIR metrics are closely related to the FAIR maturity model, which proposes a standardized way to generally assess the FAIRness of the digital objects.

Finally, at a higher level, first specifications have been established for the FAIR digital objects. These specifications are being elaborated collectively in the context of RDA, and are part also of the EOSC interoperability framework. The FAIR digital object (FDO) represents an abstraction of the actual data stored in a specific repository. The FDO represents an object that is itself FAIR according to the FAIR maturity model, and has at least minimal specifications (a PID, a type, a resolution link). The concrete implementation of FDOs will require the set-up of a dedicated architecture, probably relying on the creation of specific servers. It can be noted that, in the prospect of the FDO, publications are not considered as the major issue, as they generally already have a PID, a clear type, and a sound dissemination policy - at least considering the publications of big commercial publishers.

5. OPERAS CONTRIBUTION IN PROMOTING COMMON STANDARDS AND ORIENTATIONS FOR FUTURE WORK

Over the past years, OPERAS and its members have been involved in various projects and initiatives that contribute to the promotion of common standards. The section presents these projects and initiatives and then discusses the key axes that will guide the future work of the SIG and OPERAS members.

5.1 PROJECTS AND INITIATIVES

5.1.1 HIRMEOS

HIRMEOS (High Integration of Research Monographs in the European Open Science Infrastructure),⁵⁷ was an EU funded project dedicated to the integration of high quality scientific content in the European Open Science ecosystem, with a special focus on the Social Sciences and the Humanities. The project was undertaken by nine members of the OPERAS network (research centres, university presses, university libraries and public foundations for the promotion of research), with common orientations towards the enhancement of Open Access through the development of European-wide infrastructures for scholarly communication. The project's main goal was to improve five important OA monograph publishing platforms, by designing and implementing common operational as well as technical standards, in light of their future incorporation into EOSC.

During this ongoing process, there were certain challenges to be met, mainly related to the different technologies, functionalities and features of the participating infrastructures. To this end, HIRMEOS enhanced their interoperability, by designing common services for the identification and validation of content and its metadata, as well as tools that enriched information and entity extraction. In the future, end users will annotate, extract and share content, while content providers and infrastructure administrators will be gathering usage data and metrics. HIRMEOS also enhanced the technical capacities of DOAB to import and ingest enriched and structured metadata, and also designed peer reviewing validation processes.

As for the project's impact, HIRMEOS enhanced the platforms' capacity to serve as venues for the discovery of high quality scientific content as well as mediums for the provision of metadata to third party aggregators, such as the OpenAIRE infrastructure. In addition, HIRMEOS established criteria and standards for the validation of e-publishing platforms, along with procedures for the certification of trusted partners. Finally, the project stood next to existing initiatives, and aspired to have a catalyst effect in including more disciplines into the Open Science paradigm, and widening its boundaries towards the SSH.

The project also contributed to the emergence of new standards in two sectors, namely peer-reviewing and metrics. In particular, the work done towards the enhancement of DOAB's technical capacity enabled the Directory to assign standardized peer review types certificates to academic books that lead to a clarification of that crucial domain for scientific quality. In terms of metrics, the work undertaken entailed the creation of a metrics service open to the whole community and supported by OPERAS infrastructure.

5.1.2 TRIPLE

The TRIPLE EU-funded project (Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration)⁵⁸ was launched in October 2019 and consists of a consortium of 19 partners (research centres, research infrastructures, universities) and is coordinated from France by Huma-Num (unit of CNRS). The objective of TRIPLE is to develop an innovative multilingual and multicultural platform that will enable the discovery and reuse of data from the social sciences and humanities (SSH) and equip

⁵⁷ For further information on the HIRMEOS project: <https://www.operas-eu.org/projects/past-projects/>

⁵⁸ For further information about TRIPLE project: <https://www.gotriple.eu/>

researchers with the means to connect and build interdisciplinary projects. The TRIPLE platform will be one of the dedicated services of OPERAS.

The TRIPLE platform will provide a single access point that allows users (researchers, institutions such as universities and libraries, but also enterprises, consultancies, media and service providers) to explore, find, access and reuse open scholarly SSH materials such as literature, data, projects and researcher profiles at European scale, which are currently scattered across local repositories. It will also help to find and connect with other researchers and projects across disciplinary and language boundaries. The goal of the platform is to make use of innovative tools to support research (e.g. visualisation, annotation, trust building and recommender system) as well as to discover new ways of funding research (e.g. crowdfunding). The platform is based on the Isidore search engine developed by Huma-Num.

The GOTRIPLE discovery platform aims to have the following five impacts. 1) User-orientation – TRIPLE integrates co-design principles into research and the development of new services. Users are key to all phases of the process, from needs analysis to tool testing and evaluation, as they know best how they work and what they need. 2) Strong ties between research, industry and society – TRIPLE facilitates more efficient and effective SSH research for societies at large by involving civil society, public institutions and companies into scientific projects, thus strengthening the links between different types of stakeholders. 3) Reconnection of culture and science – TRIPLE is not only a platform that ensures the discoverability of SSH resources and facilitates collaborations, but also a cultural platform to discover, understand and highlight European diversity in terms of societies, languages and practices. It helps to promote cultural diversity in Europe. 4) Contribution to the objectives of Open Science – TRIPLE improves the access to open content and resources and facilitates collaborations across disciplinary and language boundaries. Data sharing and usage according to the FAIR principles – Findable, Accessible, Interoperable, Reusable – is fostered. 5) Integration into the European Open Science Cloud (EOSC) – Because of a close link with the SSH Open Cloud (SSHOC) project, TRIPLE will be a major component of the SSH marketplace, which will be the entry door of the EOSC for all the different SSH services. Through TRIPLE, new innovative actors will be involved in the complex EOSC ecosystem.

5.1.3 Diamond OA Study

Between June 2020 and February 2021, OPERAS led a consortium of 10 organizations that studied OA diamond journals⁵⁹ to get a better understanding of the landscape. The study showed partial compliance with Plan S requirements and the need to develop infrastructures and increasing funding. It also highlighted operational challenges as most of them rely on volunteering. The outcome of the project has taken the form of a report which includes a set of recommendations. In particular, the report elaborated a set of recommendations around 5 key topics: technical support, compliance with Plan S, capacity building, effectiveness, and sustainability. The recommendations target stakeholders with a role in the delivery of issues addressed like funders, research performing organisations, societies and infrastructures. To facilitate the implementation of the recommendations, the Report provides an action plan that involves the preparation of an international workshop and symposium, the setting up of a funding strategy and building up a capacity centre to support the implementation of technical recommendations. To enhance further the engagement with the community a series of events have taken place following the report's publication to discuss further ways in which the OA diamond sector could coordinate better and be supported further.

⁵⁹ Further information about the OA Diamond Study and the Findings and the Recommendations Reports: <https://operas.hypotheses.org/4579>

5.1.4 CO-OPERAS

CO-OPERAS aimed at providing the knowledge base and the view of the community in finding the most suitable way to make the FAIRification of SSH data possible. This led to a series of actions:

- the first step was to focus on different kinds of data in SSH and issues in implementing FAIR principles, via a thorough review of the rich and growing literature about data in SSH;
- the second step consisted of workshops, where stakeholders were engaged to find out the perception about FAIR data, the needs of different communities, the challenges of FAIRification with regards to different disciplines;
- based on the review and workshops, some suggestions to entail the directions for OPERAS concerning the FAIRification of data have been proposed, including FAIRification tools prototypes, as well as measures to further engage the community both in discussions and implementation.

The various studies and activities conducted throughout the task confirmed that the SSH are currently in a transitional period as far as the FAIR principles are concerned. On the one hand, the domain has specificities related to well-rooted practices which do not always match the genericity that the FAIR principles imply. On the other hand, even when considering fields fully integrated with the digital environment, like DH, the knowledge about the FAIR principles' content and purpose remains often scarce. More generally, in order to expand the FAIR principles' adoption within the SSH, it is the principles' finality that should be outlined, showing how they increase the quality of research and can integrate even convergent, although pre-existing, practices.

Further actions should therefore be planned, taking into account three main aspects: advocacy, contextualization, and coordination. The effort to advocate for the FAIR principles started within the task should be sustained through the CO-OPERAS IN and, more largely, through the OPERAS RI. The specificities of SSH in terms of data types, dissemination infrastructures, and existing curation practices should be preserved through the FAIR principles' contextualization. The SSH fragmented landscape and the variety of its objects will, furthermore, require coordinated actions of the various actors in the field, closely connecting the SSH infrastructures between themselves and with the research communities.

Alongside the workshops' and conferences reports already published, here is the list of CO-OPERAS main expected outcomes:

- Advocacy: Within the OPERAS SIG on advocacy. Printable cards will present the main aspects, objectives, and benefits of FAIRification.
- Training: Training activities are also envisioned through a collaboration between OPERAS, OpenAire, and the EOSC-Pillar project. The objective is to organize workshops or webinars for researchers dedicated to specific disciplines or objects.
- Support: Stemming from OpenEdition's report on the FAIRness of its platforms, a FAIR publishing toolkit has been drafted. The objective is to provide open access publishers with an online step-by-step documentation about the FAIRification of their services.
- Communication: A common space of discussion for the SSH community will be open thanks to a blog dedicated to the FAIR SSH. The blog will allow a discussion on FAIRification aspects in the

SSH, present FAIR tools, methods, and implementations, and inform on related news and events.

- Implementation: To foster FAIR implementation in the various SSH areas, it is also planned to further improve the SSH Implementation Profiles.
- Coordination: At a broader level, a closer collaboration about the SSH FAIRification should bring together OPERAS, the ERICs for SSH Cessda, CLARIN, and DARIAH, as well as projects such as SSHOC and EOSC-Pillar.

5.2 FUTURE ORIENTATIONS FOR WORK

The current situation in scholarly communication is complex and is largely marked by attempts to expand, redesign or transform the traditional concept of scholarly publishing. At the same time, the multiplicity of content types, formats and resource versions raises questions of accessibility, usability and preservation of digital scholarly output and entails new roles for stakeholders who are actively participating in the design of policies, procedures and infrastructures. This is a challenging exercise, as it involves the implementation of long-term strategies and business models for sustainable resource management, as well as the utilization of standards in a shifting policy and infrastructure landscape, where the true convergence of standards and practices is yet to happen..

On the basis of the discussion that have taken place within the SIG and taking into consideration the projects and initiatives that OPERAS members have been involved in over the past years, the SIG members have identified the following actions:

1. Identification of synergies with other initiatives and projects undertaken by OPERAS members as well as with other SIGs. As described in the previous section, OPERAS members have been involved over the past years in various projects and initiatives such as the Diamond Study, the CO OPERAS IN whose work touches upon issues discussed in the framework of the SIG Common Standards and FAIR Principles.

1.1 Cooperation with other SIGs. Between June-September 2021 the SIG members will discuss with SIG Advocacy and SIG Tools potential synergies and development of common actions as well as the type that these actions can take. Currently SIG Advocacy is developing printable cards on the benefits of FAIRification- this could for example be a first step towards the agreement on further actions.

Interoperability issues are another aspect discussed both in the SIG Common Standards and the Tools one.

1.2 Improving the legal interoperability of e-infrastructures in the OPERAS community by supporting the adoption of appropriate licensing policies and development of relevant legal documents (e.g. establishing ownership of journals, SLAs between publishers and platforms, etc.).

1.3 Aligning journal policies with Plan S (in particular rights retention, licenses, self-archiving policies);

1.4 Developing a set of discipline-specific recommendations that would guide the inclusion of research data in publishing policies (data types, formats, repositories, machine-readable interlinking between publications and data, PIDs, data citation, etc.)

2. Identification of funding sources (such as for example forthcoming Horizon Europe calls) and specification of tasks led or requiring the input of SIG Common Standards.

More specifically, the following HE forthcoming calls have been identified as relevant for the SIG Common Standards and FAIR principles:

2.1 INFRA-2022-EOSC-01-02 “Improving and Coordinating technical infrastructure for institutional open access publishing across Europe”.

2.2 WIDERA-2021-ERA-01-43 “Capacity building for institutional open access publishing” . While the call focuses on non-technological aspects, SIG members could nonetheless explore if there is room for potential contribution in the preparation of proposal.

3. Expansion of SIG Common Standards and FAIR Principles by reaching out to new members.

Call to OPERAS members not currently participating in the SIG Common Standards to participate through the identification of specific tasks they could be involved in or lead.

Taking into consideration the maturity of OPERAS not only through the accession of new members over the past years, but also through its involvement in various project and initiatives (see also point 1 above), the focus on the SIG members in the coming months should be on identifying concrete tasks thereby increasing the links and synergies among the work of the SIG and work undertaken within OPERAS. This will enhance the commitment of current SIG members in the work of the group through the assignment of specific tasks and roles. In addition, the identification of concrete tasks will allow the SIG to reach out to other OPERAS members. Therefore, work undertaken under objectives 1 and 2 as described above will be of crucial importance for the future operation of the SIG.

3.1 Disseminating to OPERAS members the work of the SIG and its future plans (through the OPERAS mailing list). In addition, explore in collaboration with OPERAS Communication Team (and SIG Advocacy) possible ways of disseminating the work done in the context of the SIG and of reaching out to new members.

3.2 Explore with current SIG members the option of contacting specific members on the basis of their expertise in contributing to specific tasks/ actions.

REFERENCES

- ‘Jussieu Call’. 2017. 2017. <https://jussieucall.org/jussieu-call/>.
- ‘Open Research Europe: Open Access Publishing Platform. Beyond a Research Journal’. 2020. 2020. <https://open-research-europe.ec.europa.eu/>.
- ‘The Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020’. 2017. European Commission. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- Alexiou, Giorgos, Sahar Vahdati, Christoph Lange, George Papastefanatos, and Steffen Lohmann. 2016. ‘OpenAIRE LOD Services: Scholarly Communication Data as Linked Data’. In *Semantics, Analytics, Visualization. Enhancing Scholarly Data*, edited by Alejandra González-Beltrán, Francesco Osborne, and Silvio Peroni, 9792:45–50. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-53637-8_6.
- Almeida, Fernando, Jose Oliveira, and Jose Cruz. 2010. ‘Open Standards And Open Source: Enabling Interoperability’. *International Journal of Software Engineering & Applications* 2 (1): 1–11. <https://doi.org/10.5121/ijsea.2011.2101>.
- Berthaud, Christine, Laurent Capelli, Jens Gustedt, Claude Kirchner, Kevin Loiseau, Agnès Magron, Maud Medves, Alain Monteil, Gaëlle Riverieux, and Laurent Romary. 2014. ‘EPISCIENCES - an Overlay Publication Platform’. In , 78–87. IOS Press. <https://doi.org/10.3233/978-1-61499-409-1-78>.
- Bosman, Jeroen, Frantsvåg, Jan Erik, Kramer, Bianca, Langlais, Pierre-Carl, and Proudman, Vanessa. 2021. ‘OA Diamond Journals Study. Part 1: Findings’. Zenodo. <https://doi.org/10.5281/ZENODO.4558704>.
- Boston, Arthur. 2020. ‘PiePlate: Proposing a Visual Peer-Review Overlay Service’. Zenodo. <https://doi.org/10.5281/zenodo.4443231>.
- Browning, Sommer, Jean-Claude Guedon, and Laurie Kaplan. 2014. ‘Metadata and Open Access: Reliably Finding Content and Finding Reliable Content’. In *Too Much Is Not Enough!*, 505–11. Against the Grain. <https://doi.org/10.5703/1288284315314>.
- Council of the European Union (2016) The transition towards an Open Science system- Council Conclusions, Brussels, 27 May 2016, <https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>
- Day, Michael. 2005. DCC Digital Curation Manual Instalment on Metadata. HATII, University of Glasgow; University of Edinburgh; UKOLN, University of Bath; Council for the Central Laboratory of the Research Councils. <https://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/metadata/metadata.pdf>.

- Di Donato, Francesca, Patrick Gendre, Elena Giglia, Arnaud Gingold, Maciej Maryl, Tom Mowlam, Ghislain Sillaume, Heather Staines, Sofie Wennstrom. 2018. OPERAS Tools and Development White Paper, OPERAS. <https://doi.org/10.5281/zenodo.1324110>.
- Di Iorio, Angelo, Silvio Peroni, and Francesco Poggi. 2019. 'Open Data to Evaluate Academic Researchers: An Experiment with the Italian Scientific Habilitation'. ArXiv:1902.03287 [Cs], February. <http://arxiv.org/abs/1902.03287>.
- EOSC (2021) Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud, Version 1.0, 15 February 2021, https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf
- European Commission (2019) European Research Area. Progress Report 2018. Report from the Commission. Luxembourg. Publications Office of the European Union.
- European Commission (2016) H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. version 3.0, 26 Kuly 2016.
- European Commission (2020) Progress on Open Science: Towards a Shared Research Knowledge System. Final Report of the Open Science Policy Platform. Luxembourg. Publications Office of the European Union.
- European Commission (2021) EOSC Interoperability Framework. Report from the EOSC Executive Board Working Groups (WG) FAIR and Architecture, Luxembourg. Publications Office of the European Union. <https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-190308283>
- European Commission. 2021. EOSC Interoperability Framework: Report from the EOSC Executive Board Working Groups FAIR and Architecture. Luxembourg: Publications Office of the European Union, <http://doi.org/10.2777/620649>. https://op.europa.eu/publication/manifestation_identifier/PUB_KI0221055ENN.
- Eyman, Douglas, and Cheryl Ball. 2015. 'Digital Humanities Scholarship and Electronic Publication'. In *Rhetoric and the Digital Humanities*, edited by Jim Ridolfo and William Hart-Davidson, 65–79. Chicago: The University of Chicago Press.
- Harley, Diane, Sophia Krzys Acord, Sarah Earl-Novell, Shannon Lawrence, and C. Judson King. 2010. *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Berkeley: Univ Of California Press. <https://escholarship.org/uc/item/15x7385g>.
- Hutchens, Chad. 2013. 'Open Access Metadata: Current Practices and Proposed Solutions'. *Learned Publishing* 26 (3): 159–65. <https://doi.org/10.1087/20130302>.
- Hyde, Adam. 2015. 'Open Access and Standards'. *The Journal of Electronic Publishing* 18 (1). <https://doi.org/10.3998/3336451.0018.117>.
- Ikoma, Tomoki, and Shigeki Matsubara. 2020. 'Identification of Research Data References Based on Citation Contexts'. In *Digital Libraries at Times of Massive Societal Transition*, edited by Emi Ishita,

- Natalie Lee San Pang, and Lihong Zhou, 149–56. *Lecture Notes in Computer Science*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-64452-9_13.
- Kepes, Sven, George C. Banks, and Sheila K. Keener. 2020. 'The TOP Factor: An Indicator of Quality to Complement Journal Impact Factor'. *Industrial and Organizational Psychology* 13 (3): 328–33. <https://doi.org/10.1017/iop.2020.56>.
- Kidwell, Mallory C., Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, et al. 2016. 'Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency'. *PLOS Biology* 14 (5): e1002456. <https://doi.org/10.1371/journal.pbio.1002456>.
- Maxwell, John W., Erik Hanson, Leena Desai, Carmen Tiampo, Kim O'Donnell, Avvai Ketheeswaran, Melody Sun, Emma Walter, and Ellen Michelle. 2019. *Mind the Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms*. 1st ed. MIT Press. <https://doi.org/10.21428/6bc8b38c.2e2f6c3f>.
- McGonagle-O'Connell, Alison, and Kristen Ratan. 2019. 'Can We Transform Scholarly Communication with Open Source and Community-Owned Infrastructure?' *Learned Publishing* 32 (1): 75–78. <https://doi.org/10.1002/leap.1215>.
- Open Science Monitor (2019) Study on Open Science: Monitoring Trends and Drivers, D.2.4 Final Report, 13 December 2019, https://ec.europa.eu/info/sites/default/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_open_science_monitor_final-report.pdf
- OPERAS-D (2017) Landscape Study on Open Access Publishing- Annex to OPERAS Design Study, <https://doi.org/10.5281/zenodo.1299151>
- OPERAS-D (2018) Developing network and e-infrastructure strategy. Report on online survey on optimizing e-infrastructure. Deliverable 2.3, February 2018. https://f-origin.hypotheses.org/wp-content/blogs.dir/2465/files/2018/05/operas_online_survey_optimizing_e-infrastructure.pdf
- Penicina, Ludmila. 2017. 'Towards Supporting Business Process Compliance with Policies'. In *Perspectives in Business Informatics Research*, edited by Björn Johansson, Charles Møller, Atanu Chaudhuri, and Frantisek Sudzina, 93–107. *Lecture Notes in Business Information Processing*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-64930-6_7.
- Ritze, Dominique, and Katarina Boland. 2013. 'Integration of Research Data and Research Data Links into Library Catalogues'. *International Conference on Dublin Core and Metadata Applications*, September, 35–40. <https://dcpapers.dublincore.org/pubs/article/view/3683>.
- Ross-Hellauer, Tony, Benedikt Fecher, Kathleen Shearer, and Eloy Rodrigues. 2019. 'Pubfair: A Distributed Framework for Open Publishing Services'. Version 2. COAR. <https://www.coar-repositories.org/files/Pubfair-version-2-November-27-2019-2.pdf>.
- Ross-Hellauer, Tony. 2017. 'What Is Open Peer Review? A Systematic Review'. *F1000Research* 6: 588. <https://doi.org/10.12688/f1000research.11369.2>.

- Shearer, Kathleen, Leslie Chan, Iryna Kuchma, and Pierre Mounier. 2020. 'Fostering Bibliodiversity in Scholarly Communications: A Call for Action'. Zenodo. <https://doi.org/10.5281/zenodo.3752923>.
- Speicher, Lara, Lorenzo Armando, Margo Bargheer, Martin Paul Eve, Sven Fund, Delfim Leão, Max Mosterd, Frances Pinter, and Irakleitos Souyioultzoglou. 2018. 'Operas Open Access Business Models White Paper', July. <https://doi.org/10.5281/ZENODO.1323708>.
- Waltman, Ludo. 2018. 'Responsible Metrics: One Size Doesn't Fit All'. In STI 2018 Conference Proceedings: Proceedings of the 23rd International Conference on Science and Technology Indicators, 526–31. Leiden: Centre for Science and Technology Studies (CWTS). <https://openaccess.leidenuniv.nl/handle/1887/65307>.
- Warren, John W. 2015. 'Zen and the Art of Metadata Maintenance'. The Journal of Electronic Publishing 18 (3). <https://doi.org/10.3998/3336451.0018.305>.
- Wilsdon, James, Judit Bar-Ilan, Robert Frodeman, Elisabeth Lex, Isabella Peters, and Paul Wouters. 2017. Next-Generation Metrics: Responsible Metrics and Evaluation for Open Science. Report of the European Commission Expert Group on Altmetrics. Brussels: European Commission - Directorate-General for Research and Innovation. <http://dx.doi.org/doi:10.2777/337729>;
- Yu, Liyang. 2014. A Developer's Guide to the Semantic Web. 2nd ed. Berlin Heidelberg: Springer-Verlag. <https://doi.org/10.1007/978-3-662-43796-4>.