

REPORT

Scoping the Open Science Infrastructure Landscape in Europe

October 2020

Scoping the Open Science Infrastructure Landscape in Europe

“Scoping the Open Science Infrastructure Landscape in Europe”

Report commissioned by:
SPARC Europe

<https://sparceurope.org/>

Contact:
Vanessa Proudman
Director, SPARC Europe
vproudman@sparceurope.org

Report authors:
Victoria Ficarra, Mattia Fosci, Andrea Chiarelli,
Bianca Kramer, Vanessa Proudman

Report DOI: 10.5281/zenodo.4159838
Dataset DOI: 10.5281/zenodo.4153742

Report dated: October 2020



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

Table of Contents

1. Executive summary	3
2. Introduction	7
3. Overview of survey respondents	9
4. The OSI Landscape in Europe.....	11
5. Technology	29
6. Governance and sustainability.....	35
7. Conclusions	50
Appendix A: Respondents by country.....	54
Appendix B: Survey questions	60

1. Executive summary

Introduction This report collates the results of a survey of infrastructure or services that are part of the European Open Science infrastructure (OSI) landscape.

The survey consisted of two parts: Part 1 assessed the general infrastructure offering; Part 2 (which was divided into two sections) considered the infrastructure's intended audience and stakeholder community, technical design, and sustainability. Part 1 of the survey was completed by 120 relevant OSIs from 28 European countries, while Part 2 was completed by 67 (part 2a) and 68 (part 2b) respondents comprising executives and senior managers, IT specialists, researchers, and contributors.

The survey was published in spring 2020 and was distributed by email and on social media to international library networks and consortia and via personal contacts at libraries in all European countries, to numerous scholarly communication lists, via EOSC networks, and amongst LIBER, SCOSS, SPARC Europe and OPERAS members. In addition, we contacted infrastructure listed in *Innovations in Scholarly Communication*. As a result, we received 353 responses, but after normalization and verification that all datasets conformed with the definition of infrastructure as provided at the beginning the survey, e.g. OSIs with a regional, national or international focus, prior to 2019, no projects and based in Europe, 120 OSIs were selected as relevant responses.

Goals and purpose of OSI: OSIs are motivated by their ambition to further OA and OS and provide access to a variety of research outputs. Discovery and archiving are the main areas of the research lifecycle which they cover. The majority integrate with external systems run by not-for-profits or RPOs.

The survey found that OSIs are predominantly motivated by their ambition to further the vision of Open Access (OA) and Open Science (OS) by making research and knowledge openly and widely available (72 out of 120). 97 out of 118 responding OSIs provide services that support OA, 66 to open data, 49 support open software. A considerable amount of respondents (35) provide support for OS evaluation.

Just over half of 116 OSIs provide access to both journals and data, followed closely by conference papers showing its significance for engaging in conversation on innovative research. Access to a variety of non-traditional and early research outputs is also provided, with 41 respondents providing access to images and photographs, and 32 providing access to software and code. Preprints are served to a lower extent than other outputs – in 22 cases – with no notable differences between disciplines. The most common areas which OSIs support in the research lifecycle were (i) enabling discovery and search services and (ii) providing storage for archiving and/or digital preservation. 95% of all OSIs provide services in three or more stages of the research lifecycle. In line with this finding, when asked to estimate the count of their total objects created, published, hosted, provided discovery of, evaluated and/or archived, OSIs reported they had provided discovery of the highest number of objects and had created and evaluated the least. Aggregation & indexing, search, storage, identity (e.g. ORCID) and persistent identifiers are central to the OSI ecosystem since they are referred to by many.

Target audience: OSIs predominantly serve researchers and libraries across all disciplines with a global focus. Most have consulted with stakeholders on needs in the last 2 years.

Technical openness: The majority of OSIs have APIs, integrate with external systems, and allow code contributors. The majority of OSIs either have fully or partly open source software.

Compliance with specific open standards and principles: Most OSIs have strategies to comply with FAIR and with open standards, EOSC service requirements and Plan S technical conditions. OSIs report maturity in the open content and open standards of the

The majority of OSIs (57 out of 67) integrate with external systems and services; the overwhelming majority of these are run by not-for-profits or RPOs. Infrastructures that support persistent identifiers are mentioned by 45% of the cohort.

Almost all OSIs are aimed at supporting researchers (114 out of 118), but a considerable proportion also support libraries (74) and research managers (66) who are often tasked with evaluating research. A large proportion of OSIs (51 out of 64) have consulted their stakeholders on their needs within the last two years. Most OSIs (72 out of 118) serve the full spectrum of subject areas. Despite being based in Europe, roughly half of 118 OSIs operate globally showing the significant contribution that European OSIs are making to supporting the OS sector worldwide. OSIs that have a national focus (46 out of 118) tend to be based at research organisations, whereas globally-focused OSIs tend to be either based at not-for-profits or research organisations. The majority of for-profits are also globally-focused.

The large majority of OSIs (51 out of 64) have consulted their stakeholders on their needs within the last two years which shows their customer service orientation-focused OSIs tend to be either based at not-for-profits or research organisations.

Principles of openness also apply to the technology underpinning OSIs. The vast majority of OSIs (50 out of 65) have application programming interfaces (APIs), typically to support data harvesting (12 out of 46) and metadata exchange (7 out of 46), which shows the OSIs' inclination towards openness. Many OSIs (57 out of 67) also integrate with external systems or services. Of the tools and services that three or more OSIs integrate with (n=19), the majority are either run by a not-for-profit organization or a research performing organization. This points to the important role of not-for-profit organisations and the infrastructure they provide in the open science landscape .

Over half of OSIs are built using fully open source software (34 out of 64) or partly open source (19), which shows a clear trend towards open source. Collaboration is key to the success of many OSIs since 33 out of 53 of respondents allow others to contribute their code to the OSI. Most commonly, the number of code contributors used on open source software ranged between one and five (14 out of 27) although seven reported using more than 25.

A clear majority of responding OSIs (45 out of 54) have a strategy to comply with FAIR data principles which shows that FAIR is becoming a strong standard. Across the board, responding OSIs tend to prioritise FAIR (Findable, Accessible, Reusable and Interoperable) principles around discoverability, documentation and accessibility. 36 out of 61 OSIs noted usage and compliance with at least some open standards, whilst a large minority (23 out of 61) stating that the OSI they represent only use open standards. Finally, most OSIs state they meet the European Open Science Cloud (EOSC) service requirements and Plan S mandatory technical conditions, which demonstrates that many OSIs are adaptable and up-to-date with recent standards.

COAR/SPARC Principles most. They see most challenges with good governance, sharing open content and open standards.

Governance: The majority of OSIs have a board, steering group or advisory committee and are guided by a formal vision, plan or roadmap.

Financial sustainability: OSIs are generally run on low resources despite offering a range of services. One third of OSIs start the year with no approved budget. Most OSIs do operate sustainably but their ability to do so is reliant on grant income.

OSIs report maturity most in the open content and open standards of the COAR/SPARC principles. Effectively implementing good governance, sharing open content and applying and following open standards are the areas that see the greatest amount of challenges. Most good practices are mentioned in the areas of open content and open standards.

Four in five OSIs have a board, steering group or advisory committee, and stakeholders such as researchers and libraries are frequently represented on these bodies. This illustrates that a large majority exercise some good governance. Almost all OSIs are also guided by mission/vision documents (94 out of 118), with considerably fewer utilising strategic plans or roadmaps (68 out of 118), illustrating that many know where they want to go, but considerably less have formulated and described plans on how to get there. Following good governance is a key challenge for OSIs.

Services are commonly run by a small staff of 2-5 full-time equivalents (FTE) (25 out of 65) or less FTEs by not for-profits and RPOs and with varying levels of reliance on volunteers. The sustainability of an OSI can thereby be considered somewhat under threat when dependent on a small core team. Some of the services run with limited personnel resources provide a range of services which means that it is not always the case that more resources are necessary to provide a rich service offering.

Two thirds of respondents (42 out of 64) begin each fiscal year with an approved budget, which is most commonly used to cover salaries (57 out of 66), travel expenses (47 out of 66), equipment (44 out of 66) and – for half of respondents – marketing. One-third, however, start the year without an approved budget which can impact the stability of an OSI.

Two in five respondents (n=53) report an annual revenue of less than €50,000, but 12 OSIs report a revenue above €500,000. Infrastructure income sees revenues matching the outgoings to a large extent. Most OSIs operate sustainably and at least break even on their costs, with roughly half (19 out of 36) noting that their operational deficit is covered by grants or sponsored projects, while another half (18 out of 36) breaks even thanks to earned income. Approximately one-third of respondents (22 out of 64), however, begin each fiscal year without an approved budget.

OSIs most commonly report national government grants as their most important source of income (22 out of 62) with the European Commission and membership fees coming in second and third respectively (with 14 responses each). It is worth noting that an additional 30 respondents commented that their OSI does not rely on grant funding including most of the for-profits, over 15 RPOs and a handful of not-for-profits. However, thirteen out of 32 respondents stated that without grant funding their OSI could only remain viable for less than a year which is a clear concern for a user community that might depend upon it. The greatest sustainability challenges relate to costs and funding, a lack of resources such as staff or equipment and the ability to keep up with technological development and OA/OS.

Conclusions

We see a diverse, interconnected, open, professional and viable OSI ecosystem developing in Europe on solid ground – one that is worth investing in. It is a system that is made up of valuable service providers, both large and small, serving the global research community. Nonetheless, OSIs still have a range of issues to contend with in their organisations and strategies, particularly as they move towards more openness and a sustainable future.

Sharing lessons learnt, developing communities of practice, developing guidance, pooling resources and working with initiatives such as Invest in Open Infrastructure (IOI) and JROST will help them grow even further.

Additionally, despite a strong commitment to open source and open standards by many, challenges remain for some in good governance, sharing open content and applying open standards. This ecosystem will thrive if OSIs follow good governance practices, ensuring the community it will be steered by their needs, and will stay true to the values of research.

To sustain themselves, OSIs will need to continue to diversify their fund-raising efforts and upskill to embrace a range of business or revenue models in the future to spread the risk to their financial stability. Funding agencies, governments, institutions, charities and other funders need to consider strategies on how to effectively fund this rich and important landscape more structurally. We also call on governments to maintain and increase support for both development activities and for sustaining operations. Making smart choices on what to invest in will be essential.

2. Introduction

2.1 Background and objectives

Everything we have gained by opening content and data will be under threat if we allow the enclosure of scholarly infrastructures. – Geoffrey Bilder, Jennifer Lin, Cameron Neylon¹

The last two decades have seen a surge in Open Access and Scholarship/Science policies and activities in Europe with research performing organisations (RPOs), not-for profits (NFPs) and for-profits contributing to the Open Science infrastructure (OSI) offering to support researchers, libraries, research managers, governments, service providers and others. By infrastructure we mean the structures and services needed for Open Science/Scholarship to operate, e.g. services, protocols, standards and software that the academic ecosystem needs in order to perform its functions during the research lifecycle. ESFRIs (the European Strategy Forum on Research Infrastructures) are excluded in this report.

The dependence of open research and scholarship on digital infrastructure has grown and, while the current and emerging open infrastructure efforts are promising, they are currently operating independently of one another and without collective funding or coordinated long-term community support for sustainability. Furthermore, while many OSIs projects have been generously funded in this area over the years, and proven their value over time, many later struggle with operational costs in the absence of dependable mid-term or long-term funding solutions.

While many governments are investing in open infrastructure to support their policies, a recent SPARC Europe study into European research funders' OS policies and practices also reported that in general European research funders are not yet consistently funding the operations of essential scholarly communication and OS as much as they might.² This puts a number of important infrastructures at risk and as a consequence, the products and services that comprise open infrastructure are increasingly being tempted by buyout offers from large commercial enterprises. This threat affects both not-for-profit open infrastructure as well as closed, and is evidenced by the buyout in recent years of commonly relied on tools and platforms such as SSRN, bepress, Mendeley, and Github. In the case that infrastructure has been developed by the OS community and would ideally remain independent, governed by it and closely aligned with the values of research, how does such infrastructure sustain itself?

In their *Principles for Open Scholarly Infrastructures* of 2016, Bilder, Lin and Neylon underline the importance of systematically addressing a range of elements to help run and sustain open infrastructures successfully.³ Other principles have been built upon these to support the research process and its values like the COAR/SPARC *Good Practice Principles for Scholarly Communication Services*.⁴ However, to what extent are such principles being applied, and how do they work in practice?

This research project, funded by the Open Society Foundations and in collaboration with Invest in Open Infrastructure, aims to make concerted steps to help scholarly communications infrastructure developed in Europe become and stay open and independent.^{5 6} This work is part of a larger project, some of which was published in Sept 2020, which includes advocating for the development of a more open infrastructure in Europe following open principles and understanding some of the sustainability journeys and choices of certain key services worldwide.⁷

¹ Bilder G, Lin J, Neylon C (2015) *Principles for Open Scholarly Infrastructure-v1*, retrieved [date], <http://dx.doi.org/10.6084/m9.figshare.1314859>

² Insights into European research funder Open policies and practices, SPARC Europe, 2019, <https://zenodo.org/record/3401278>, p 17-18

³ idem

⁴ Good Practice Principles for Scholarly Communication Services, 2019, <https://www.coar-repositories.org/files/COAR-SPARC-Good-Practice-Principles.pdf>

⁵ Open Society Foundations: <https://www.opensocietyfoundations.org/>

⁶ Invest in Open Infrastructure: <https://investinopen.org/>

⁷ 10 key interviews. Insights into the sustainability of open infrastructure services: <https://www.sparceurope.org/joiinterviews>

In spring 2020, SPARC Europe commissioned a survey to investigate the current state of the OSI landscape in Europe to inform the strategic path ahead for OSIs and their funders – with a particular focus on their sustainability. This document presents the results of a three-part survey, analysed by Research Consulting, SPARC Europe, and OPERAS, with a significant contribution made by Bianca Kramer.^{8 910} The study investigates the OSI landscape in Europe, seeking to identify the scope of OSIs currently in operation and a range of features such as their audience, openness and sustainability. It also analyses the practical feasibility of certain open scholarly communications infrastructure principles. OSIs address the unmet needs of researchers and other stakeholders involved in the scholarly communications landscape or offer alternatives to existing services.

The findings in this report will help funders identify investment strategies, and we expect that ongoing initiatives such as Invest in Open Infrastructure, and The Global Sustainability Coalition for Open Science Services (SCOSS) will be able to build on this work to further their missions in the OSI landscape.¹¹¹² This document summarises the findings at a pan-European level; it does not attempt to draw a connection between responses and the national context, which could be part of a separate analysis.

2.2 Definitions

Open Access (OA): Open Access is the free online availability of research articles, books, or other published content, combined with licensing that allows reuse with limited or no restrictions (from the definition used in a previous SPARC Europe report *Insights into European research Funder Open policies and Practices*).¹³

Open Science/Scholarship (OS): The practice of science and scholarship in such a way that others can collaborate and contribute, where research data, lab notes and other research outputs and processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods (adapted from the FOSTER definition).¹⁴

Open Access and Open Science/Scholarship Infrastructure (OSI): We define OA & OS Infrastructure as sets of services, protocols, standards and software contributing to the research lifecycle – from collaboration and experimentation through data collection and storage, data organization, data analysis and computation, authorship, submission, review and annotation, copyediting, publishing, archiving, citation, discovery and more (IOI definition).

2.3 Limitations

Please note that this research may show the effects of self-selection bias. All participating OSIs were volunteers and autonomously decided whether they wished to contribute to our work: therefore, our findings may not represent the entire population of European OSIs.

We received 353 responses, but after normalization and verification that all datasets conformed with the definition of infrastructure as provided at the beginning the survey, e.g. OSIs with a regional, national or international focus, prior to 2019, no projects and based in Europe, 120 OSIs were selected as relevant responses.

2.4 Survey question set and data availability statement

The question set used in the survey is available in the Appendix and in the Zenodo repository: <https://www.doi.org/10.5281/zenodo.4048270>. We encourage you to re-use the survey instrument.

⁸Research Consulting: <https://www.research-consulting.com/>

⁹SPARC Europe: <https://sparceurope.org>

¹⁰OPERAS: <https://operas.hypotheses.org/>

¹¹Invest in Open Infrastructure: <https://investinopen.org>

¹²SCOSS: <https://scoss.org/>

¹³Fosci, Mattia, Richens, Emma, & Johnson, Rob. (2019, September 30). *Insights into European research funder Open policies and practices*. Zenodo. <http://doi.org/10.5281/zenodo.3401278>, p9

¹⁴Open Science definition, FOSTER: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

Where survey questions were marked as ‘optional’, numbers of responses may vary. The text and figures indicate the total number of responses to each question, n.

The dataset generated and analysed during this study is available in the Zenodo repository under a CC0 licence: <https://doi.org/10.5281/zenodo.4153742>. We encourage you to reuse the dataset.

It contains no personal data. Data has been collected following the [SPARC Europe privacy policy](#). Data for part 2b, Governance and Sustainability have been kept confidential for those who requested it.

2.5 Acknowledgements

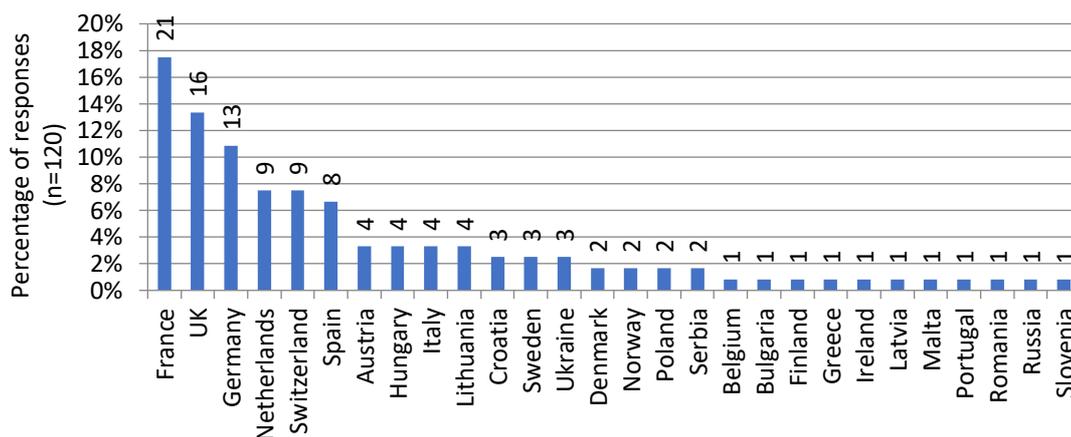
This report was prepared by Research Consulting on behalf of SPARC Europe. We are thankful to Vanessa Proudman for her leadership and guidance, and we would also like to thank the members of the project’s advisory group Jeroen Bosman, Mériam Cheikh, Bianca Kramer, Pierre Mounier and Kaitlin Thaney for their input. We would like to give special thanks to Bianca and Mériam for their contribution to the analysis undertaken.

3. Overview of survey respondents

This study consisted of two online surveys. Part 1 of the survey was distributed to OSIs across Europe between May and August 2020 and it received 120 responses. All these respondents were asked to complete Parts 2a and 2b of the survey, which were run between June and August 2020. A subset of these did so: receiving 67 and 68 responses, respectively.

Survey respondents were based in 28 countries. The five most common countries were France (21 out of 120), UK (16), Germany (13), Switzerland (9) and the Netherlands (9). The large French representation may be explained by the fact that the French national OS policy depends on a range of infrastructures being the first European country to list its OSIs in its Research Infrastructure national roadmap¹⁵. Eleven countries saw responses from a single OSI, which were often national in scope.

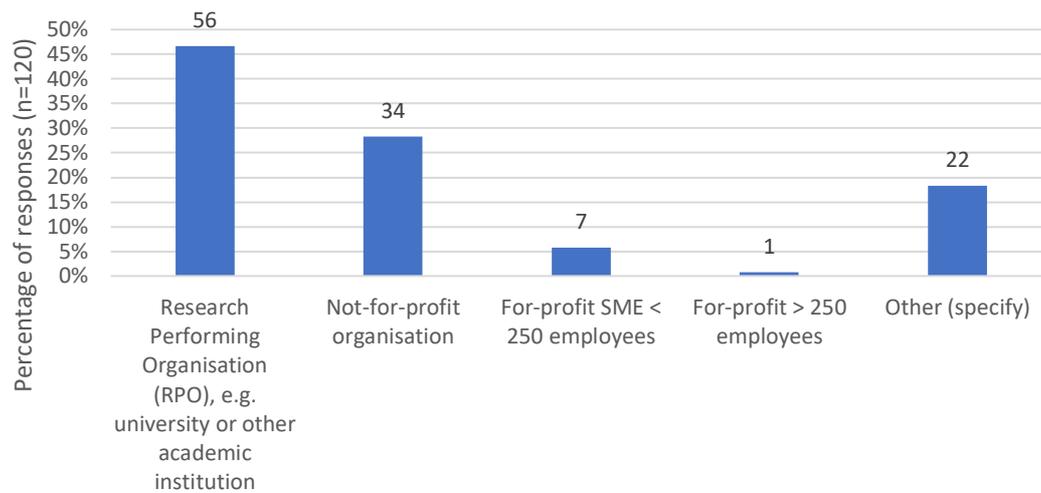
Figure 1. Survey respondents by geographical location (n=120)



¹⁵ French national strategy on research infrastructures, Edition 2018. Retrieved on 1 October 2020: https://cache.media.enseignementsup-recherche.gouv.fr/file/Infrastructures_de_recherche/04/6/Brochure_Infrastructures_2018_UK_1023046.pdf

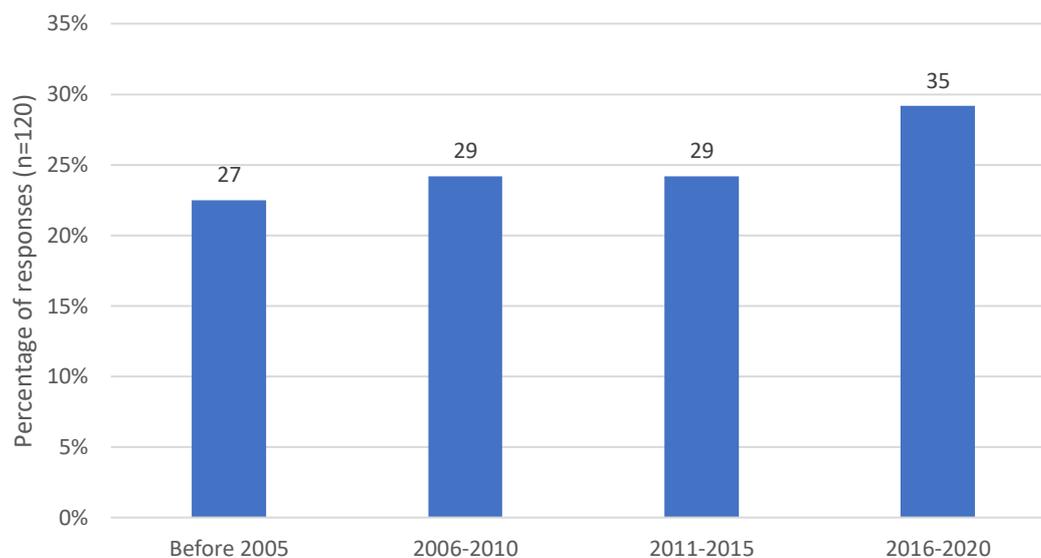
Almost half of respondents are Research Performing Organisations, i.e. RPOs or universities or other academic institutions (56 out of 120), whilst the second largest group is not-for-profit organisations (34) as seen in Figure 2. This shows the significant contribution that these types of organisations are making to Open Science. The third largest group, with 22 responses was the ‘Other’ category, which include a research infrastructure for funding organisations, a library consortium, an intergovernmental organisation and an R&D National Agency. Despite contributing significantly to the OSI market, there were very few for-profits which suggests that OSIs that responded to this survey are largely government or institutionally backed.

Figure 2. OSI organisation (n=120)



Most commonly, participating OSIs (35 out of 120) were established more recently in 2016-2020. The remaining 85 OSIs are still going strong with 27 of these being the more mature group established before 2005 as seen in Figure 3.

Figure 3. Year of establishment (n=120)



4. The OSI Landscape in Europe

4.1 In brief

This section summarises the key characteristics of OSIs in Europe, focusing on responses gleaned from Part 1 of the survey. Particularly, it focuses on the scope and purpose of OSIs, their support for OA and OS, their target market, maturity, and dependency on other infrastructure. It also looks at the challenges in being open as against the [COAR/SPARC Principles for Scholarly Communication Services](#).

Overall, the main findings are that:

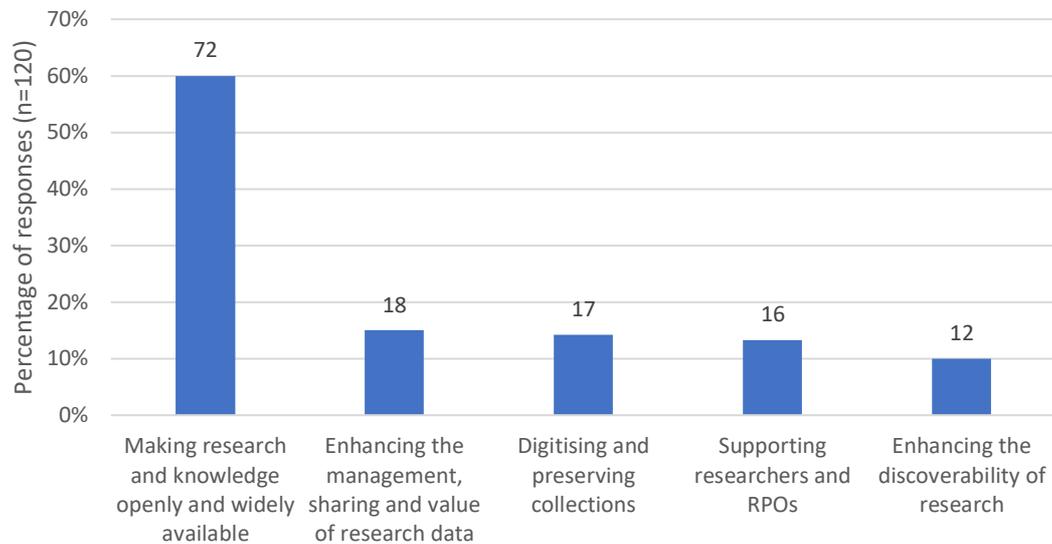
- OSIs are predominantly motivated by their ambition to further the vision of OA and OS by making research and knowledge openly and widely available (72 out of 120);
- researchers are the main target audience of OSIs (114 out of 118) with libraries coming in second and research managers, third;
- among the OSIs consulted, 95% of all OSIs provide services in three or more stages of the research lifecycle demonstrating that OSIs generally offer a comprehensive breadth of services.
- the most common activities central to the OSI ecosystem are aggregation & indexing, search, storage, identity (e.g. ORCID) and persistent identifiers underscoring the importance of sustaining the infrastructures that support them
- discovery and archiving are the most commonly supported types of services;
- the type of content OSIs most commonly provide access to are traditional research outputs such as journals (59 out of 116) and data (59) but also cover a variety of non-traditional and early research outputs;
- OSIs report maturity most in the open content and open standards COAR/SPARC principles. Effectively implementing good governance, sharing open content and applying and following open standards are the areas that see the greatest amount of challenges. While at the same time, most good practices are mentioned in the areas of open content and open standards.

4.2 Goals and scope of OSI

Goals

Respondents were asked to provide the overarching goals of the OSI they represented. By the means of qualitative coding, individual responses were categorised into themes, and the resulting top five are reported in Figure 4. For the vast majority of respondents, the most commonly cited goal was furthering the vision of OA and OS by making research and knowledge openly and widely available (72 out of 120), which was followed by enhancing the value of research data (18), digitising and preserving collections (17), supporting researchers and RPOs (16) and enhancing the discoverability of research (12).

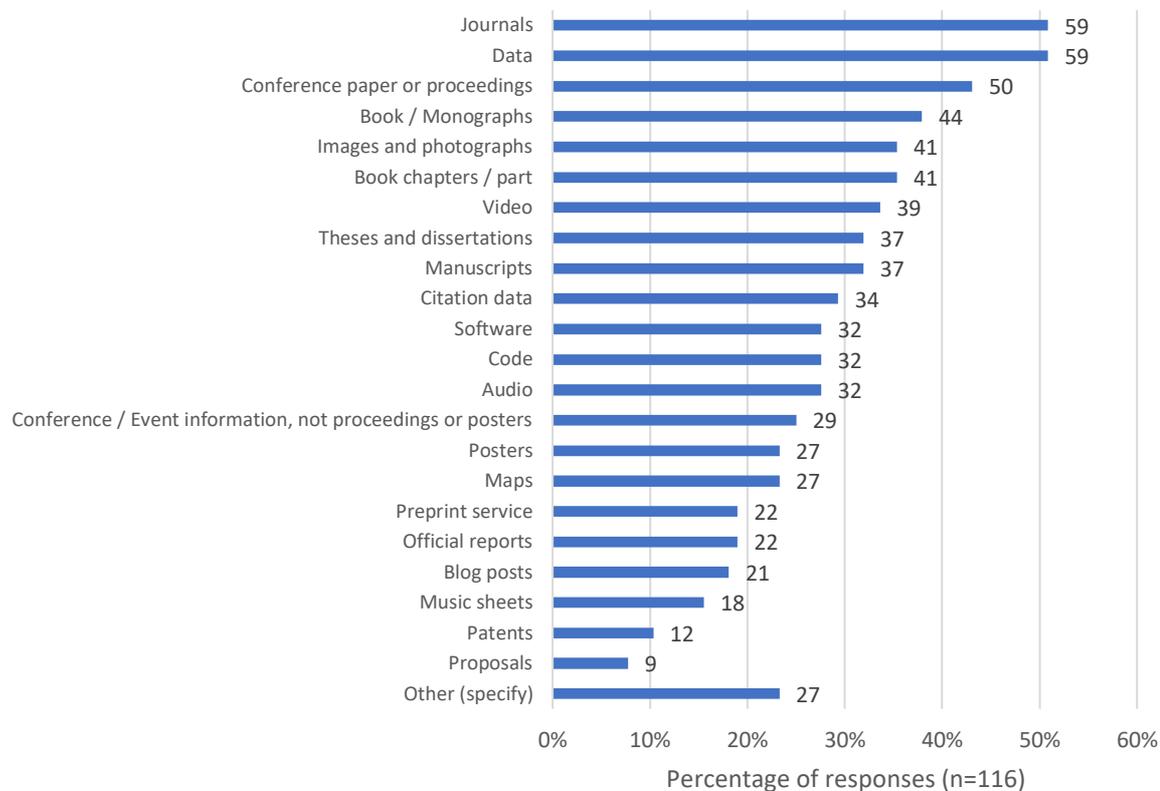
Figure 4. Goals of OSIs (n=120)



Services

Respondents were then asked a series of questions on the services provided by the OSI they represented in order to meet their goals and objectives (see Figure 5). Just over half of 116 respondents provide access to both journals and data, followed closely by conference papers or proceedings (50). This shows the significance of the OA conference paper to research, or at least its potential. Service analysis shows that conference papers are found most in Discovery services, Archiving/Preservation and Hosting and Access services. Books/monographs (44), images and photographs (41), and book chapters (41) are the next most commonly covered outputs. Only three respondents selected 'None'. This illustrates that whilst on the one hand traditional research outputs are most commonly provided by OSIs, they also provide access to a variety of non-traditional research outputs such as data (59), images (41), software (32), and code (32) and early research outputs, such as preprints (22), posters (27) and proposals (9). Preprints are served to a lower extent than other outputs with no notable differences between disciplines. A minority of respondents (27) selected 'Other' in response to this question, citing outputs such as data management plans, grey literature and learning objects.

Figure 5. Type of content provided by OSIs (n=116)



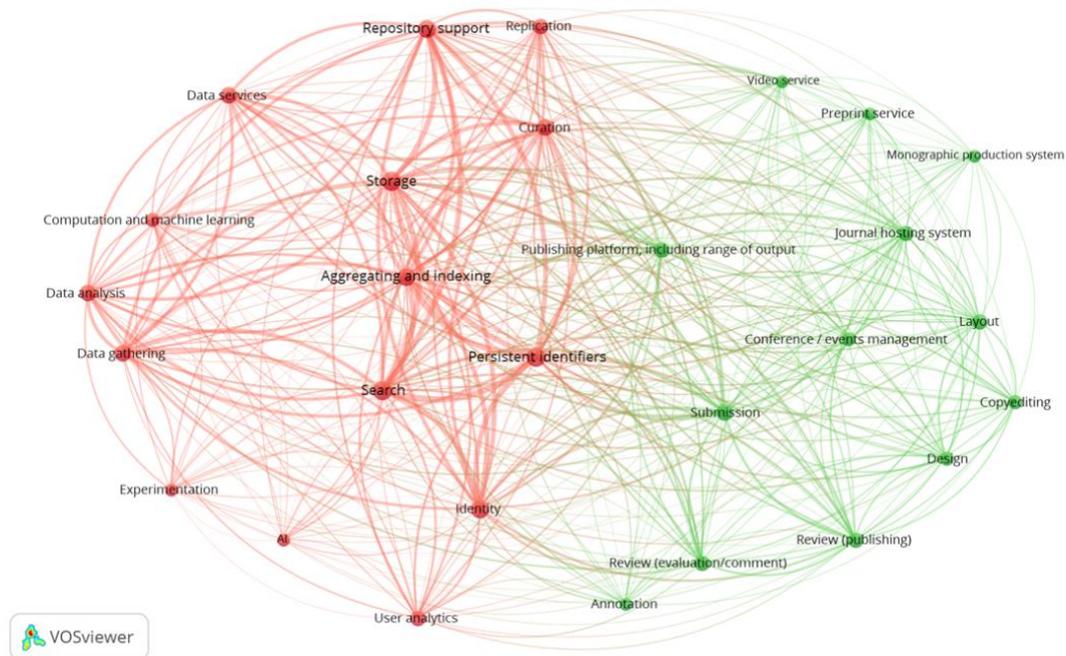
Respondents were asked which services the OSI provides across the research lifecycle, including the following stages with Discovery, Hosting and Archiving/Preservation in the top three:

- Creation** – supported by 71 OSIs (e.g. [Huma-Num](#), [Materials Cloud](#))
- Evaluation and commenting** – supported by 57 OSIs (e.g. [Plaudit](#), [Kriterium](#))
- Publishing** – supported by 70 OSIs (e.g. [HRČAK](#), [Ubiquity Press](#))
- Hosting** – supported by 97 OSIs (e.g. [OpenEdition](#), [OAPEN](#))
- Discovery** – supported by 100 OSIs (e.g. [OpenAIRE](#), [Open Knowledge Maps](#))
- Archiving/preservation** – supported by 85 OSIs (e.g. [Persée](#), [4TU.ResearchData](#))

Many OSIs indicate they provide services across multiple areas, with 95% (110) mentioning providing services in three or more stages. This suggests that many OSIs offer comprehensive services that extend beyond supporting one particular research lifecycle stage.

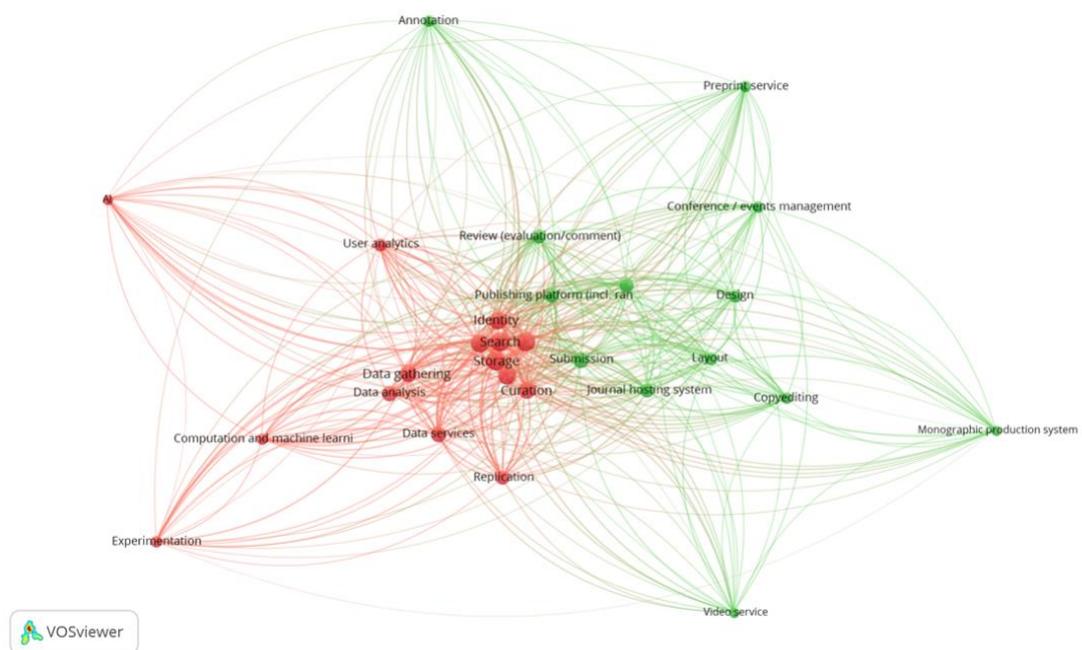
In an attempt to characterize the OSI landscape further, we looked at the specific activities OSIs indicate supporting within each stage. A network visualisation shows significant interconnectivity (i.e. OSIs support many different combinations of research activities) (Figure 6). Notwithstanding this, cluster analysis reveals two main clusters of activities, depicted in red and green: OSIs associated with publishing and hosting traditional text formats, and OSIs concerned with processing and storing research outputs, particularly data. Services concerned with search and persistent identifiers are included in the second group (likely due to the fact that this holds for all types of output), and the same goes for repositories.

Figure 6. Network visualization of activities supported by OSIs (created with VOSViewer)



A more condensed view of the network shows which activities are most central in the ecosystem of OSIs (**aggregation & indexing, search, storage, identity (e.g. ORCID) and persistent identifiers**) (Fig 7). These are mentioned by OSIs in combination with a wide variety of other activities – indicating they constitute central activities across the range of OSIs. This underscores the importance of these activities and the infrastructures supporting them and the need to sustain them.

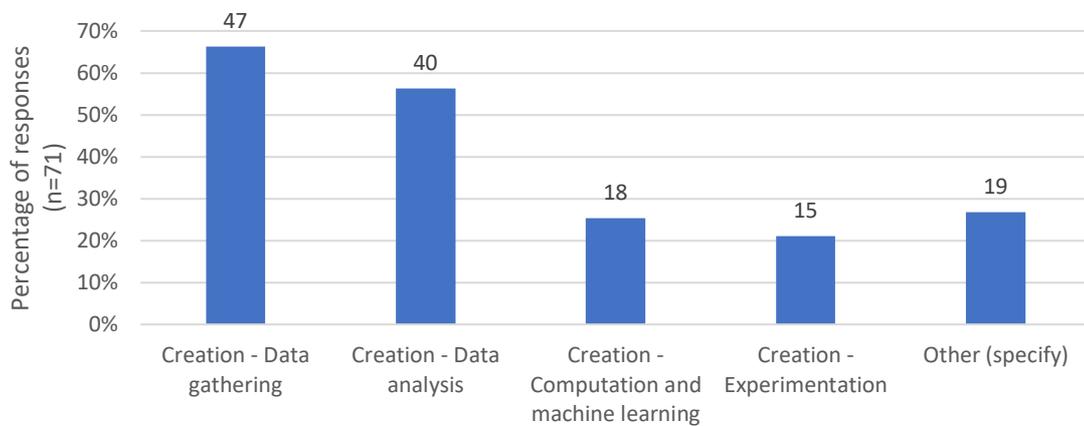
Figure 7 Network visualization of activities supported by OSIs (created with VOSViewer)



The most common activities cited by participating OSIs were enabling discovery and search services (79), storage for archiving and digital preservation (74) and persistent identifiers (72) showing where the concentration of OS infrastructure currently lies. A full breakdown of the activities supported in each stage in the research lifecycle can be seen in Figures 8-13.

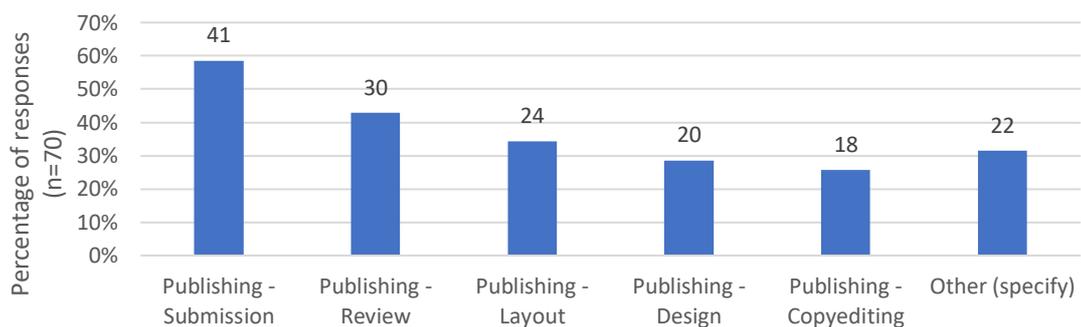
Of those OSIs providing support at the creation stage in the research lifecycle, data gathering (47 out of 71), and data analysis (40) were the most commonly reported activities. Computation and machine learning (18) and Experimentation (15) were roughly half as common. Nineteen responding OSIs selected 'Other' to highlight additional activities they provided, e.g. data management planning and data encoding.

Figure 8. OSIs which support the 'Creation' portion of the research lifecycle (n=71)



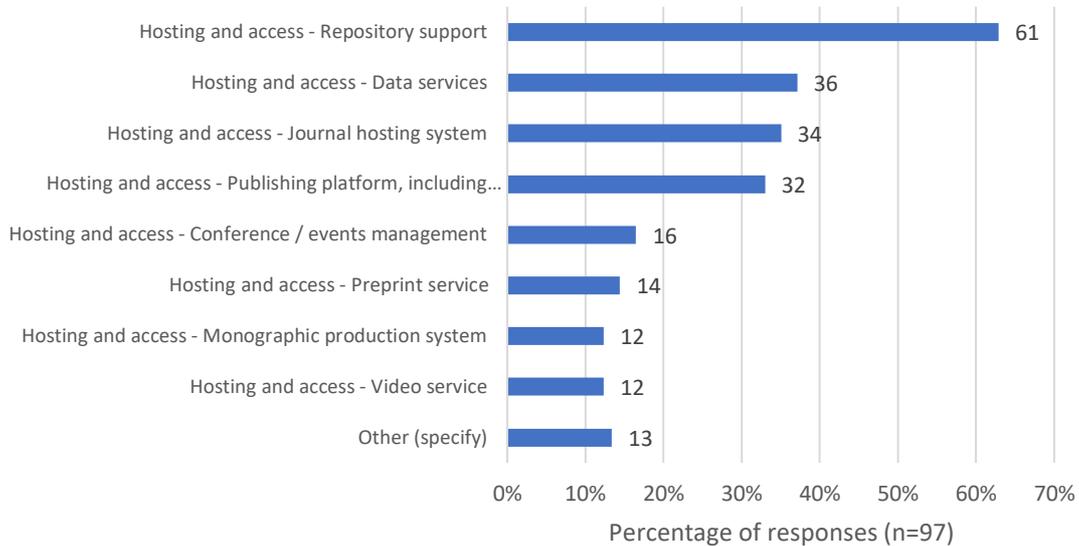
Of those OSIs providing support at the publishing stage, paper submission (41 out of 70) and review (30) were the most commonly reported activities. Twenty-two respondents selected 'Other', with a minority supporting additional activities such as providing tools for identifying trustworthy journals and books to publish in, plagiarism checks, dissemination, and training on editorial best practices.

Figure 9. OSIs providing 'Publishing' support (n=70)



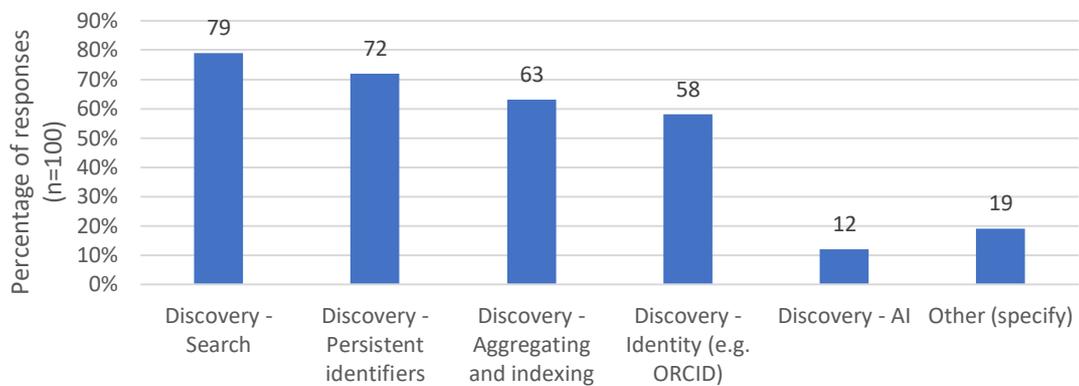
Of those OSIs providing support at the hosting and access stage, repository support (61 out of 97) was by far the most common activity, followed by data services (36) and journal hosting (34). Provision of monographic production systems (12) and video services (12) were the least commonly supported.

Figure 10. OSIs providing 'Hosting and access' (n=97)



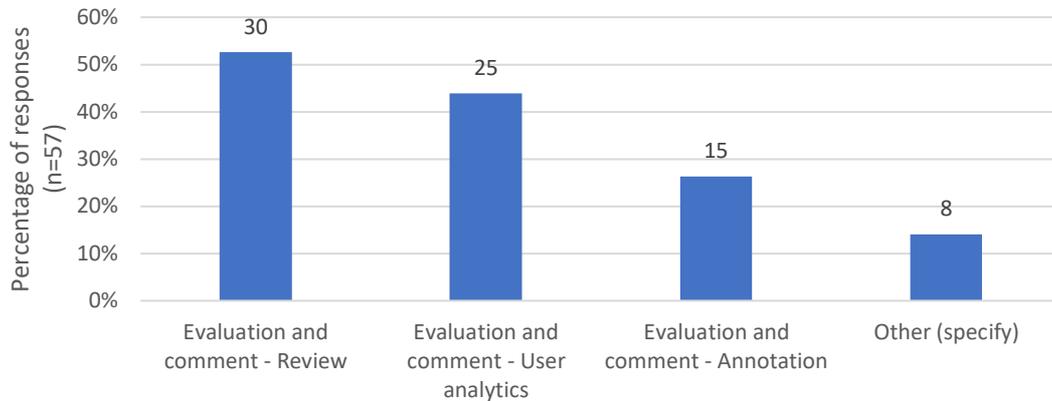
Of those OSIs providing support at the discovery stage, search (79 out of 100) and the provision of persistent identifiers (72) were the most commonly supported activities. Use of artificial intelligence (12) for discovery purposes was the least supported.

Figure 11. OSIs providing information 'Discovery' (n=100)



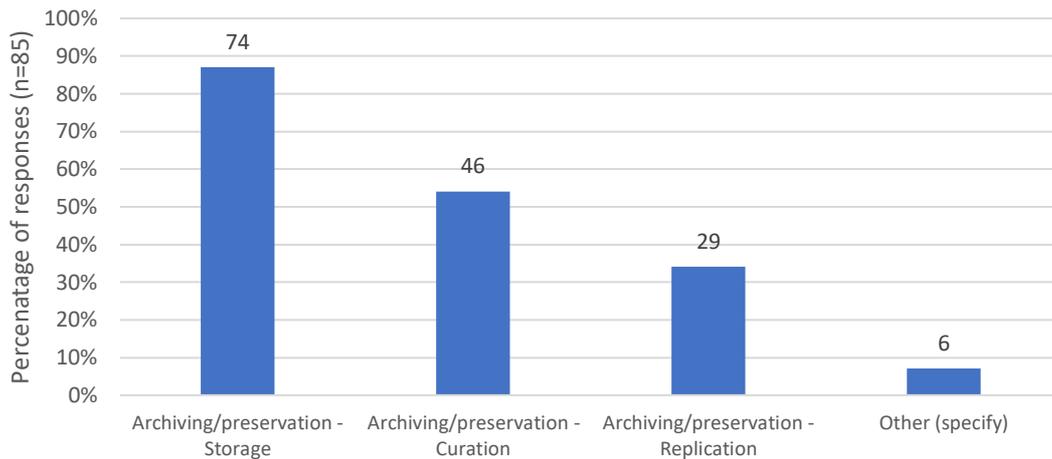
As seen in Figure 12, of those OSIs providing support at the evaluation and comment stage, over half tend to support review (30 out of 57) and nearly half provide user analytics (25). A further eight respondents selected 'Other', with one respondent noting that they offer users the option to give feedback on data quality, and another reporting that they provide certification.

Figure 12. OSIs providing 'Evaluation and comment' services (n=57)



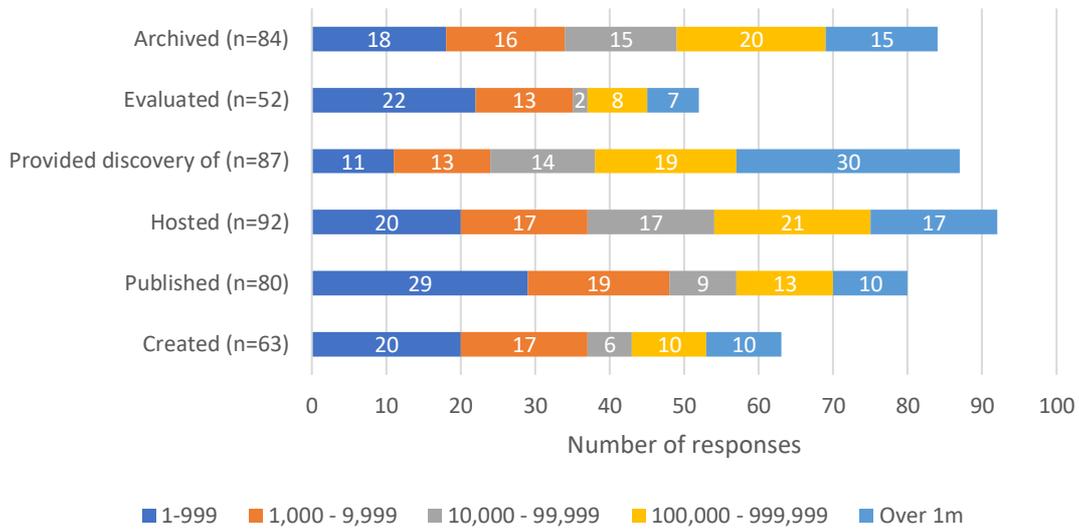
Lastly, of those OSIs providing support at the archiving and preservation stage, storage (74 out of 85) and curation (46) were the most commonly supported activities. A further six respondents selected 'Other', commenting that they were either not yet supporting this phase, or did so externally through their links to institutional repositories/libraries.

Figure 13. OSIs providing Archiving/Preservation (n=85)



As to the size of OSIs, respondents were asked to estimate the count of total objects (i.e. research outputs) that their OSI has created, published, hosted, provided discovery of, evaluated and/or archived as shown in Figure 14. Of 116 respondents, hosting and discovery are the two most common services provided by OSI, followed by publishing and archiving. Discovery of research outputs also has the greatest number of large-scale services with over 1 million objects (30), followed by hosting (17) and archiving (15). By contrast, small-scale services with fewer than 1,000 objects are concentrated in publishing (29), evaluation (22) and creation (20).

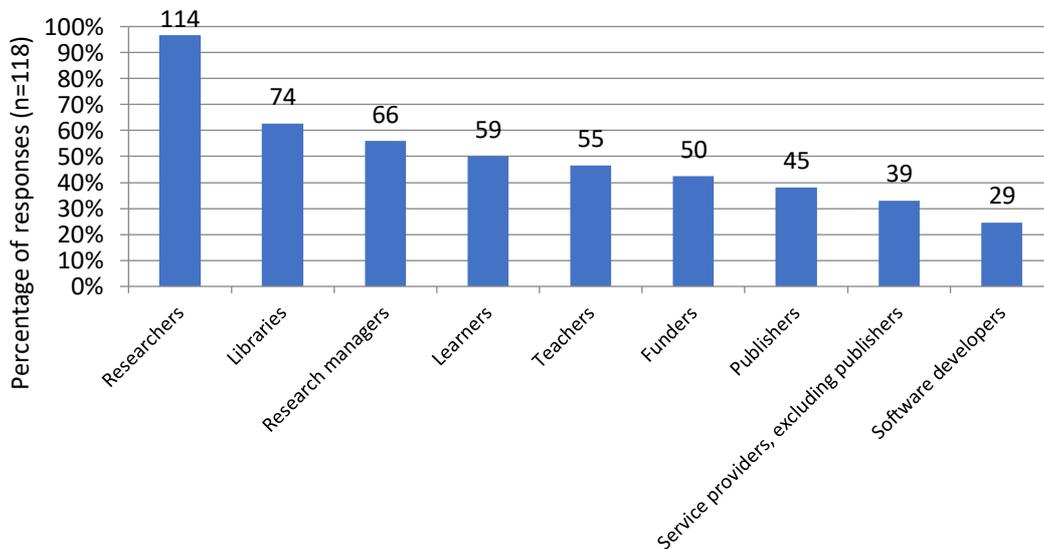
Figure 14. Estimation of total objects that OSI has created, published, hosted, provided discovery of, evaluated and/or archived (n=116)



4.3 Target audience

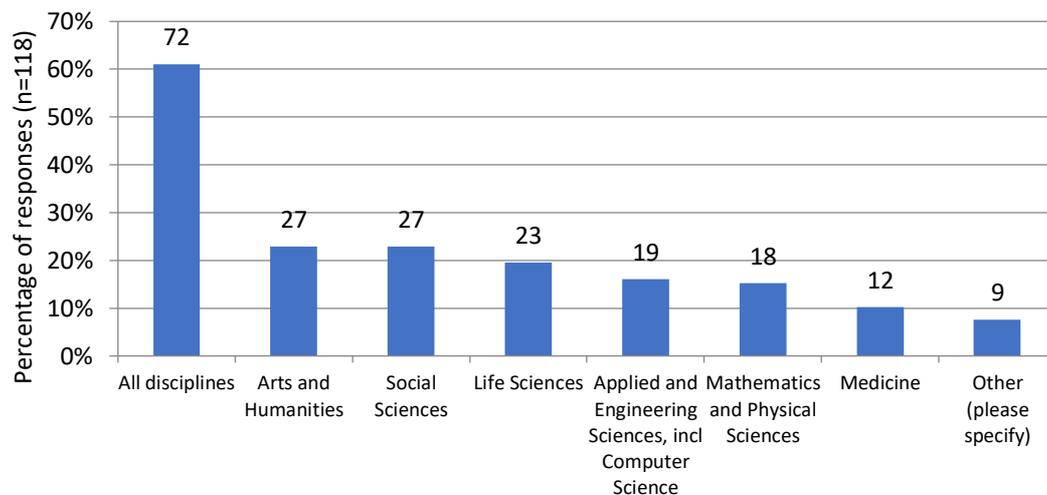
To gain an understanding of who is served by the current OSI offering, OSIs were asked a series of questions with regards to their target audience in terms of users, disciplines and geographical scope. As seen below in Figure 15, when reporting on the OSIs’ target audience (n=118), almost all OSIs report serving researchers (114 out of 118), followed by libraries (74) and research managers (66), who are often tasked with evaluating research. Teachers and learners are also popular target audiences, which indicates the benefits of OS for these communities. The least commonly served audience are software developers (29). This shows that OSIs target and serve a wide range of stakeholders.

Figure 15. OSIs Target audience (n=118)



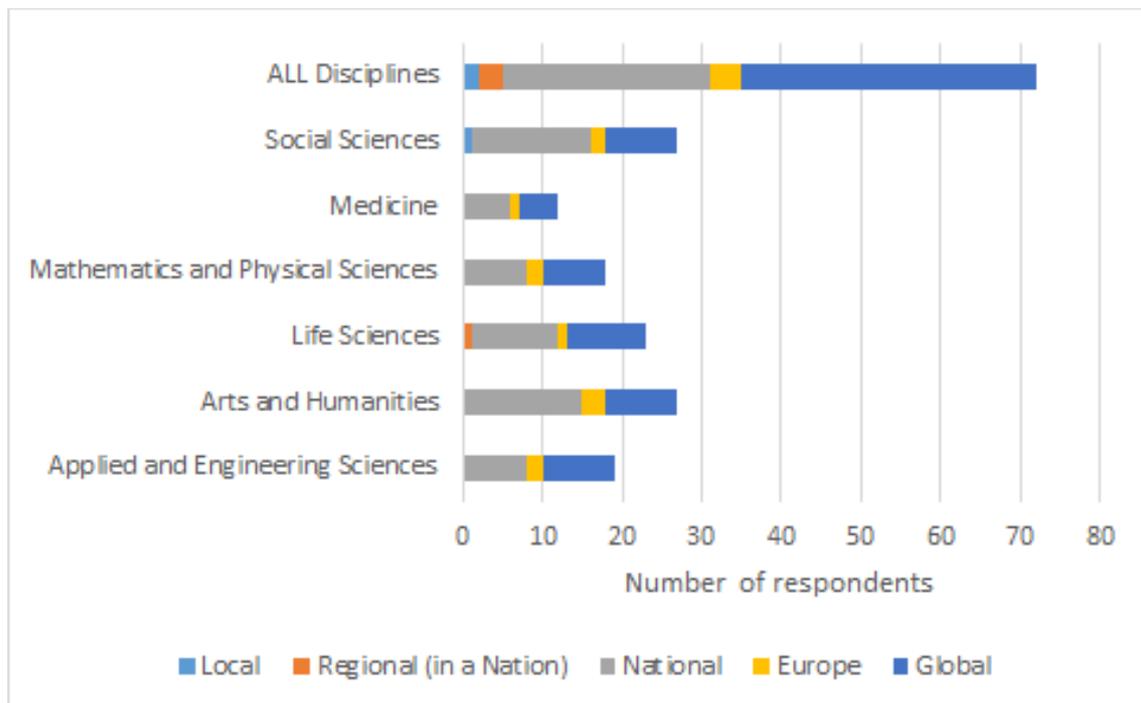
Honing in on the disciplinary focus of OSIs, Figure 16 shows that 72 out of the 118 OSIs report serving the full spectrum of subject areas. This shows that many OSIs are developed as generic systems that serve a multitude of domains or multidisciplinary research. The Social Sciences and the Arts and Humanities come in equal second place with 27 each. This might be explained by the fact that the survey was distributed widely by the OPERAS network. Few OSI respondents serve the medical sector.

Figure 16. Disciplines served by OSIs (n=118)



European OSIs most frequently operate at the global scale (57 out of 118) which clearly demonstrates the significant contribution that European OSIs are making to support the OS sector worldwide. Numerous OSIs have a national focus (46) showing the significant commitment made to OS infrastructure in European countries. OSIs operating at the national level tend to be based at research performing organisations (RPOs). Less than 10 OSIs reported serving only Europe, and less than five report serving only a regional and local audience. Figure 17 shows that there are no significant differences between the domain offering by geographical scope.

Figure 17. Disciplines served by OSIs by geographical region (n=118)



For the vast majority of OSIs, the primary language is English (97 out of 118). Other languages than English were reported in 131 cases. Three respondents noted that they had measures in place to ensure that their OSI could be translated into other languages.

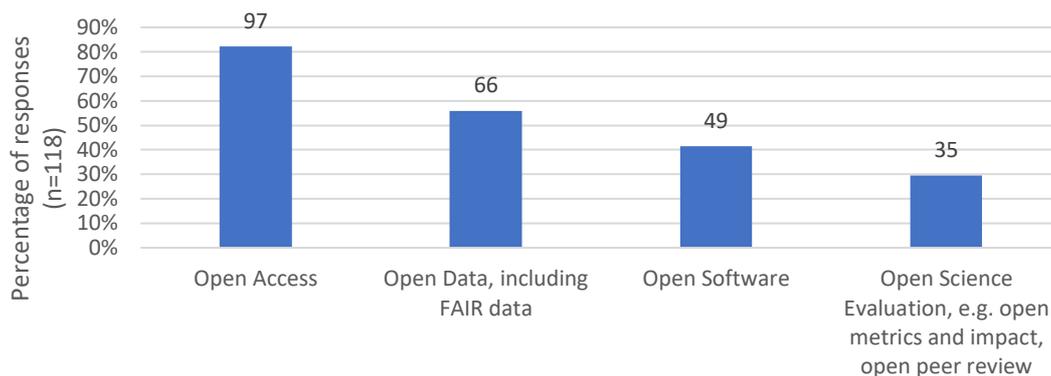
The language of the material provided by OSIs follows a similar pattern, showing that the primary OSI language is English with 77 out of 118 also providing materials in English. However, OSIs also reported material being provided in a range of languages in 217 cases demonstrating that OSIs provide access to a range of language content of local and international significance.

4.4 Support for OA and OS and the COAR/SPARC Principles

4.4.1 General

OSIs were asked as to how they supported Open Science. 97 of 118 responding OSIs support OA, 66 support open data, 49 support open software and 35 support open science evaluation (i.e. open metrics and impact, open peer review) which shows action to innovate in this important sector. Only two respondents reported no support for these areas.

Figure 18. Support of Open Access, Open Science or Open software (n=118)



4.4.2 How far the COAR/SPARC Principles for Scholarly Communication Services apply to OSIs

We have seen a range of principles for scholarly communications discussed and developed over the last few years. Many are based on the Bilder, Lin and Neylon Principles for Open Scholarly Infrastructures.¹⁶ In 2017, Paul Peters of Hindawi also discussed an open approach to developing infrastructure for Open Science in favour of a radically open approach to prevent private companies from owning and controlling OS infrastructure.

In that same year, COAR and SPARC published a joint statement pledging that they would work to ensure that scholarly communication is better aligned with research goals. In 2018 they developed the Principles for Scholarly Communication Services, and published them in January 2019.¹⁷ Figure 19 shows a palatable set of seven principles that encourage services to improve their open practices, follow good governance, be transparent and help services align with the aims of scholarship. This study explores openness using the lens of these principles also to understand the challenges in applying them in practice.

¹⁶ Idem

¹⁷ Principles for Scholarly Communication Services <https://www.coar-repositories.org/files/COAR-SPARC-Good-Practice-Principles.pdf>

Figure 19.: COAR/SPARC Good Practice Principles for Scholarly Communication Services

Good Practice Principles for Scholarly Communication Services

COAR and SPARC have developed seven good practice principles to ensure that scholarly communication services are transparent, open, and support the aims of scholarship. These principles can be used by users to make decisions about which services they will contract with, and by service providers to improve their practices and governance

GOOD GOVERNANCE

The service has strategic governance that allows community input on the direction of the service and operational governance with community representation and decision making power.

TRANSPARENT PRICING AND CONTRACTS

The service's contract conditions and pricing are transparent and equitable, with no non-disclosure agreements included.

EASY MIGRATION

User-owned or generated content can be easily migrated to another platform or service upon termination of contract, without any additional fee from the service provider.

OPEN STANDARDS

The service uses open APIs to enable interoperability, and adheres to open standards. Ideally, the platform is based on open-source software, but in cases where it is not, user-owned content is managed according to well-established, international standards.

SUCCESSION PLANNING

If the service is a nonprofit, the organization's bylaws state the conditions and terms governing how the organization may be transferred or wound down. If the service is provided by a for-profit entity, the contract/agreement should not be assignable to another entity without the client's express permission.

FAIR DATA COLLECTION

Only data necessary for the service's provision are collected from users and the type of the data collected and how they are used is clearly and publicly articulated.

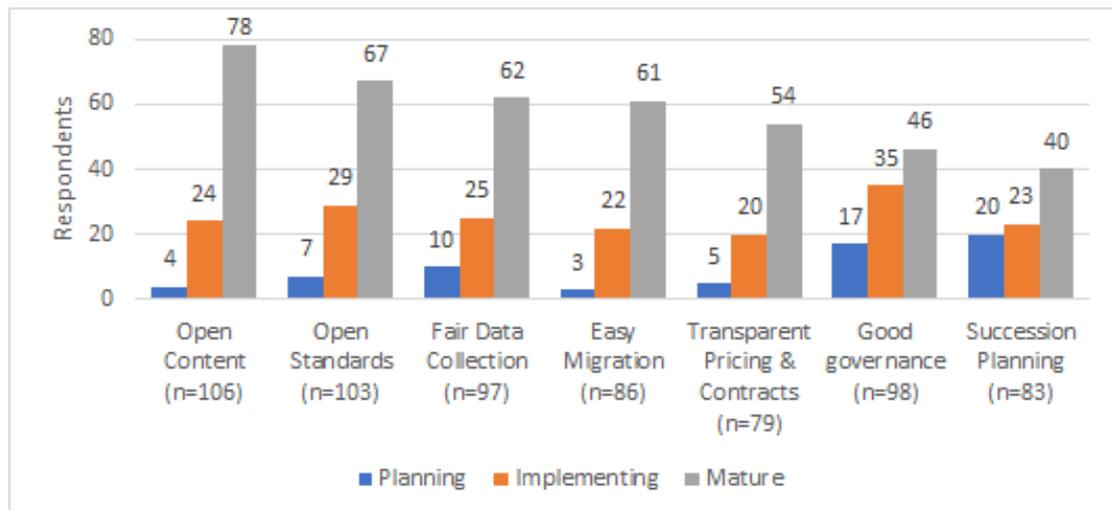
OPEN CONTENT

Content, metadata and usage data are immediately, openly and freely available in machine-readable format via open standards, and using licenses (like CC0 or similar) which facilitate reuse.

These principles are informed by Principles for Open Scholarly Infrastructure-v1 by Bilder G, Lin J, Neylon C (2015) © 2019 COAR and SPARC, subject to a Creative Commons Attribution 4.0 International License

Respondents were asked about their estimated level of maturity with regards to the Good Practice Principles. Across the range of principles, as seen in Figure 20, more OSIs report being at the mature stage, with the top scoring areas being open content and open standards. The areas with the lowest levels of maturity are governance and succession planning.

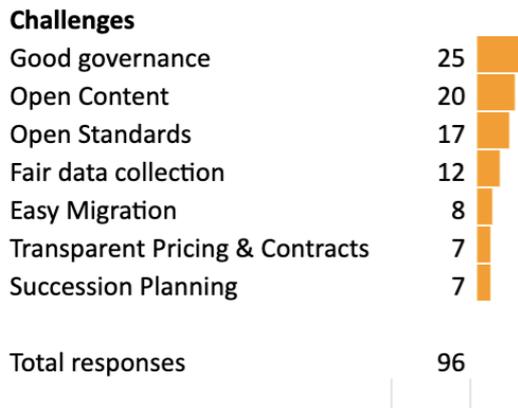
Figure 20. Status of maturity against SPARC/COAR principles



4.4.3 Challenges reported by OSIs in relation to the COAR/SPARC Principles

Respondents were asked about the challenges they encounter on various levels of ‘open’ referring to the Good Practice Principles. Forty-six respondents report a total of ninety-six challenges. Through qualitative coding, themes were identified from the individual responses. The majority of challenges reported related to good governance followed by open content and open standards as shown in Figure 21. A more detailed breakdown of the challenges can be found below.

Figure 21. Number of reported challenges by COAR/SPARC Principle and frequency



Good governance (25)

As stated in the Principles: *“The service has strategic governance that allows community input on the direction of the service and operational governance with community representation and decision-making power.”*

The majority of challenges (9) relate to **ensuring sound, equitable and community-relevant decision-making**. One specific challenge identified was the tension between serving the needs of the community of users versus prioritising the needs of clients that provide financial support to the OSI. Matching needs with management principles was also mentioned as a challenge. Some share concerns with setting up and working with various new board structures.

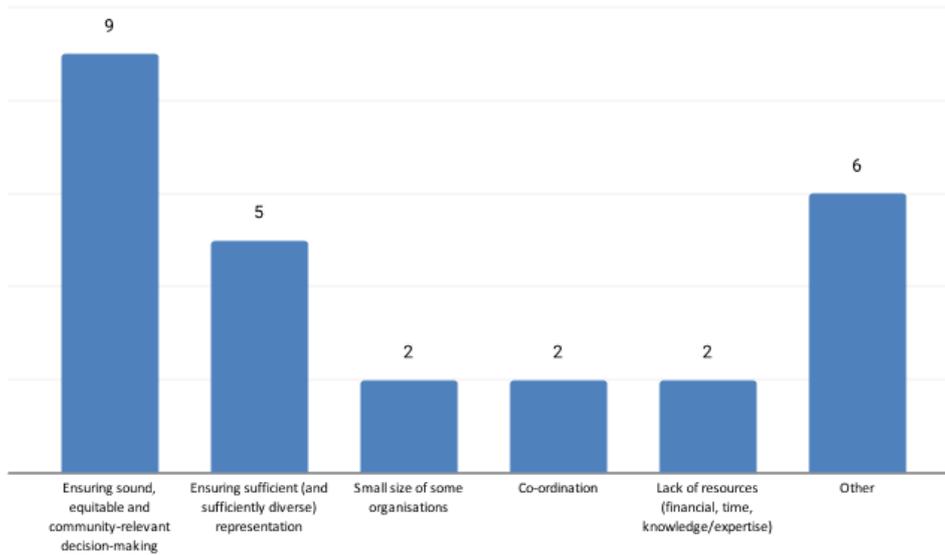
Five respondents report challenges around **ensuring sufficient (and sufficiently diverse) representation**. This included concerns around the need for broader stakeholder engagement or more diverse representation, getting the right representatives of the (research) community or having certain groups more involved such as researchers and libraries. One respondent also stated the concern about ensuring sufficient member commitment in decision-making meetings.

Challenges regarding governance in relation to size are felt at both ends: with a couple of one-person initiatives experiencing barriers in the initial establishment of governance structures due to their small size and limited resources. Larger initiatives often mention challenges regarding co-ordination, negotiating with existing power structures and ensuring inclusive representation.

Challenges mentioned by other respondents included the **co-ordination** of organisations when multiple organisations are involved (2) and the **lack of resources** as reasons for either not having a governance structure or for managing a formal incorporation (2).

Six respondents reported a range of other challenges, including a lack of time, the legal structure of the organisation and the effort needed to agree on governance terms and values.

Figure 22. Reported challenges in good governance (n=25)



Open standards (17)

As stated in the Principles: *“The service uses open APIs to enable interoperability, and adheres to open standards. Ideally, the platform is based on open-source software, but in cases where it is not, user-owned content is managed according to well-established, international standards.”*

Respondents share a diverse range of challenges related to this principle.

Five respondents report resource issues related to APIs and open source including concerns related to the **time, effort, expertise and costs required to build or switch to open source solutions or provide APIs**. Three **question the value of the API and its actual (re)use** showing the need for information on how the infrastructure is used.

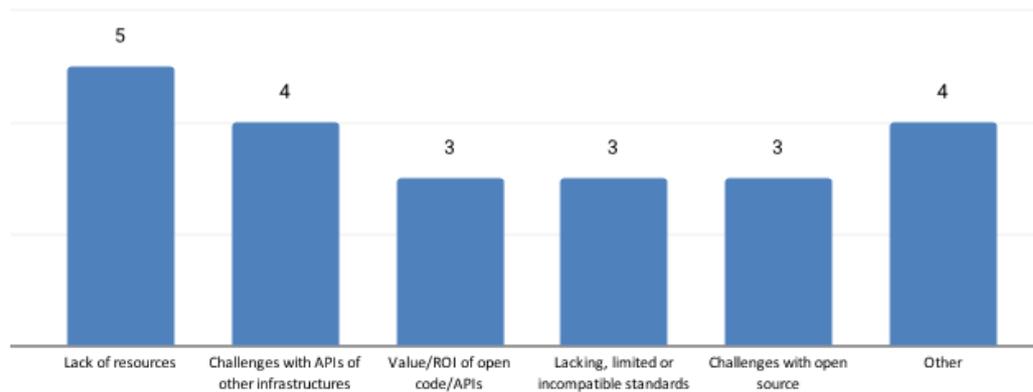
Three respondents report **concerns around standards**: a lack in compatibility between existing standards, missing and incomplete standards or keeping up with changing standards. Issues with standards prevent access to information.

Four respondents mention **challenges with the APIs of other infrastructures**, including term disambiguation, the need to work with different data types and changing APIs causing broken imports. One respondent was disappointed by the fact that no API existed to enrich their metadata sufficiently.

Three respondents report **challenges with open source**: finding it difficult to find open source solutions for one. A respondent also shared the frustration in not being able to open the code of in-house developed tools and services and one stating that this was due to how their services are funded.

Four respondents mentioned a range of other concerns including having to work with aged systems or reporting that an OSI does not use open source software. One comment also mentioned politics as an inhibitor to open standards.

Figure 23. Reported challenges in Open Standards (n=17)



Fair data collection (12)

As stated in the Principles: *“Only data necessary for the service’s provision are collected from users and the type of the data collected and how they are used is clearly and publicly articulated.”*

Responses to this question seem to indicate that the question is often considered by respondents to be about FAIR data (Findable, Accessible, Reusable and Interoperable), rather than about fair data collection in the afore-mentioned definition, and as highlighted in the survey but relevant to data collection in general.

Several respondents point out that (especially in the case of user-owned or -generated content) that **user data policies are set by publishers which limits what can be made available** resulting in some closing off data.

Three respondents mentioned challenges around **personal data collection or GDPR-compliance** and another three cited **copyright issues**. In one case, the need for legal support to formulate official terms and conditions was articulated.

One OSI also specifically mentioned the need for implementing an authentication system to allow restricted access to data.

Figure 24. Reported challenges in fair data collection (n=12)



Transparent pricing and contracts (7)

As stated in the Principles: *“The service’s contract conditions and pricing are transparent and equitable, with no non-disclosure agreements included.”*

Seven OSIs point out that their service is, and will remain, free of charge, so this principle does not apply to them. Some highlight their specific funding model (e.g. institutional or state funding) and/or their challenges in securing (and disbursing) such external funding.

Of the OSIs that do charge fees, challenges reported were mainly in the area of **setting fair and/or transparent fees** including setting fees based on the differing size of institutions and needs of clients, since fixed fees are not usually deemed feasible or appropriate. In one case, these fees are determined by others (e.g. APCs determined by participating journals or societies). Using transparent membership fee tiers is named by one respondent, and the challenge here is determining what an equitable tier system looks like. This echos concerns shared by another respondent, who mentions the difficulty in evaluating real costs. A further respondent states that their business model is currently being defined.

Two respondents also mention **negotiation** as a challenge: either to gain access to data or to set pricing or contracts since client organizations vary in size, with different needs.

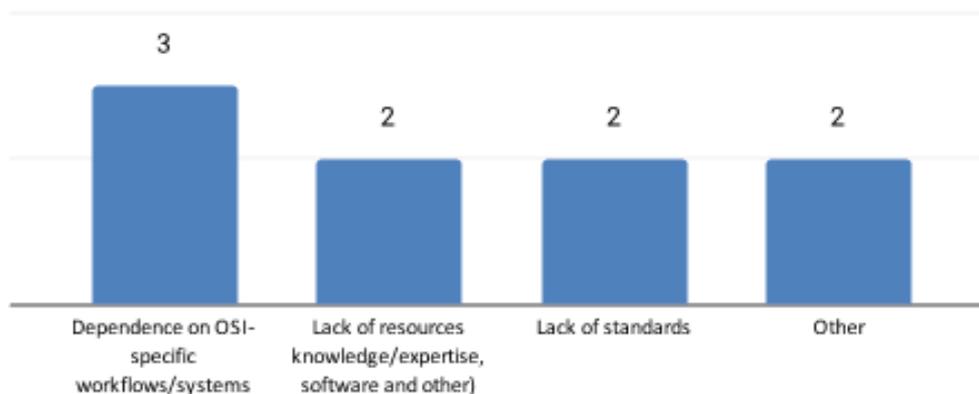
Easy migration (9)

As stated in the Principles: *“User-owned or generated content can be easily migrated to another platform or service upon termination of contract, without any additional fee from the service provider.”*

Whereas some OSIs consider content migration as out of scope, as they ‘only provide software and storage space’, three respondents cite the dependence on OSI-specific workflows/systems as a challenge: **the use of centralized or self-built systems and the resulting limitations to data portability or integration with other systems**. One respondent reports requiring ‘a lot of supporting software’ to run their OSI. Another raises issues with older (proprietary) formats or software. For one OSI, it is not clear where users could migrate their specific content to.

Four respondents mention more general issues such as a **lack of resources as a challenge (2), including expertise, software and other resources**. Two further respondents raise a lack of standards as an issue for some domains or generally for the migration of data between data repositories.

Figure 25. Reported challenges in Easy migration (n=9)



Succession planning (7)

As stated in the Principles: *“If the service is a nonprofit, the organization’s bylaws state the conditions and terms governing how the organization may be transferred or wound down. If the service is provided by a for-profit entity, the contract/agreement should not be assignable to another entity without the client’s express permission.”*

Two OSIs indicate that succession planning is formally documented in statutes or bylaws. One respondent reports that **despite succession planning being documented, they question how it works in practise** also with a clear reliance on the importance of informal, personal networks. Another respondent states that as a small organisation, **upscaling (which would require resources)** would allow for more robust succession planning. One respondent confesses that no succession planning was in place whilst another cited it as a sensitive subject.

Note that other responses to this question primarily concern continuity of the OSI itself, rather than in transferring or winding down operations when needed. This might indicate that there is little experience with succession planning, which can be a concern for users.

Open content (20)

As stated in the Principles: *“Content, metadata and usage data are immediately, openly and freely available in machine-readable format via open standards, and using licenses (like CC0 or similar) which facilitate reuse.”*

The main challenge reported is that in six cases, OSIs cite that they **do not themselves decide on the openness of content since they are dependent on the policies of content providers**, e.g. publishers or researchers. It is also reported that researchers are not used to sharing data openly and in machine-readable form, which is expected to take time to change.

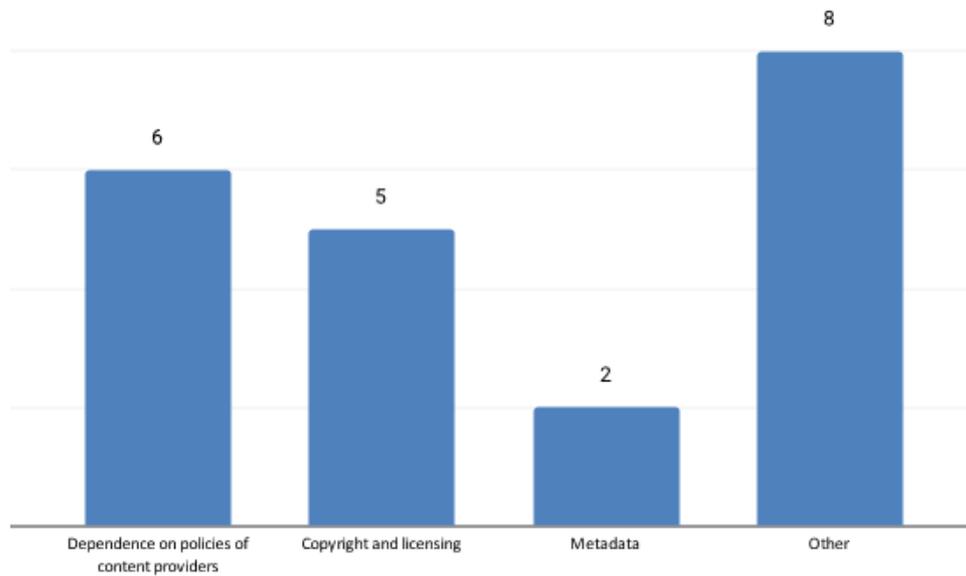
Five respondents raise **copyright and licensing** as a demanding and challenging area, although certain publishers call for a Creative Commons (CC) licence. One OSI reports only allowing reuse under a CC-BY-NC-ND licence. Related challenges include how authors and publishers agree on the licence. Another concern relates to few researchers being willing to publish OA or assign a standard licence to the content shared. Two respondents report apathy from researchers towards either assigning licences to publications or to sharing data openly and in a machine-readable format.

Another respondent reports that although the platform might be openly licensed, the licence status of the content aggregated is either mixed or unclear. This prevents data service providers from openly licensing that data the way they want to or it prevents them from providing data dumps. A further respondent reported that they are not ready to provide metadata under the CC0 license in fear of having their work discredited although they are working to resolve this issue.

Two respondents share challenges with **metadata**. One respondent is concerned about the quality and completeness of metadata they require, as well as a lack of open references. Another OSI mentions that it is not able to share all metadata via OAI-PMH unless they are fully compliant with the service’s standards.

Where OSI’s do have control (or can set policies) over their data, a small handful of respondents express **concerns over privacy (including GDPR) and fear (or previous experience) of misuse**. One respondent shares the issue of linking publications to research data. And one service reports that despite opportunities to do business with non-open publishing, they have declined this to stick to their values.

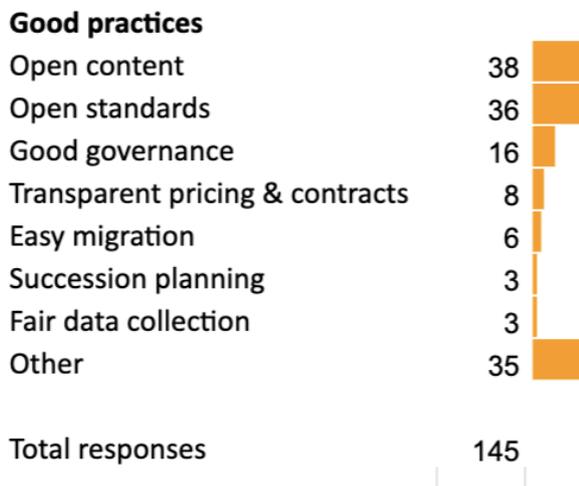
Figure 26. Reported challenges in Open content (n=20)



4.4.4 Good practices reported by OSIs in relation to the COAR/SPARC Principles

Good practices were reported by 71 respondents resulting in 145 references to good practices. The principle that saw the most good practices was open content, followed by standards and then good governance as seen in Figure 27. A more detailed analysis of good practices will be shared in a separate publication related to Open OSI Champions out in late 2020.

Figure 27. Number of reported good practices by COAR/SPARC Principle and frequency (n=145)



5. Technology

5.1 In brief

This section summarises the technological aspects of OSIs in Europe, focusing on responses gleaned from Part 2a of the survey. Particularly it focuses on the technical management of OSIs, integration with external systems, API availability, open source, compliance with FAIR principles, and compliance with EOSC and Plan S and technical standards and requirements.

Overall, the main findings with regards to technology are that:

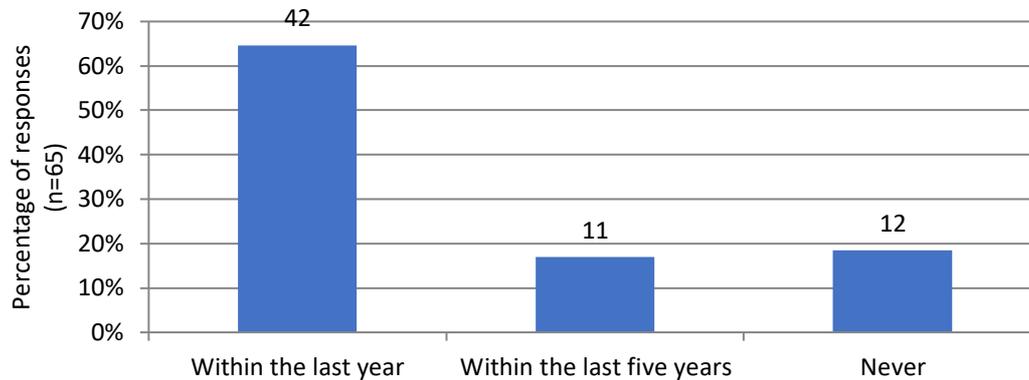
- the majority of OSIs (55 out of 66) have a dedicated technical lead and have formally evaluated their technical environment within the last five years (53 out of 65) showing that they are up-to-date and innovating;
- the majority of OSIs (57 out of 67) integrate with external systems and services; the overwhelming majority of these are run by not-for-profits or RPOs. The most commonly cited are ORCID, Crossref, DOAJ, BASE, OpenAIRE, Altmetric, and Datacite. Infrastructures that support persistent identifiers are mentioned by 45% of the cohort showing their importance as critical infrastructure;
- 50 out of 65 responding OSIs have an API and common use cases are data harvesting (12 out of 46) and metadata exchange (7);
- over half of OSIs are either fully (34 out of 64) or partly (19) built with open source code showing how essential good collaboration is to the success of an OSI. Over half (33 out of 53 respondents) have code contributors with most OSIs having between one and five code contributors (14 out of 27) which shows some fragility in sustainability;
- over half of all OSIs comply with *some* open standards (36 out of 61). Another large proportion *only* use open standards (23) which shows the strong adoption of open standards by OS infrastructures;
- the vast majority of respondents (45 out of 54) report that they comply with FAIR data principles showing the strength of the FAIR standard (where relevant);
- most OSIs meet the EOSC service requirements and Plan S mandatory technical conditions which demonstrates that many OSIs are adaptable and up-to-date with key OS community standards.

5.2 Management

Respondents were asked questions around the management and continuous assessment of their technological solutions. The majority of OSIs (55 out of 66) report that they have dedicated technical leads.

The majority of respondents have also evaluated their technical environments either within the last year (42 out of 65) or the last five years (11), as seen in Figure 28 showing that most OSIs have an interest in not becoming obsolete by keeping up with the state of the art and by innovating. However, we note that 12 respondents have never formally evaluated their technical environment.

Figure 28. Last formal evaluation of technical environment (n=65)



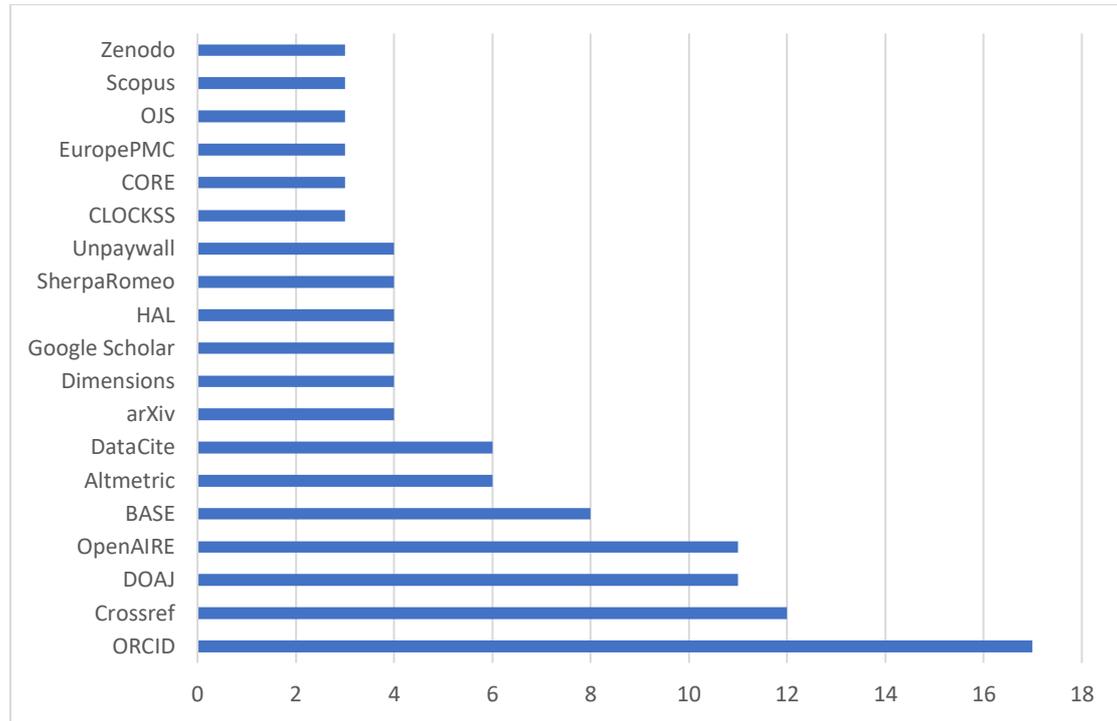
5.3 Dependency on other infrastructure

The vast majority of OSIs (50 out of 65) have APIs, and typical use cases include data harvesting (12 out of 46) (e.g. for analysis or discovery purposes) and metadata exchange (7).

Most OSIs (56 out of 67) report that they integrate with external systems or services, 57 OSIs provided free text responses discussing the external systems and services they integrate with, whether to provide information to or to harvest information from. As shown in Figure 24, the most commonly noted were ORCID (17 out of 56) and Crossref (12). This highlights the importance of persistent identifiers (PIDs) as a critical layer for interoperable OS services. Together with DataCite (6), these PID providers are mentioned by 25 of 56 respondents (45%).

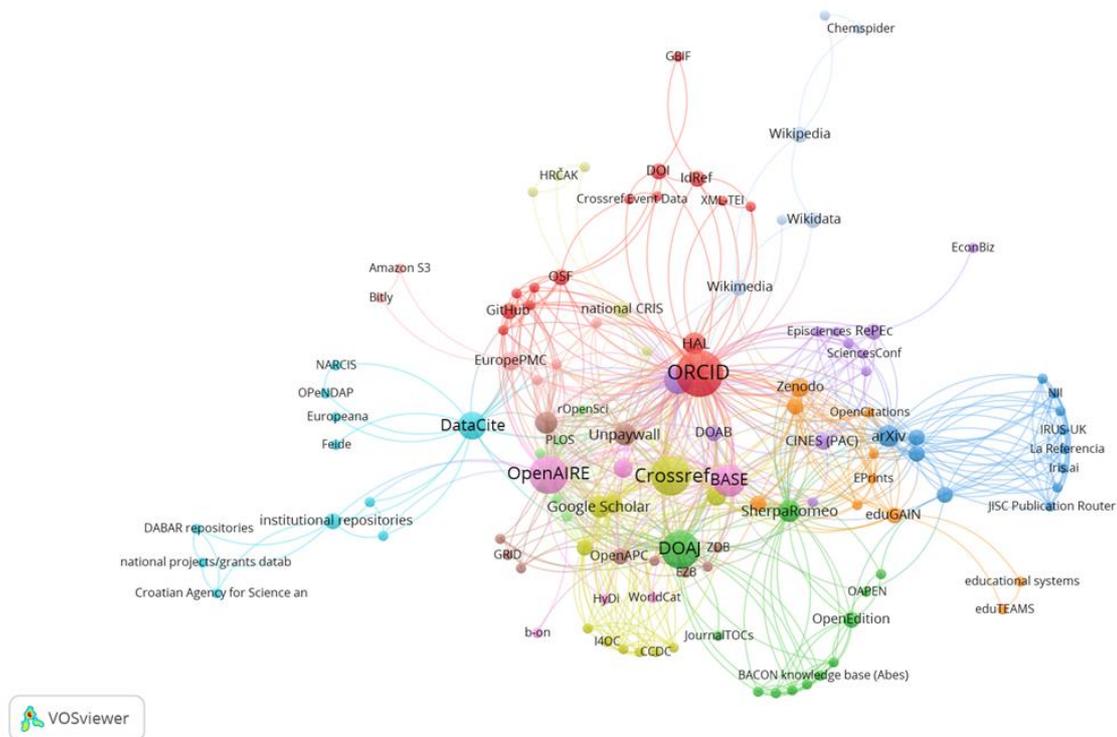
Of the tools and services that three or more OSI's integrate with (n=19), the majority is either run by a not-for-profit organization or a research performing organization, and only four are for-profit (Altmetric, Dimensions, Google Scholar and Scopus). This highlights the important role of not-for-profit organizations and the infrastructure they provide in the open science landscape.

Figure 29. External systems and services that OSIs commonly integrate with (n=57)



A network visualisation of all tools and services that the OSIs in our survey report interoperate with (Fig 30) again shows the centrality of services like ORCID, Crossref, OpenAIRE, DOAJ and BASE, and also shows more distinct clusters of services that are used by specific OSIs or smaller groups of OSIs. Examples of such clusters are: services built on Wikimedia (Wikipedia, Wikidata), a cluster including services provided by GÉANT such as eduGAIN and eduTEAMS (linked with educational systems) that also includes other European infrastructure like Zenodo and EOSC. Other examples are a cluster consisting of institutional and national repository services that is specifically connected to DataCite and OpenAIRE, and a cluster of national infrastructures in a specific country (Croatia). Overall, the results make clear that OSIs do not operate as single entities, but rely on (or are relied on by) other infrastructures in an interconnected web of services that provide researchers with support for their research activities.

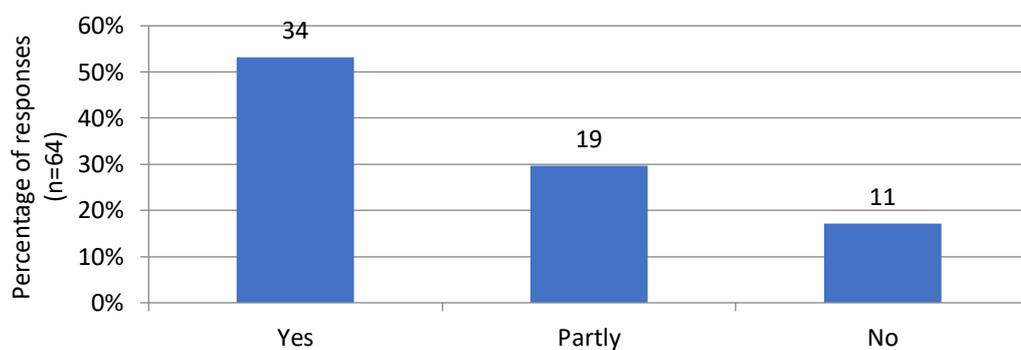
Figure 30. Network visualization of tools and services that OSIs interoperate with (created with VOSViewer)



5.4 Open source code and community maintenance

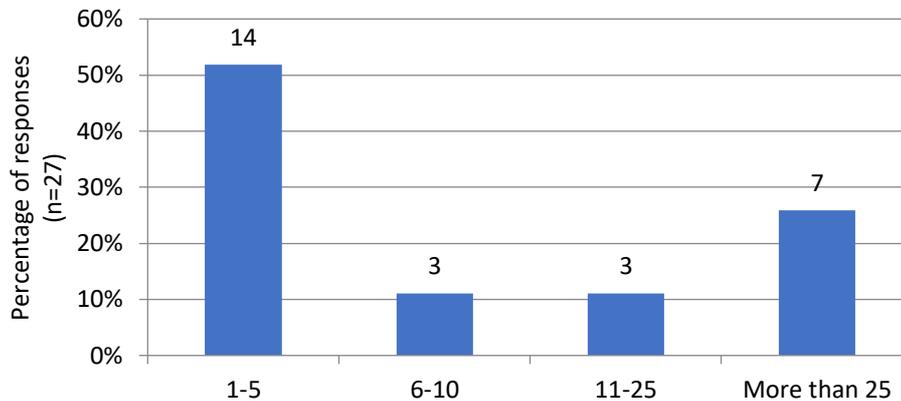
Figure 31 shows that the majority of responding OSIs are either fully (34 out of 64) or partly (19) built with open source code. Four of the five for-profits reported their software being partly open source, with none being fully open source. RPOs are the most likely to go fully open source (17), with 11 not-for-profits coming in second and six others coming in third demonstrating more of a commitment to full open source in research performing organisations and not-for-profits. This clearly demonstrates the strong adoption of open source and it becoming a norm of sorts. 33 out of 53 responding OSIs note that others can contribute their code to the OSI. In practice, this is enabled by sharing guidelines (16 out of 30) and/or by using GitHub pull requests (9). Open source OSIs were asked to provide information on the licence, if any, they used in sharing code: Apache 2.0 and MIT were the most commonly cited.

Figure 31. Is code OSI is built with open source? (n=64)



With regards to the numbers of code contributors, Figure 32 shows that OSIs report between one and five in 14 out of 27 cases which demonstrates the fragility of code production in some cases. However, a large minority of OSIs (7) noted that they have over 25 code contributors, including large data repositories, new data software, publishing platforms, and a bibliographic service.

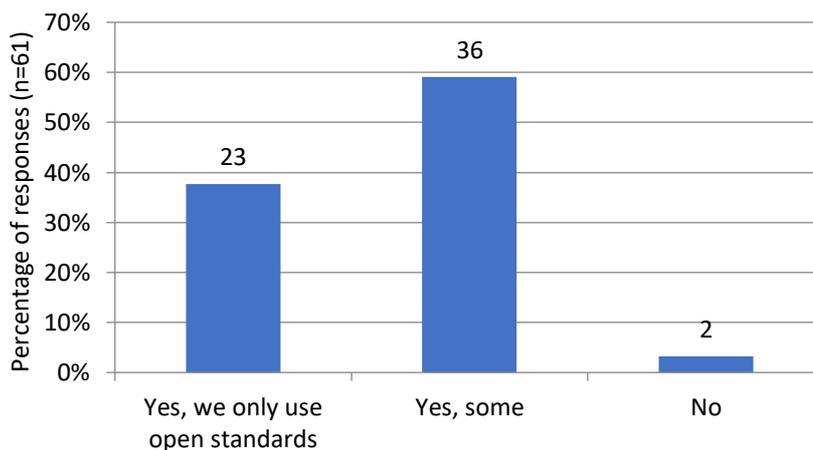
Figure 32. How many code contributors does your OSI have? (n=27)



5.5 Compliance with open standards, FAIR, EOSC and Plan S

36 out of 61 OSIs note compliance with at least *some* open standards, whilst a large minority (23) state that the OSI they represent *only* uses open standards, as seen in Figure 33. This clearly shows a commitment to openness by OSIs in this cohort although more open standards could be used in future to contribute to a more sustainable and interconnected infrastructure. Further, respondents were asked to list any relevant open standards that their OSI adheres to, with OAI-PMH and Dublin Core being the most commonly cited with each eliciting over 10 responses.

Figure 33. Does your OSI utilise and comply with open standards? (n=61)



The majority of responding OSIs (45 out of 54) report that they comply with the [FAIR data principles](#), with an additional 12 OSIs noting that this question was not applicable to them since this is tightly coupled to the type of service the OSI is. Forty-two free text responses were analysed to understand what FAIR principles OSIs comply with. The most common principles in our sample were (meta)data are assigned a globally unique and persistent identifier (23 out of 42), (Meta)data are released with a clear and accessible data usage licence (22), (meta)data are registered or indexed in a searchable

resource (20) and The protocol is open, free, and universally implementable (15). Among these, the first three are concerned with the use of metadata and identifiers, including licensing, while the last one is about the use of an open communication protocol. These figures indicate while FAIR is part of the OSI strategy for many, it does not prove that they are FAIR compliant.

Fifteen OSIs provide details with regards to their compliance with **EOSC service requirements**. The vast majority of requirements are reported by the OSIs as already being met or that plans are in place to address many of the currently unmet requirements. This shows that OSIs are up-to-date with current important standards and indicates their readiness for EOSC integration. Of these 15, five report following all EOSC requirements, while others are still planned. Despite being data-provider infrastructures, three OSIs, including a national data archive, report not planning certain requirements.

In terms of **Plan S mandatory technical requirements**, the survey made a distinction between journal publishing venues and OA repositories. The majority of OSIs either already comply with all requirements (5 of 14 publishing venues and in 8 of the 20 repositories) or have plans in place to address unmet requirements. This shows that OSIs are up to date with current important community funder publishing standards. Five repositories report that they are not planning to implement certain requirements as opposed to only two of the publishing venues. The most notable exclusion is the provision of machine-readable information on OA status and licensing conditions in the case of OA repositories, as four OSIs note no plans to enable this in the future.

With regards to data policies, only 19 out of 44 responding OSIs have a data management plan (DMP), compared to 25 who report that they do not, which is considered low since DMPs are increasingly becoming standard practise. Twenty-two respondents report that this question was not applicable despite thirteen of them reporting that their OSI supports Open Data, including FAIR data in a previous question, which may show that more awareness on DMPs is necessary.

5.6 Terms of use

38 out of 51 OSIs report having terms of use related to their data which could be higher when an OSI provides data. Terms of use are key to allow others to know under what conditions data can be used. It is unclear as to how transparent these terms of use are on the sites of OSIs. Numbers may be low since efforts may focus on providing technical services with less resources; this would be a clear oversight since interoperable OS services are dependent on understanding a system's terms of use. Limitations caused by third party content provider terms of use was noted as a challenge when providing open content by some OSIs. For more on this, see the subsection on Open Content in 4.4.3.

6. Governance and sustainability

6.1 In short

This section summarises the governance and sustainability of OSIs in Europe gleaned from Part 2b of the survey. It focuses specifically on OSI stakeholders, OSI governance structures, their resources, costs and income, funding, and sustainability. It should be noted that 68 of the total respondents to Part 1 completed Part 2b of the survey.

Overall, the main findings are that:

- OSIs largely have formal documentation such as missions/visions (94 out of 118), strategic plans or roadmaps (68 out of 118) and statutes (79 out of 120) which shows their professional capacity and a strategic approach to developing their OSIs;
- almost half of all OSIs have conducted a market analysis in the last four years, whereas the other half have not: showing a blind spot for comparing the OSI's value compared to others. However, respondents report OSI evaluation occurring in the last five years in 71% of all cases, which does show reflection and assessment exercises.
- most commonly, OSIs cite researchers (67 out of 68) and libraries (54) and research managers (43) as their most important stakeholders. A large proportion of OSIs (51 out of 64) have consulted their stakeholders on their needs within the last two years which shows their customer service orientation although there is room for more user-focus by some OSIs;
- the majority of OSIs (56 out of 68) have a board, steering group or advisory committee or equivalent and most commonly stakeholders such as researchers (30 out of 53) and libraries (27) are represented on these. This illustrates that a large majority exercise some good governance.
- approximately one-third of respondents (22 out of 64) begin each fiscal year without an approved budget which is a risk to an OSIs sustainability;
- with regard to the sustainability of OSIs and their costs: 21 out of 53 OSIs report spending less than €50,000, with two-thirds of the total cohort running on less than half a million euros showing that many OS infrastructures run at a relatively low cost. It also shows that many small infrastructures exist and the need to scale small and to connect these services to build a thriving ecosystem. The most commonly reported expenses were salaries/benefits (57 out of 66), travel and meetings (47), equipment (44) and marketing and communications (33);
- OSIs most commonly have between two and five FTE (25 out of 65) demonstrating their low costs. About half of the OSIs also report benefitting or relying on volunteers although the other half does not. This can be seen as both a strength and weakness to an OSI's sustainability;
- infrastructure income sees revenues matching the outgoings to a large extent. OSIs most commonly report national government grants (26 out of 65) as their main source of income with in-kind contributions, and membership fees coming second, and third. National grants are reported as the most important source of income with the second most important being the European Commission. Donations and private funding are far less prevalent with each source mentioned 8 times.
- OSIs most commonly identify potential funders through personal networks (25 out of 52);
- infrastructures report a wide range of funders (106) who have funded development on the one side and operations on the other (95);
- as to their reliance on grants, 17 out of 32 respondents state that without grant funding their OSI could only remain viable for less than a year, while 15 note that they would remain viable for more than a year. Thirty respondents report not relying on grants however;

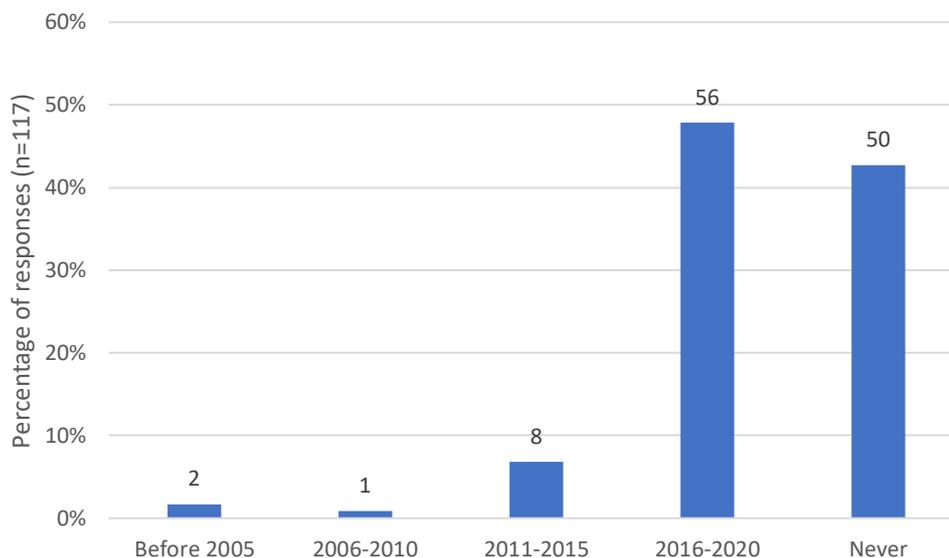
- infrastructures operate financially sustainably and at least break even on their costs, with 19 out of 36 noting that their operational deficit is covered by grants or sponsored projects, and 18 stating that they operate with a break-even thanks to earned income. Three OSIs generate a surplus, only one of which is for-profit, and
- finally, the greatest sustainability challenges relate to costs and funding, a lack of resources such as staff or equipment and the ability to keep up with technological development and OA/OS.

6.2 Mission, vision and strategy

94 out of 118 have a documented mission and/or vision and only 68 out of 118 respondents document strategic plans and roadmaps. 79 out of 120 OSIs report the use of statutes or similar documentation describing their governance structure. This shows their professional capacity and a strategic approach to developing OSIs.

To understand to what extent OSIs follow their place in the OS market, OSIs were asked about their market analysis activities, with most OSIs commonly having conducted a market analysis between 2016-2020 (56 out of 117) as seen in Figure 44 below. We do, however, note that almost as many respondents (50) report that they have never conducted a market analysis. This could be explained by the fact that the OSI may be more used to a reporting culture than an evaluation one; one involving external resources. Furthermore, OSIs may not consider themselves as part of the 'market' as such. More work would be necessary to verify these assumptions. To conclude this point, more OSIs would need to conduct a market analysis to gauge their added value as compared to others.

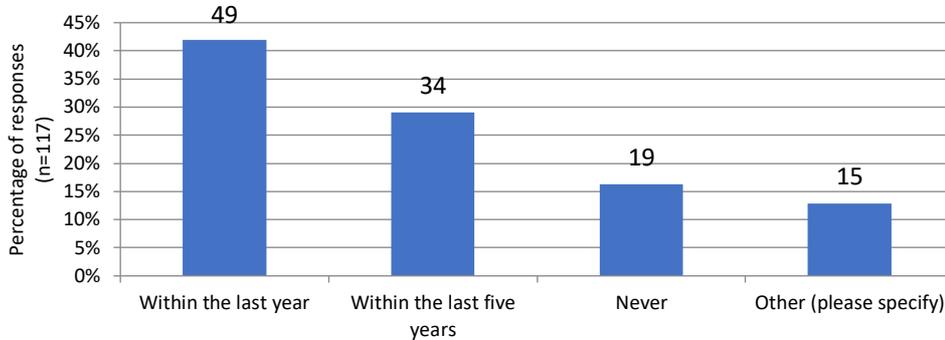
Figure 44. Last time a market analysis was conducted (n=117)



However, the majority of OSIs report that they had been evaluated either within the last year (49 out of 117) or within the last five years (34), totalling 71%, which shows their service-orientation and interest in ensuring that their OSI is up to date and future fit. Figure 45 shows, however, that 19 respondents report that their OSI infrastructure has never been evaluated. A small proportion of respondents (15) reported the last time their OSI was evaluated as 'Other', with three respondents noting evaluation had taken place within the last 10 years and three respondents commenting that

their last evaluation had taken place more regularly than the past year. It would be advisable that more OSIs would conduct more regular evaluations to gauge their relevance.

Figure 45. Last time OSI was evaluated (n=117)



6.3 Stakeholder communities

Respondents were asked to identify **who their most important stakeholders were**, meaning those who they work with or those who are impacted by their work. Nearly all respondents (67 out of 68) report researchers as being one of their stakeholders, followed closely by libraries (54) and research managers (43), who often use OSIs when evaluating research. Journalists (6) were the least common stakeholders, while volunteers, for both developmental (13) and non-developmental (11) work represent a fairly substantial group of stakeholders. Seventeen respondents selected ‘Other’, providing stakeholders such as research or medical institutions, repository vendors, journal editors, pre-print servers and the general public.

With regard to **the most important stakeholders** for OSIs (n=68), Figure 46 shows that few substantial differences exist between different types of OSIs. Those supported by research performing organisations most commonly support end users directly (researchers, libraries, research managers and students) while not-for-profit OSIs also commonly support other service providers, publishers and their own members. Interestingly, for-profit OSI respondents of this survey do not provide any service to students, advisory boards, journalists or volunteers outside of the development sector.

Figure 46. Most important stakeholders for OSIs (n=68)

(Legend: NFP=Not-for-profit, RPO=Research Performing Organisation, For-profit=For profit Organisation)

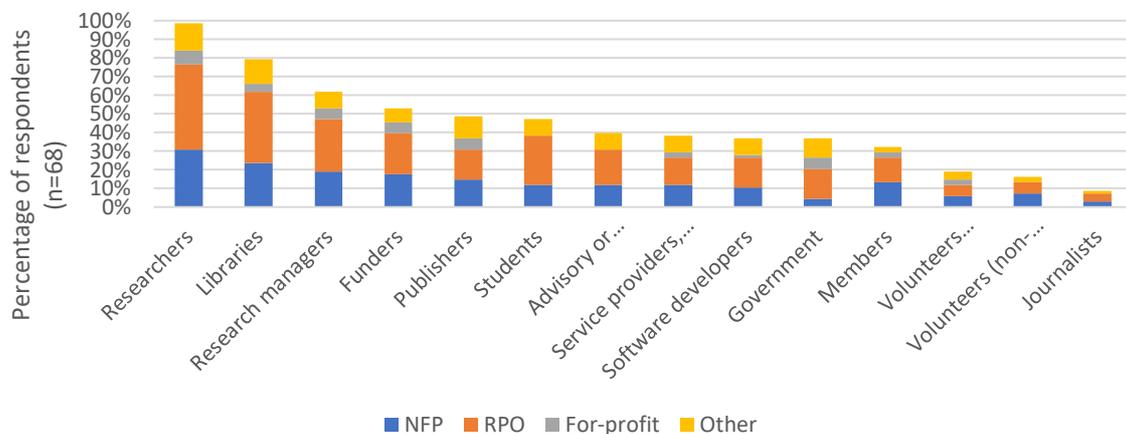
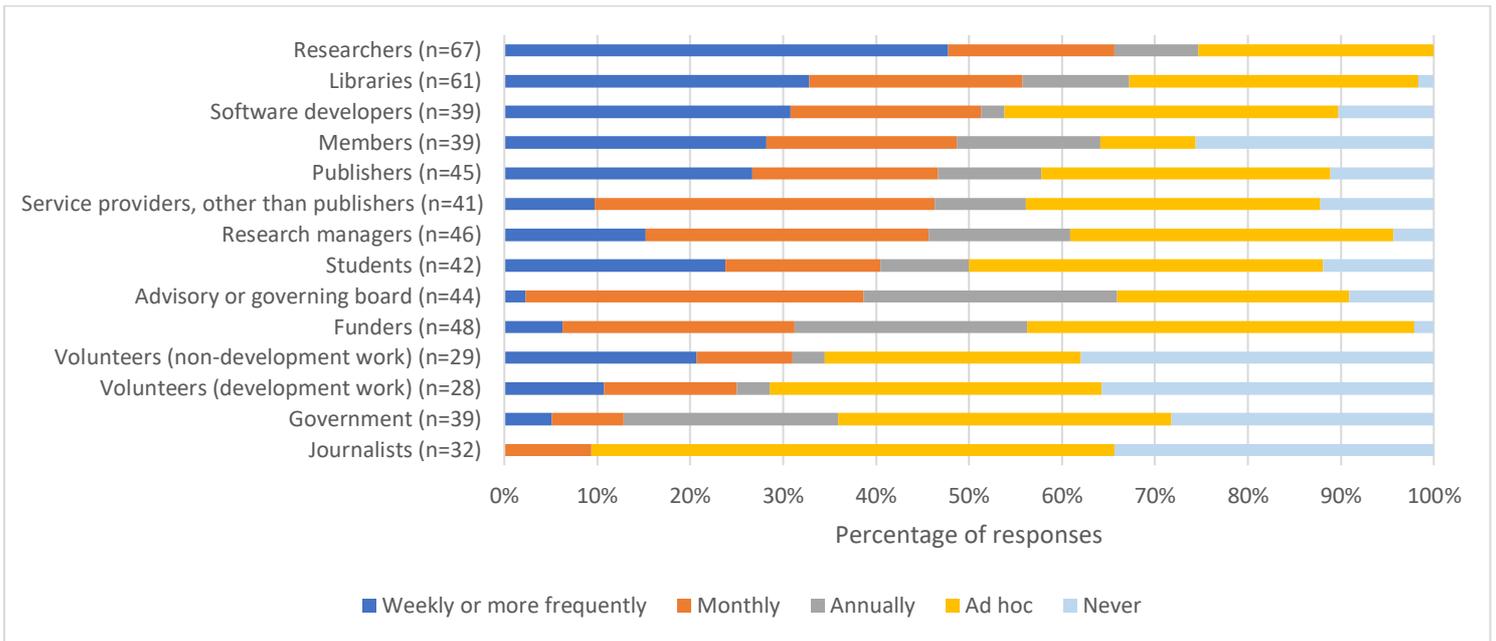


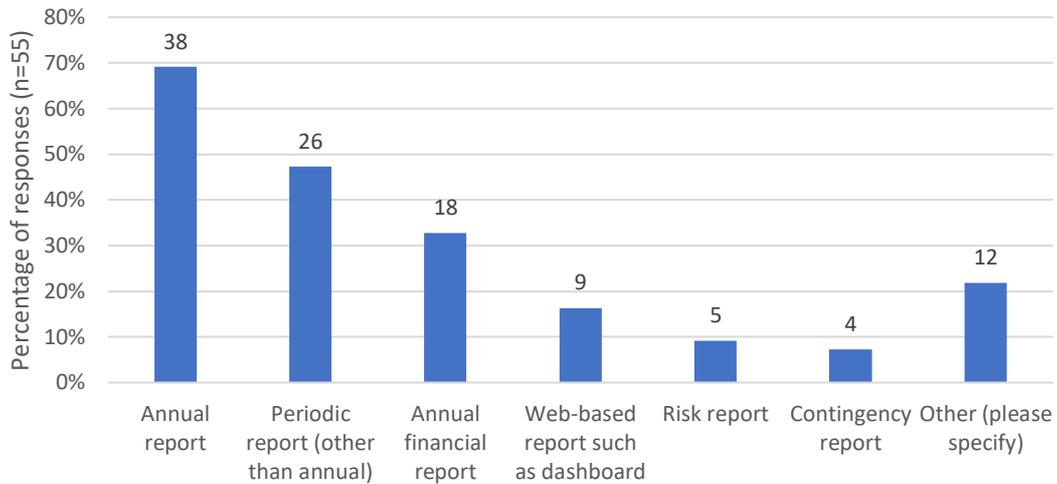
Figure 47 shows **how frequently OSIs engage with stakeholders** and the frequency of engagement for different stakeholders. For instance, there is more ad-hoc engagement with journalists, volunteers (developmental and non-development) and government who would presumably have less regular input although this might also show that OSIs may not engage in transferring knowledge to society. There is more regular monthly and even weekly engagement for those more likely to be regularly served by OSIs such as researchers, libraries, advisory or governing boards, and research managers. Six respondents list ‘Other’ stakeholders, citing ad hoc engagement with stakeholders such as research institutions and repository vendors, and more regular weekly engagement with journal editors and daily engagement with client organisations.

Figure 47 How frequently OSIs engage with stakeholders



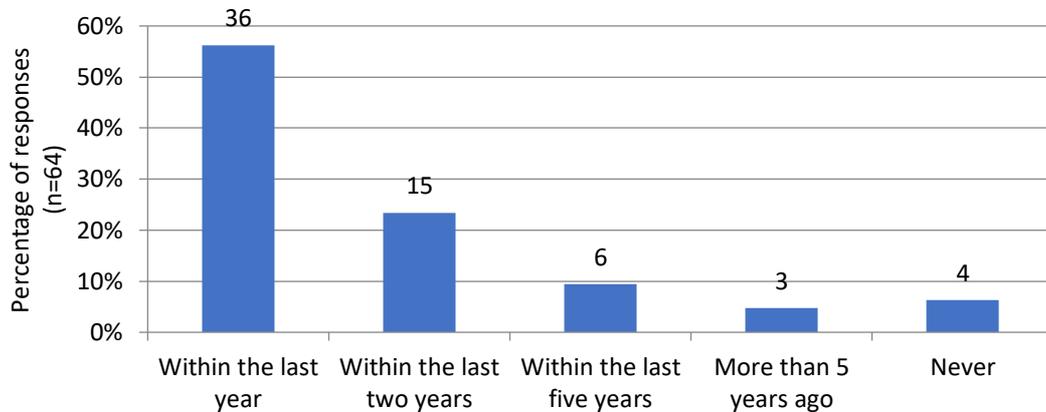
With regards to reporting, 55 out of 68 participating OSIs state that they regularly report to internal and external stakeholders on the organisation’s activities and finances, most frequently via annual (38 out of 55) or periodic reports (26) showing an active commitment to OSI stakeholders as seen in Figure 48. A minority of respondents provided ‘Other’ reporting formats such as updates at annual meetings, blog posts and presentations.

Figure 48. Community reporting mechanism (n=55)



A large proportion of OSIs (51 out of 64) note that they have consulted their stakeholders with regards to needs within the last two years showing a high level of customer service orientation. Only a handful (4) of responding OSIs say they have never engaged their stakeholder communities to assess their needs as shown in the following figure.

Figure 49. Last time stakeholder communities were surveyed about their needs (n=64)

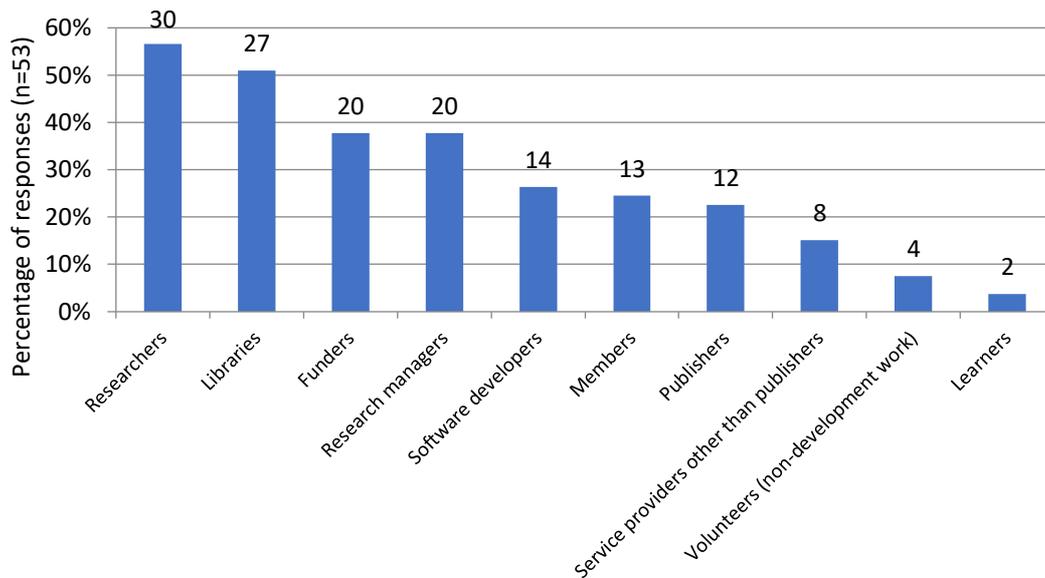


6.4 Governance

56 out of 68 participating OSIs note that they are governed by a board, steering group, advisory committee or similar showing concerted measures to govern their systems using formally agreed structures. Via free text comments, respondents report that decisions are usually taken by vote or through consensus, and only rarely do they come unilaterally from a director or president. In 11 out of 54 cases, participating OSIs mention the role of an advisory board in decision-making, for example by providing additional technical or scientific expertise.

Respondents were also asked which stakeholders are represented in their governance structures: researchers (30 out of 53) and libraries (27) tend to be the most represented categories, followed by research managers (20) and funders (20). The least represented stakeholder groups are learners (2) and volunteers (4), as seen in Figure 28. This closely coincides with those named as the most important stakeholders as seen in Figure 50 although publishers and service providers are less represented.

Figure 50. Stakeholder group represented on the board/governance groups (n=53)



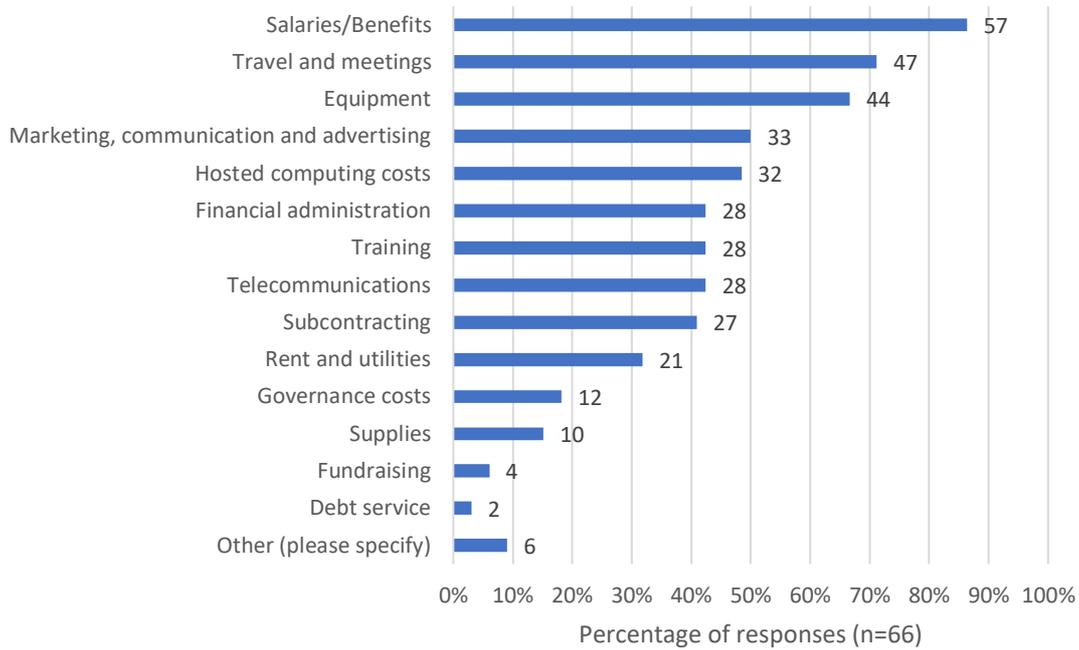
6.5 Sustainability

6.5.1 Costs and resources

Expenses

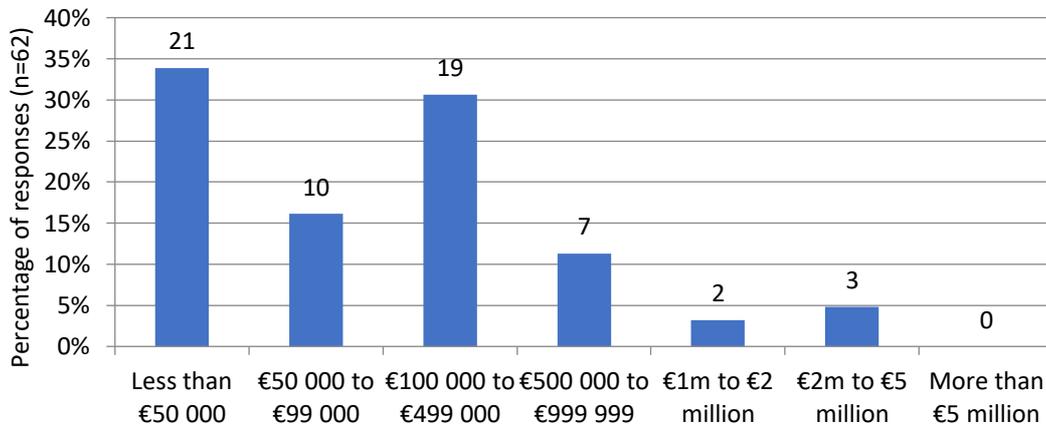
When asked about expenses, salaries/benefits (57 out of 66), travel and meetings (47), equipment (44) and marketing and communications (33) topped the list, as seen in Figure 51. Travel and meetings are a high expense and can be explained by many OSIs being maintained by in-kind contributions and this being an extra cost to the organisation. Furthermore, meetings are a way to stay connected to users and to gather user needs. A minority of respondents note 'Other' expenses such as university overheads and membership fees for systems their OSI relies on. When asked to identify the three most important expenses, the picture changed only slightly, with more importance being placed on equipment than on travel and meetings.

Figure 51. OSI expenses (n=66)



In terms of outgoings, Figure 52 shows that most OSIs report spending less than €50,000 in the last fiscal year (21 out of 62), but a similar share spent between €100,000 and €499,000 (19) while 12 OSIs report spending over €500,000. Seven of these serve the global OS community, and five are nationally focussed. Half of these OSIs serve all disciplines, with others focussing on the Arts and Humanities and/or Social Sciences, and two Applied and Engineering Sciences, including Computer Science and/or Mathematics and Physical Sciences. One of these services is a research data service and provides access to less than 10,000 objects. Generally, one can observe that many infrastructures have low expenses, also often running on in-kind contributions where many expenses are covered internally. This shows how important RPOs and not-for profits are in sustaining the OS sector.

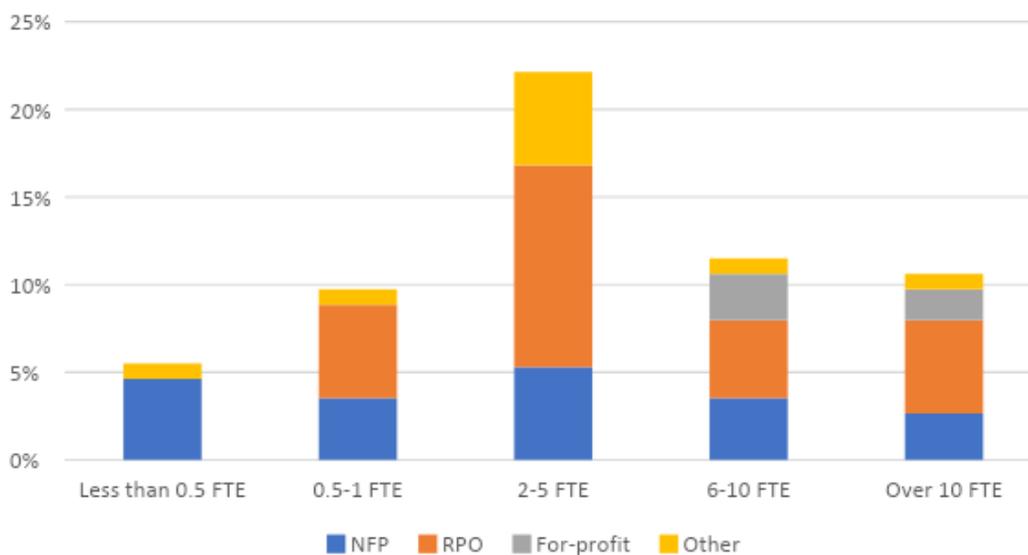
Figure 52. Expenses in last fiscal year (n=62)



Personnel resources

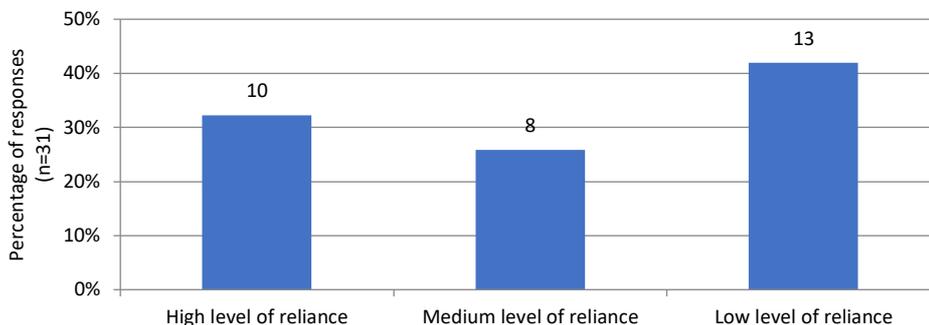
With regards to personnel resources and capacity, it is the not for-profits and RPOs who mainly work with limited personnel: 25 of 65 participating OSIs reported being supported by two to five FTEs mainly being reported by RPOs (19) and not-for-profits (14). Despite limited resources, of the not-for-profits, two OSIs provide extensive services across the publishing lifecycle, while another provides extensive services in hosting and access and discovery, a further offers three services in the creation phase, and another three OSIs offer a number of discovery services. A minority (4) have less than 0.5 FTEs (not-for-profits). Limited resources raises risks for the sustainability of an OSI since it can be dependent on a small core team. Ten OSIs have over 10 FTE: nine supporting research data, and seven of those also OA. Half of these offer a very broad or broad range of services as shown in the following figure.

Figure 53. The number of FTE currently supporting OSI by organisation type (n=65)



Most OSIs are split on to what extent they rely on voluntary efforts with 35 out of 66 not dependent on volunteers; and 31 reporting they do as shown in Figure 54. They were asked about their level of reliance on volunteers (high/medium/low), but responses to this question were relatively uniform across the scale. This is explained by the wide variety of tasks undertaken by volunteers, which often include code writing, providing feedback on activities, and data entry and management. Engagement with volunteers clearly shows staff commitment to the OSI. However, reliance on volunteers – if not in-kind contributions – could be considered a weakness or threat to the OSI since having no budgeted personnel resources might mean that volunteers can be more easily withdrawn when priorities change.

Figure 54. Reliance on volunteer effort (n=31)

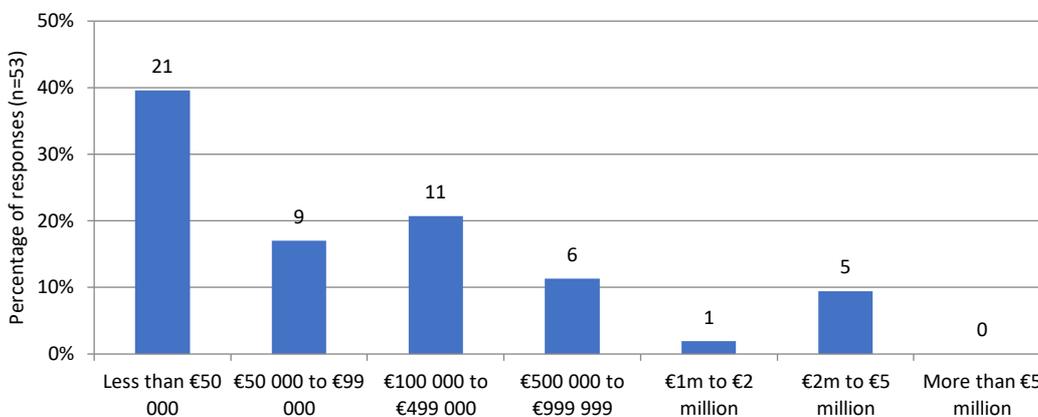


6.4.2 Income

Participating OSIs were asked a series of questions around the costs they incur and any income they generate to cover these costs. 42 out of 64 respondents state that the OSI they represent has an approved budget at the beginning of each fiscal year. However, one-third start the year without an approved budget bringing a possible risk to the stability of such OSIs.

About 21 out of 53 report an annual revenue last year of less than €50,000 whereas 12 OSIs report a revenue above €500,000. Most of the OSIs with over a half million Euros in revenues last year offer a broad or very broad range of activities. Eleven of these provide hosting and access, discovery, archiving/preservation services.

Figure 55. Revenue in last fiscal year (n=53)



Who funds OSIs?

Over the last two years, OSIs report covering their costs via national government grants (26 out of 65) or in-kind contributions (20), followed by membership fees (18), host subsidy (16) and European Commission grants (14) as seen in Figure 56. Membership and service fees are rather high since they help sustain the OSI.

Twenty-five respondents report ‘Other’ income with around half of these mentioning specific grants or support from institutions, partners or via research organisation alliances. In addition, ‘in-kind’ was not always understood by participants since clear examples of such contributions were mentioned under Other. Four other OSIs reported individual donations or voluntary contributions as income streams, whereas two others report income through training or book revenues. Two more mentioned SCOSS specifically and sponsorships in one case.

Figure 56. How the OSI has covered the costs of funding over the last 2 years? (n=65)

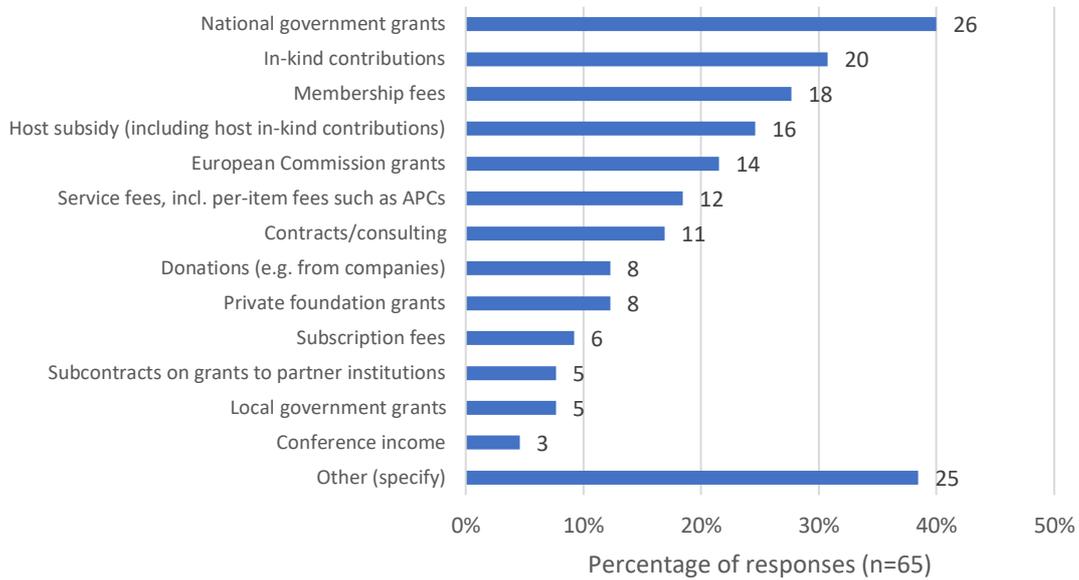
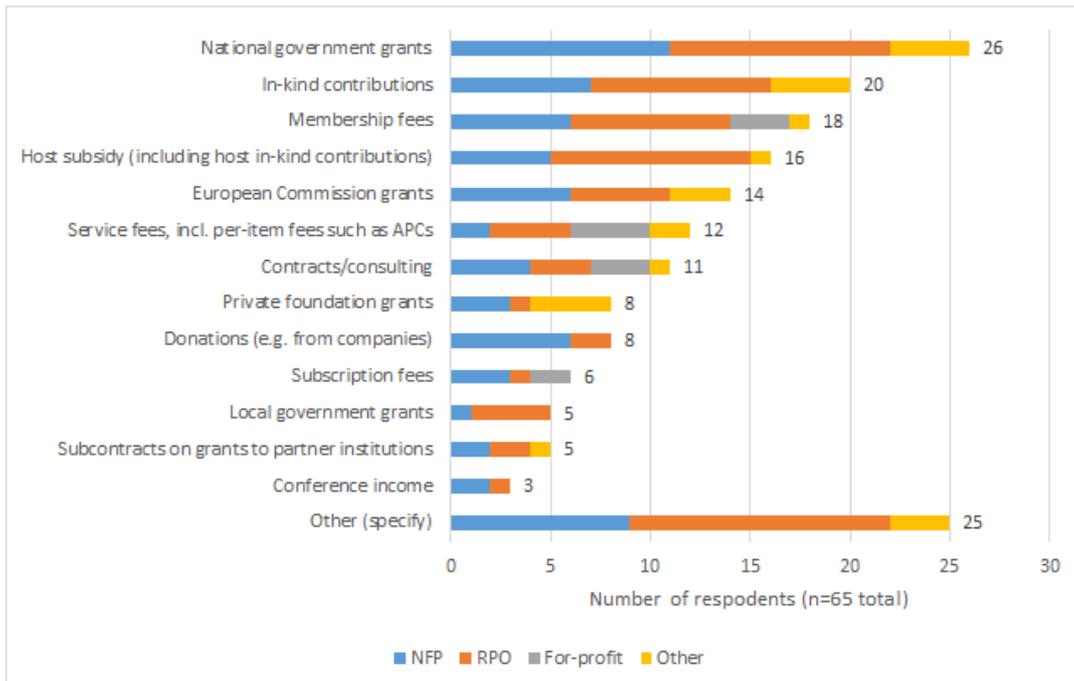


Figure 57 shows that RPOs and not-for-profits use a wide range of revenues to cover costs. Certain kinds of income are only seen in RPOs or not-for-profits such as national or local government grants, in-kind contributions, host subsidies, EC grants, private grants or donations, subcontracts on grants or conference income. For-profits cover their costs through service fees, membership fees, contracts or consulting and through subscriptions.

Figure 57. How OSIs have covered the costs of funding over the last 2 years by type of organisation (n=65)



To gain better insights into how OSI **covered their costs over the previous 2 years**, we analysed the types of revenues reported. Of the 29 who are RPOs, nine reported only one source of income of which three were government funded and four were in-kind contributions. The majority of RPOs (13) report having three or more sources of income. Of the 21 not-for-profits, only three report depending on one source of income. The majority (11) report 3 or more sources of income. One respondent reports a wide range of seven sources of income. This shows that the majority are dependent on multiple sources of income to sustain themselves. For-profits SMEs with less than 250 employees never reported more than three income streams, including membership fees, service fees and contracts. This illustrates that not-for-profits or RPOs are more dependent on a multitude of income streams although many are also dependent on one income stream, which for-profits are not.

As shown in Figure 58 below, 22 respondents report national grants as the **most important source of income**, which is also reported by most as a source in the last 2 years with the European Commission coming in second place with only 14. Almost all of those who report European Commission funding as a funding source in the last two years (12/14), report it as one of their most important sources of income.

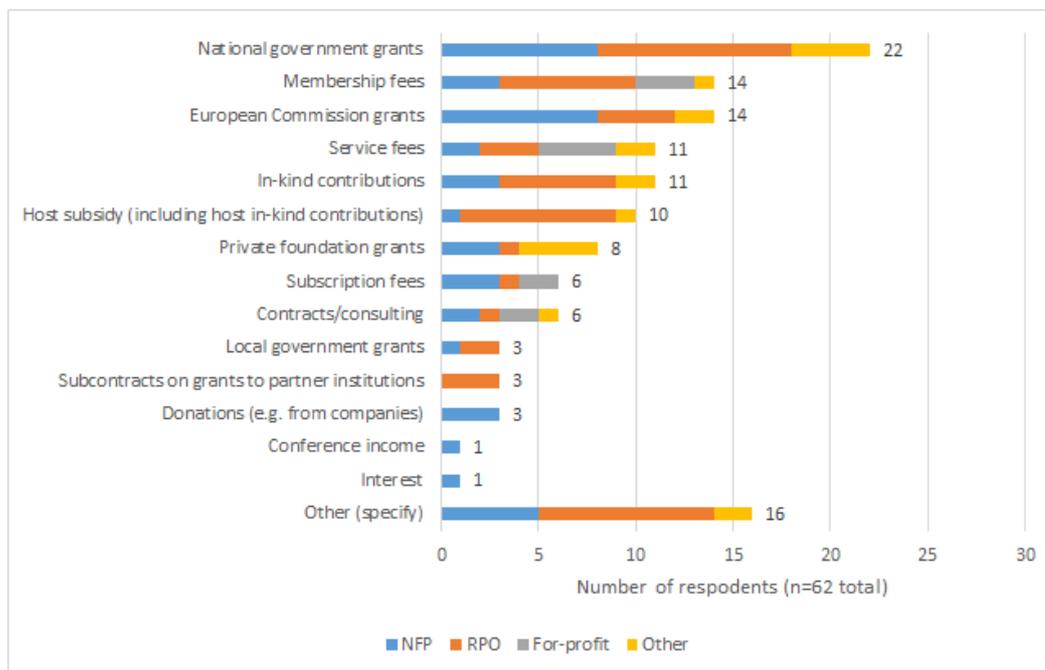
Conversely, private funding is far from being at the centre of investment in OSIs in Europe. Only three respondents report donations from companies as an important source of income although private foundation grants feature higher with eight. When private funding is involved, it is OSIs that relate to STEM disciplines or used by both STEM and SSH (as in the case of infrastructures dedicated to the discovery phase of the research workflow) who tend to benefit from this type of funding. This clearly shows the insignificant role that private funding plays in OSIs at present, something that could be further explored.

While in-kind contributions or a host subsidy are reported by many as a source of income in the last two years (20 and 16 respectively), it is curious to find they are mentioned by significantly fewer OSIs as important streams of income (11 and 10 respectively) although they are in the top six income streams. This may be due to their dependence on external funds to make ends meet.

Membership fees are a more popular stream of revenue for OSIs ranking third both as a popular income stream and in being considered as an important one. Conference income, local government grants, subcontracts on grants, subscription fees, contracts and consulting remain of lower importance in this cohort both as a source of income and of importance. This illustrates that many OSIs rely on external contributions to fund their activities.

Figure 58 shows that national government grants are most important for both not-for-profits and RPOs. Interestingly, membership fees are more important to RPOs than not-for-profits. The most important type of income for for-profits is service fees.

Figure 58. The three most important sources of income that have supported OSIs over the last two years (n=62)



Funding development work vs. operations

To understand the range of funders upon whom OSIs depend, OSIs were asked who had granted them funding for *development work* over the last five years, and who had granted them *support for operations*. Names were provided, and these were then grouped for analysis.

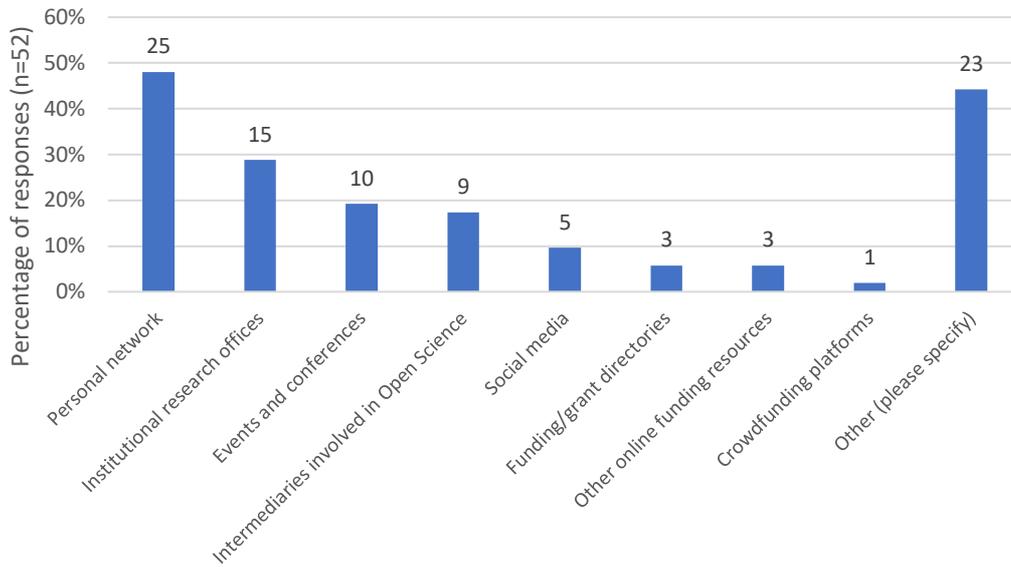
Respondents shared the names of 106 identifiable funders who funded *development*, of which 99 were unique. The funder mentioned most frequently was the European Commission (8) followed by three French research organisations, which relates to the larger proportion of French OSI respondents. For *operational funding*, 95 identifiable funders were provided, of which 82 were unique, with CNRS mentioned most frequently (8), and the European Commission coming in second (5). This is curious since the Commission generally funds innovation rather than maintenance costs, perhaps showing that OSIs use external funding to make ends meet out of necessity.

The types of funders identified were similar for both funding development and operations. OSIs report public institutions as the main financial backers of OSIs, whether to cover development costs or operations. Public institutions include a variety of organisations operating at international, national and regional levels. These include ministries of education and research, research agencies, universities, libraries, administrative regions and the European Commission. Investment is made through specific grants, through a permanent budget or via salaries.

How OSIs identify funding opportunities

Responding OSIs were asked how they identified their sources of funding to better understand how they locate funds to sustain themselves. The most common response was through personal networks (25 out of 52), followed by institutional research offices (15) and via events and conferences (10) as shown in Figure 59. This shows that funding acquisition does not seem to be a routine activity for OSIs in a less than mature sector, which is a risk to the OSI's sustainability. Eleven respondents report other ways to identify funding opportunities, including institutional contacts more generally, government or EU contacts, the OPERAS network or by getting informed through newsletters or web announcements. One indicated that they identify funders through author affiliations or acknowledgements. A few OSIs report that funders approach them directly.

Figure 59. How funders were identified (n=52)

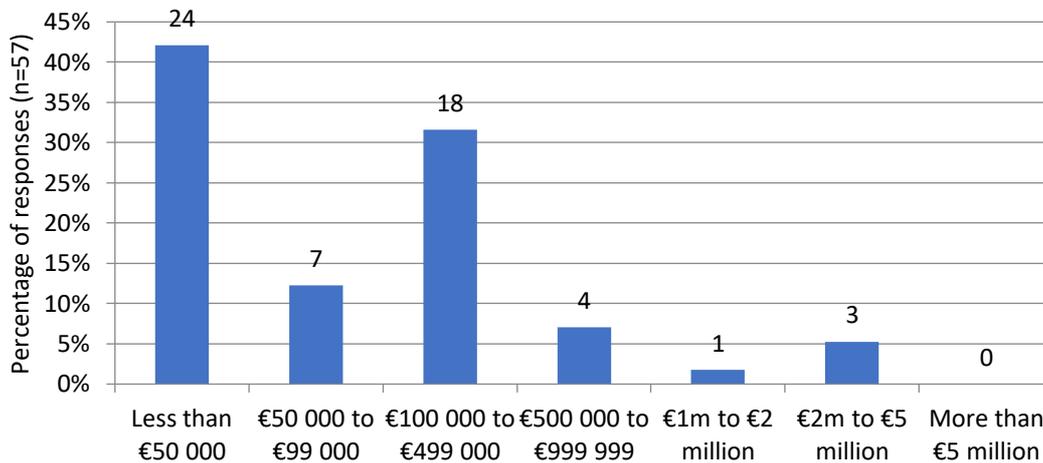


6.4.3 OSI financial health

Investments or grant income

In their last fiscal year, participating OSIs most frequently were invested or granted sums lower than €50,000 (24 out of 57) or between €100,000 and €499,000 (18), shown in Figure 60. Just over half of OSIs state that they were not planning for substantial changes in revenue streams in the future (34 out of 65).

Figure 60. How much was invested/granted in last fiscal year (n=57)



OSIs were also asked if they anticipated any financial impact due to COVID-19. While 33 respondents indicated that they did not – meaning that they are reasonably stable infrastructures – 11 others did fear negative consequences from Covid-19. Five respondents felt that there would be a small impact if any on their OSI.

When asked how long an OSI might remain viable without grant funding, 30 of the 68 respondents report that their infrastructure is not dependent on subsidies coming from grants, indicating that these OSIs continue to work without grant funding, the ideal case for an infrastructure. Most for-

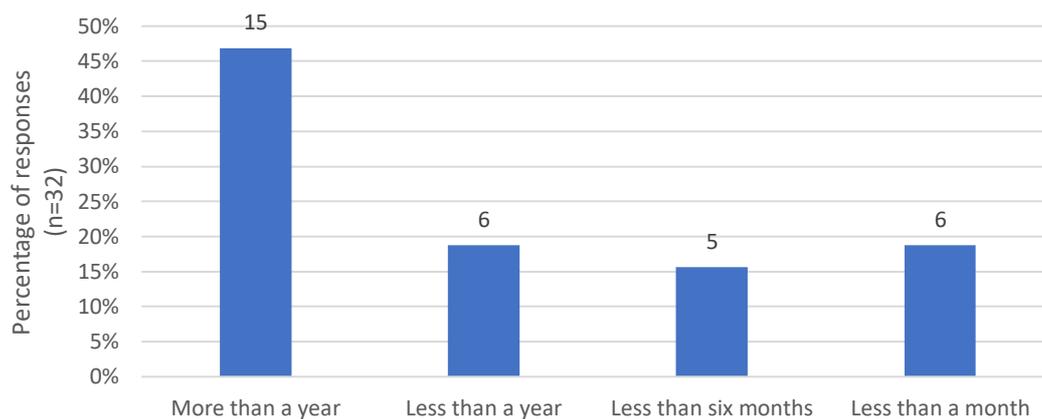
profits, 18 RPOs, five not-for profits and four other organisations report this fact. However, six of these respondents say that without grant funding in the future, they would have difficulty developing their OS.

A further six OSIs indicate that if grants were to stop, that this might make them obsolete or would impact them gravely. Twenty-three other respondents report that this would have an impact on their development. Some also report them losing their international network and that it would be slow to conform to new standards, making them ill-equipped to face competition. For those who rely on grants, very few respondents (2 of the 68) report that no grant funding would have any negative consequences. Only three of 68 are not concerned by grant funding since their main funding comes from their affiliation to public institutions, the sustainable business model for OS infrastructure.

Many also report that if grants were to come to a halt, that this would not affect operations since they have other funding models to maintain their work.

As seen in Figure 61 below, of the 32 respondents who depend on grants, 15 out of 32 respondents said they could last for more than a year without grants. However, six report they could remain viable for less than one year, with five more reporting less than six months, and a further six surviving less than one month without grants. This clearly shows the unstable footing of some OS infrastructure upon which some of us depend.

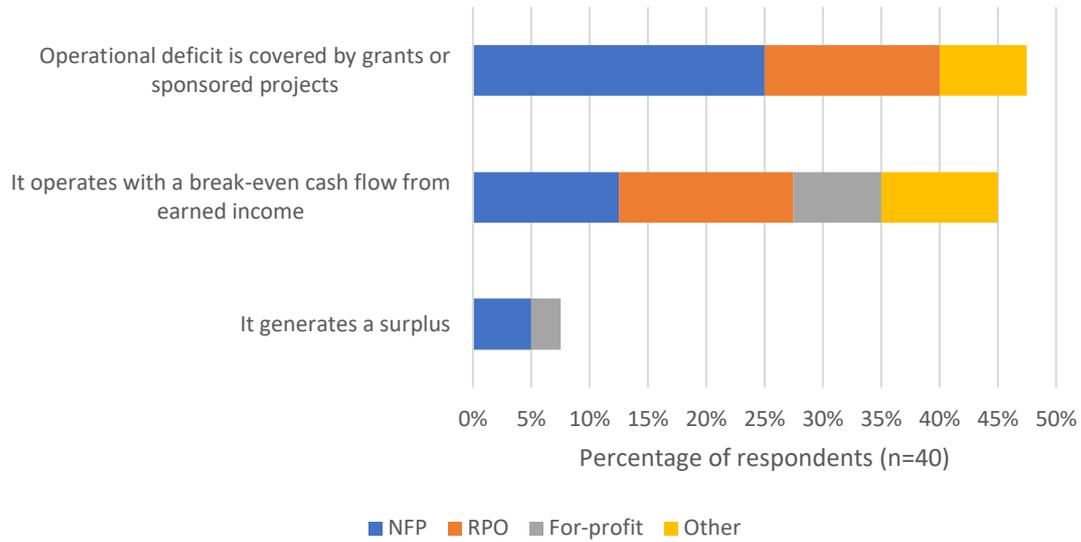
Figure 61. Without grant funding, how long the OSI reported it could remain viable (n=32)



OSIs were also asked to describe their current situation in terms of financial sustainability and 19 out of 40 noted that their operational deficit is covered by grants or sponsored projects, while 18 stated that they operate with a break-even thanks to earned income, which is illustrated in Figure 62. This shows that the majority of responding OSIs operate sustainably and manage to generate revenues at least covering their costs although many do not have a surplus to deal with eventualities: if even permitted to due to their legal structure.

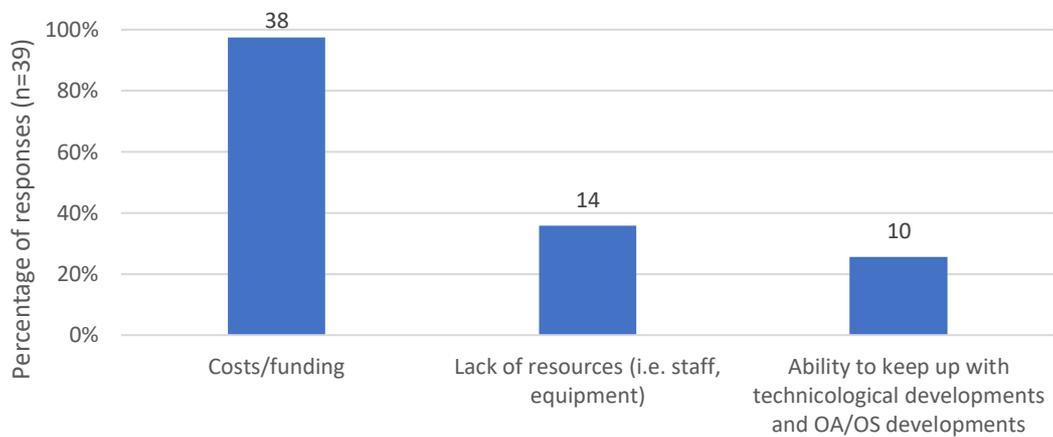
Notably, there is a variance in this data by organisational type. Although none of the RPOs in this sample report a surplus, an equal number of RPOs covers deficits with grant funding as they break even with earned income. Conversely, one not-for-profit generates a surplus through commercial activities, but the vast majority operate at a loss and rely on grant funding or sponsored projects to break even. Predictably, for-profits either break even or generate a surplus and do not rely on grant funding.

Figure 62. OSI's current sustainability health



When asked for their three greatest challenges related to sustainability, OSIs most commonly cite costs/funding (38 out of 39), followed by a lack of resources (14) (including both staff and equipment) and their ability to keep up to date with technological developments and developments pertaining to OA and OS (10), which is low considering that they generally do keep up with the state of the art in their field.

Figure 63. Three greatest challenges related to sustainability (n=39)



7. Conclusions

This study provides us with various key insights into a range of Europe's Open Science Infrastructures (OSIs), including what they are, how they function, and where they need further support.

The OSI offering

European Open Science Infrastructures (OSIs) essentially have the ambition to make research and knowledge openly and widely available primarily by targeting researchers, libraries and research managers. RPOs and not-for-profits contribute significantly in this sector since the overwhelming majority of OSIs are run by them. Most OSIs offer services in three or more stages of the research lifecycle: Creation, Evaluation and Commenting, Publishing, Hosting, Discovery and Archiving/Preservation, with Discovery and Archiving/Preservation as the most commonly supported types of service. OSIs provide access to a wide range of research outputs: most commonly, to journals and research data with many OSIs also offering a variety of non-traditional outputs such as images, software or code, and early research outputs in preprints, posters and proposals.

Whilst using the COAR/SPARC *Principles for Scholarly Communication Service* lens, we observe how OSIs run their infrastructures and to what extent they are open. Most OSIs report a level of maturity in providing open content and in abiding by open standards; the majority of their good practices are also reported in these areas. Nevertheless, for many OSIs, **challenges remain in good governance, sharing open content and applying open standards. Sharing good practices, developing communities of practice and developing guidance would help to further support OSIs and their missions.**

The majority of OSIs surveyed demonstrate up-to-dateness and forward-looking focus since they have generally evaluated their technical environment in the last five years at least once. Most infrastructures have a dedicated technical lead at the helm. OSIs also frequently integrate with external systems and services, which helps to build and strengthen an interconnected OS infrastructure landscape, which is greater than the sum of its parts. The most commonly cited systems – and thus essential infrastructure for many – are ORCID, Crossref, DOAJ, BASE, OpenAIRE, Altmetric, and Datacite, most of which are not-for-profit. Furthermore, aggregation and indexing, search, storage, identity (e.g. ORCID) and persistent identifiers (PIDs) seem to be vital components of OS. PIDs in particular since half of all respondents mention integrating with them. **Prioritising investment in such critical nodes of infrastructure would have a positive impact on the OS and research community.**

As far as open content and open standards are concerned, over half of all OSIs comply with *some* open standards while a lower but important proportion *only* use open standards, which shows a strong adoption of open standards by OS infrastructures. Furthermore, a good majority of OSIs encourage the re-use of their data by having an API. Open source is also a standard that is widespread amongst OSIs with over half of all OSIs being fully open source or partly open source to a lesser extent. Good collaboration is essential to the success of an OSI: with over half reporting multiple code contributors, though a good proportion have fewer than five, which could indicate some fragility in sustaining some OSIs. Many OSIs show up-to-dateness, reporting that they largely comply with specific OS community standards set either by the research or research funder communities, such as FAIR, EOSC or Plan S. **Despite a strong uptake of open source and open standards by many, other OSIs could benefit from community support to get them up to speed in this area.**

Governance and community

OSIs are largely professional infrastructures, steered by missions/visions, with many following strategic plans or roadmaps. Two-thirds are governed by statutes. However, although just over 70 per cent of respondents report OSI evaluation occurring in the last five years, over half of all OSIs have not conducted a market analysis in the last four years. OSIs could **prioritise spending resources on objectivising the value of their work by evaluating and reflecting more on the value they contribute to the OS and research community as compared to others.**

The large majority of OSIs demonstrate some good governance with a board, steering group or advisory committee or equivalent in place with important stakeholders such as researchers and libraries represented on these. Other OSIs, however, still grapple with ensuring sound, equitable and community-relevant decision-making or acquiring adequate and sufficiently diverse representation. For OSIs to suitably support their communities and to remain independent, it would be helpful to **learn about the good practices of good governance** from peers.

OSIs demonstrate sound customer service-orientation since a large proportion of OSIs have consulted their stakeholders on their needs within the last two years although **additional user-focus by some OSIs would be beneficial**.

Financial sustainability

OSI expenses (excluding in-kind contributions) are generally low, with two-thirds spending less than half a million Euros last year. Almost half spend less than €50,000, showing that many OS infrastructures run at relatively low cost. OSIs also operate with limited personnel, most commonly between 2-5 FTE and approximately half of the OSIs report benefitting from or relying on volunteers. This can be considered as both a strength and weakness to an OSI's sustainability. All in all, our data show that numerous small infrastructures exist and co-exist. **The OS and funding communities can help scale small to build a thriving ecosystem: connecting OSIs to share resources and upskill.**

When comparing outgoings with revenues, revenues seem to balance with outgoings to a large extent. For those who answered questions on sustainability and finances, OS infrastructures report operating in a financially sustainable way and at least break even on their costs. Nearly half of the respondents cover their operational deficit through grants or sponsored projects, and the others operate with a break-even, thanks to earned income. Three OSIs generate a surplus, only one of whom is for-profit. Nevertheless, despite such sustainability health checks, OSIs still report the greatest sustainability challenges with a lack of resources (such as staff or equipment), the ability to keep up with technological development and OA/OS but above all related to costs and funding.

Many challenges lie in the area of financial sustainability for OSIs. First of all, **OSIs can improve on managing their budgets** since about one-third of respondents begin each fiscal year *without* an approved budget, which certainly imposes risks to an OSI's sustainability. To sustain themselves, infrastructures manage a range of revenues and business models. National government grants are central to the funding of many with OSIs reporting them as their main source of income and most important one in many cases. The European Commission is also a priority source for numerous OSIs for both funding development and operations. In-kind contributions and membership fees are also key sources of income for some while OSIs benefit from donations and private funding to a far lesser extent. Whereas many OSIs report not relying on grants, some OSIs are partly dependent on grant funding for their survival with 17 reporting that without grant funding their OSI could only remain viable for less than a year, which is a clear concern if the community depends upon them.

OSIs report over 100 funders that have funded development on the one side and almost as many who have funded operations on the other, indicating that there is a range of funding opportunities if OSIs know how to locate them. When exploring who might fund their OSIs, most use personal networks, showing funding acquisition does not seem to be a routine activity for OSIs in a less than mature sector. **OSIs will need to continue to diversify their fund-raising efforts and upskill to embrace a range of business or revenue models in the future to spread the risk to their financial stability.**

To sum up

As we move towards the research community being increasingly open in nature, with measures such as Plan S soon coming into full effect and with the growing demand for openness resulting from COVID-19, we anticipate that the importance of OSIs will magnify as they become the backbone of new open workflows globally for stakeholders in the scholarly communications landscape. Although we have a rich and diverse OS infrastructure offering, this sector could become even more of an open ecosystem, one that is connected through open content and open standards, and one that draws on the benefits of being open, and is enabled to do so. **This ecosystem will thrive when following good governance practices to ensure the community has the confidence that it will be steered by the needs of the community, and will stay true to the values of research.**

It is, however, important to note that for some OSIs, their sustainability is intrinsically linked to their ability to receive government or grant funding. Further, COVID-19 places risks on the continued availability of such funding in the future. To further the efforts of OSIs in this regard, **funding agencies, governments, institutions, charities and other funders must strategise on how to effectively fund this rich and important landscape more structurally. We also call on governments to maintain and increase support** – similar to the report on Supporting the Transformative Impact of ESFRIs – **for both development activities and for sustaining operations.**¹⁸ It will be important for OSIs to **continue to have a range of business models and revenues at their disposal to reduce the risk to sustainability.**

We see a diverse, interconnected, open, professional and viable OSI ecosystem developing in Europe on solid ground – one that is worth investing in. It is a system that is made up of valuable service providers, both large and small, serving the global research community. Nonetheless **OSIs still have a range of issues to contend with in their organisations and strategies, particularly as they move towards more openness and sustainability. Sharing lessons learnt and pooling resources through initiatives such as Invest in Open Infrastructure (IOI) and JROST will help them grow even further.**^{19,20}

This study is the first step towards gaining a better understanding of the OS ecosystem and how it is sustained. Further research into the challenges of open and the OSI's financial health through qualitative research would help to get to the core of the exact issues at stake and would shed more light on where specific assistance could be provided by the OS and funder communities in the future. In addition, gaining more insights into the lifecycle of each OSI from start to finish, and the challenges along that path as seen with the *2020 IOI interviews* would be valuable before designing ways to strengthen OSIs to make them more open and futureproof.²¹ Understanding how OSIs connect to other infrastructures such as Europe's ESFRIs, and what the opportunities are for alignment, collaboration and mutual learning are essential to build the OS infrastructure ecosystem of the future. We look forward to continuing on that journey.

¹⁸ Supporting the Transformative Impact of Research Infrastructures on European Research, European Commission, 2020 https://ec.europa.eu/info/sites/info/files/research_and_innovation/strategy_on_research_and_innovation/documents/ec_rtd_transformative-impact-ris-on-euro-research.pdf

¹⁹ Invest in Open Infrastructure: investinopen.org

²⁰ JROST: <https://investinopen.org/community/jrost-2020-conference/>

²¹ IOI Interviews: <https://sparceurope.org/ioiinterviews/>

Appendix A: Respondents by country

Below is a list of respondents and their country of origin. All respondents completed part 1 of the survey, covering the OSI landscape in Europe; of these, over half completed part 2a (focusing on the technologies underpinning OSIs) and part 2b (focusing on OSI governance and financial sustainability).

Respondents of Part 1:

Scoping the OA & OS Infrastructure (OSI) Landscape in Europe

OSI name	Country
4TU.ResearchData	Netherlands
Advances in Combinatorics	UK
BEEP (Bibliothèque électronique en partenariat)	France
Beilstein-Institut	Germany
Bielefeld Academic Search Engine (BASE)	Germany
Bridge of Knowledge (Most Wiedzy)	Poland
BulDML	Bulgaria
CCDC	UK
Centre Mersenne	France
CINECA	Italy
CORE (core.ac.uk)	UK
Croatian Scientific Bibliography CROSB	Croatia
Crossref	UK
Crystallography Open Database (COD)	Lithuania
Data and Service Center for the Humanities (DaSCH)	Switzerland
DataCite	Germany
DataverseNO	Norway
dblp computer science bibliography (DBLP)	Germany
Digital Academic Archives and Repositories – DABAR	Croatia
Digital Repository of Ireland (DRI)	Ireland
DIGITAL.CSIC	Spain
Directory of Open Access Books, DOAB	Netherlands
Directory of Open Access Journals (DOAJ)	UK
Dissemin	France
DiVA – Digitala Vetenskapliga Arkivet/ Academic Archive Online	Sweden
DMPonline	UK
e-cienciaDatos; Research Data Repository from the Consortium Madroño (Consortium of Public University Libraries from Madrid Region)	Spain
EconStor	Germany
EGI	Netherlands
Episciences	France
eTalk	Switzerland
Fair Open Access Alliance (FOAA)	Netherlands
Figshare	UK
FORS	Switzerland
Frictionless Data (FD)	UK
GBIF: Global Biodiversity Information Facility	Denmark

OSI name	Country
GRICAD	France
HAL	France
Horizon Pleins Textes	France
Huma-Num	France
Hypergraph	Germany
Information and reference system "Sort" (IRS "Sort")	Ukraine
Integrated Carbon Observation System, ICOS	Sweden
IOSIN	Romania
Istituto per il Lessico Intellettuale Europeo e Storia delle Idee, Consiglio Nazionale delle Ricerche; ILIESI-CNR	Italy
Jisc UK ORCID Consortium	UK
Juelich Shared Electronic Resources (JuSER)	Germany
KIFÜ	Hungary
Kriterium	Sweden
Lithuanian Academic Electronic Library (eLABa)	Lithuania
Lithuanian Data Archive for Social Sciences and Humanities (LiDA)	Lithuania
LUDITA	Hungary
Materials Cloud	Switzerland
MDC (Memòria Digital de Catalunya / Digital Memory of Catalonia)	Spain
Mir@bel : a free gateway to journals contents	France
Mundi Web Services	Germany
NARCIS – National Academic Research and Collaborations Information System	Netherlands
National Aggregator of Open Access Repositories, NEICON & NORA	Russia
National Open Access Research Data Archive (MIDAS)	Lithuania
NewsEye	France
Nord Open Research Data	Norway
numerev	France
OAPEN	Netherlands
OAR@UM	Malta
Observatory of International Research (OOIR)	Austria
Open Access Monitor (OAM)	Germany
Open Access Publications	Switzerland
Open Commons of Phenomenology	Switzerland
Open Knowledge Maps (OKMaps)	Austria
OpenAIRE	Greece
OpenCitations	Italy
OpenDOAR	UK
OpenEdition	France
openRDM.swiss	Switzerland
openSNP	Germany
Pangloss Collection	France
Peer Community In (PCI)	France
Pépinière de Revues en Open Access (PREO)	France

OSI name	Country
Pépinière de revues et 560le d'appui	France
Persée	France
Phaidra	Austria
Plaudit.pub	Netherlands
PoPuPS	Belgium
Portal of Croatian Scientific and Professional Journals – HRČAK	Croatia
Prairial	France
Publications Router	UK
Publisso	Germany
QOAM	Netherlands
Quantum	Austria
RACO (Revistes Catalanes amb Accés Obert / Catalan Journals in Open Access)	Spain
RECOLECTA (Spanish National OA harvester)	Spain
RECYT	Spain
Renku	Switzerland
Repository services at the National Library of Finland	Finland
RERO DOC	Switzerland
research data management tool RE:COLL	Germany
Researchfish by Interfolio	UK
Reseau de Pepinières de Revues Scientifiques (REPÈRES)	France
Riviste unimi	Italy
RUIdeRA	Spain
RUMAK	Poland
Scholia	Denmark
ScienceOpen	Germany
Sciencesconf	France
Scientific Conferences of Ukraine	Ukraine
Scientific Periodicals of Ukraine	Ukraine
SciPost	Netherlands
Semmelweis University Central Library Open Access & Open Science	Hungary
Serbian Citation Index (SCIndeks)	Serbia
Sherpa Services	UK
Software Heritage (SWH)	France
Software Sustainability Institute (SSI): Training, Guidance and Evaluation Services	UK
Tesis Doctorals en Xarxa (TDX) – Theses and Dissertations Online (TDX)	Spain
The National Digital Library of Latvia, LNDB	Latvia
Think. Check. Submit.	UK
TRAP-RCUB Repository Network	Serbia
Tudasportal	Hungary
Ubiquity Press (UP), Ubiquity Repositories (UR)	UK
UC Digitalis	Portugal
ZRC SAZU DARIAH group	Slovenia

Respondents of Part 2a Technology: Scoping the OA & OS Infrastructure (OSI) Landscape in Europe

OSI name	Country
4TU.ResearchData	Netherlands
Centre Mersenne	France
CORE (core.ac.uk)	UK
Croatian Scientific Bibliography CROSBIB	Croatia
Crystallography Open Database (COD)	Lithuania
Data and Service Center for the Humanities (DaSCH)	Switzerland
DataverseNO	Norway
dblp computer science bibliography (DBLP)	Germany
Digital Academic Archives and Repositories - DABAR	Croatia
Digital Repository of Ireland (DRI)	Ireland
DIGITAL.CSIC	Spain
Directory of Open Access Books, DOAB	Netherlands
Directory of Open Access Journals (DOAJ)	UK
Dissemin	France
DiVA - Digitala Vetenskapliga Arkivet/ Academic Archive Online	Sweden
e-cienciaDatos; Research Data Repository from the Consortium Madroño (Consortium of Public University Libraries from Madrid Region)	Spain
EconStor	Germany
Episciences	France
eTalk	Switzerland
Figshare	UK
Frictionless Data (FD)	UK
GRICAD	France
HAL	France
Hypergraph	Germany
Istituto per il Lessico Intellettuale Europeo e Storia delle Idee, Consiglio Nazionale delle Ricerche; ILIESI-CNR	Italy
Juelich Shared Electronic Resources (JuSER)	Germany
KIFÜ	Hungary
Lithuanian Academic Electronic Library (eLABa)	Lithuania
Mir@bel : a free gateway to journals contents	France
NARCIS - National Academic Research and Collaborations Information System	Netherlands
National Aggregator of Open Access Repositories, NEICON & NORA	Russia
National Open Access Research Data Archive (MIDAS)	Lithuania
numerev	France
OAPEN	Netherlands
OAR@UM	Malta
Open Access Monitor (OAM)	Germany
Open Commons of Phenomenology	Switzerland
Open Knowledge Maps (OKMaps)	Austria
OpenAIRE	Greece

OSI name	Country
OpenCitations	Italy
OpenEdition	France
openRDM.swiss	Switzerland
openSNP	Germany
Peer Community In (PCI)	France
Persée	France
Phaidra	Austria
Plaudit.pub	Netherlands
PoPuPS	Belgium
Portal of Croatian Scientific and Professional Journals - HRČAK	Croatia
QOAM	Netherlands
RECYT	Spain
Repository services at the National Library of Finland	Finland
Researchfish by Interfolio	UK
Riviste unimi	Italy
RUMAK	Poland
Scholia	Denmark
ScienceOpen	Germany
Sciencesconf	France
Scientific Conferences of Ukraine	Ukraine
Scientific Periodicals of Ukraine	Ukraine
SciPost	Netherlands
Semmelweis University Central Library Open Access & Open Science	Hungary
Serbian Citation Index (SCIndeks)	Serbia
Think. Check. Submit.	UK
TRAP-RCUB Repository Network	Serbia
Tudasportal	Hungary
UC Digitalis	Portugal

Respondents of Part 2b Governance and Sustainability: Scoping the OA & OS Infrastructure (OSI) Landscape in Europe

OSI name	Country
4TU.ResearchData	Netherlands
Centre Mersenne	France
Croatian Scientific Bibliography CROSBI	Croatia
Crystallography Open Database (COD)	Lithuania
Data and Service Center for the Humanities (DaSCH)	Switzerland
DataverseNO	Norway
dblp computer science bibliography (DBLP)	Germany
Digital Academic Archives and Repositories - DABAR	Croatia
Digital Repository of Ireland (DRI)	Ireland
DIGITAL.CSIC	Spain
Directory of Open Access Books, DOAB	Netherlands
Directory of Open Access Journals (DOAJ)	UK
Dissemin	France
DiVA - Digitala Vetenskapliga Arkivet / Academic Archive Online	Sweden
e-cienciaDatos; Research Data Repository from the Consortium Madroño (Consortium of Public University Libraries from Madrid Region)	Spain
EconStor	Germany
Episciences	France
eTalk	Switzerland
Figshare	UK
Frictionless Data (FD)	UK
GRICAD	France
HAL	France
Hypergraph	Germany
Istituto per il Lessico Intellettuale Europeo e Storia delle Idee, Consiglio Nazionale delle Ricerche; ILIESI-CNR	Italy
Jisc UK ORCID Consortium	UK
Juelich Shared Electronic Resources (JuSER)	Germany
KIFÜ	Hungary
Lithuanian Academic Electronic Library (eLABa)	Lithuania
Materials Cloud	Switzerland
Mir@bel : a free gateway to journals contents	France
NARCIS - National Academic Research and Collaborations Information System	Netherlands
National Aggregator of Open Access Repositories, NEICON & NORA	Russia
National Open Access Research Data Archive (MIDAS)	Lithuania
numerev	France
OAPEN	Netherlands
OAR@UM	Malta
Open Access Monitor (OAM)	Germany
Open Commons of Phenomenology	Switzerland
Open Knowledge Maps (OKMaps)	Austria

OSI name	Country
OpenAIRE	Greece
OpenCitations	Italy
OpenEdition	France
openRDM.swiss	Switzerland
openSNP	Germany
Peer Community In (PCI)	France
Persée	France
Phaidra	Austria
Plaudit.pub	Netherlands
PoPuPS	Belgium
Portal of Croatian Scientific and Professional Journals - HRČAK	Croatia
QOAM	Netherlands
RECYT	Spain
Repository services at the National Library of Finland	Finland
Research data management tool RE:COLL	Germany
Researchfish by Interfolio	UK
Riviste unimi	Italy
Scholia	Denmark
ScienceOpen	Germany
Sciencesconf	France
Scientific Conferences of Ukraine	Ukraine
Scientific Periodicals of Ukraine	Ukraine
SciPost	Netherlands
Semmelweis University Central Library Open Access & Open Science	Hungary
Serbian Citation Index (SCIndeks)	Serbia
Think. Check. Submit.	UK
TRAP-RCUB Repository Network	Serbia
Tudasportal	Hungary
UC Digitalis	Portugal

Appendix B: Survey questions

See

Proudman, Vanessa; Mounier, Pierre; Kramer, Bianca; & Bosman, Jeroen. (2020, September 24).

Survey instruments: For the Scoping the Open Science Infrastructure Landscape in Europe Study.

Zenodo. <http://doi.org/10.5281/zenodo.4048270>