publications

Toward a New World in Scholarly Communication

The 9th PUBMET2022 Conference on Scholarly Communication in the Context of Open Science

Edited by
Jadranka Stojanovski and Iva Grabarić Andonovski

mdpi.com/journal/publications

MDPI

Toward a New World in Scholarly Communication: The 9th PUBMET2022 Conference on Scholarly Communication in the Context of Open Science

Toward a New World in Scholarly Communication: The 9th PUBMET2022 Conference on Scholarly Communication in the Context of Open Science

Editors

Jadranka Stojanovski
Iva Grabarić Andonovski

Editors

Jadranka Stojanovski   Iva Grabarić Andonovski
University of Zadar   University of Zagreb
Zadar, Croatia   Zagreb, Croatia

This is a reprint of articles from the Special Issue published online in the open access journal *Publications* (ISSN 2304-6775) (available at: https://www.mdpi.com/journal/publications/special_issues/PUBMET2022).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

Cover image courtesy of Ivana Končić

Contents

About the Editors

Jadranka Stojanovski

Jadranka is an associate professor at the University of Zadar. She has played a pivotal role in shaping open access and open science infrastructure and services while effectively curating knowledge generated by the Croatian research community. Jadranka boasts a rich interdisciplinary background, holding an MSc in Physics and earning an MA and a Ph.D. in Information Sciences from the University of Zagreb. Jadranka's commitment to advancing open scholarly communication is underscored by her involvement in numerous international initiatives, projects, and communities, such as OPERAS, OpenAIRE, NI4OS Europe, PEERE, ENRESSH, UNESCO, OS MOOC, and EASE. Jadranka represents Croatia in the European Commission Expert Group on National Points of Reference on Scientific Information. Her research interests are firmly rooted in the realm of open scholarly communication, with a specific focus on peer review and research assessment, emerging trends in scholarly publishing, editorial quality, ethical considerations, and research data management.

Iva Grabarić Andonovski

Iva obtained her master's degree in Horticulture and Landscape Design at the University of Zagreb's Faculty of Agriculture and completed her postgraduate specialist study in Food Management at the Faculty of Food Technology and Biotechnology. She found the perfect ground for fulfilling her interest in scholarly communication as the Editor of the *Food Technology and Biotechnology* journal. She is the Vice-President of the EASE Association of Science Editors, the Chair of the EASE Croatian Regional Chapter, and the Co-Chair of the EASE Environment and Sustainability Committee. Iva is very active in the Croatian scientific community as one of the founders and the Secretary of the Croatian Association for Scholarly Communication (CROASC), a member of the HRČAK (Portal of Croatian Scientific and Professional Journals) Advisory Board, and a member of the Organizing Committee of the PUBMET Conference. She is also a member of the Croatian Society of Biotechnology (HDB). A keen believer in open access, she is one of the signees of the Croatian Open Access Declaration. Her current focus is on the promotion of the UN SDG Publishers Compact initiative.

Editorial

Toward a New World in Scholarly Communication: The 9th PUBMET2022 Conference on Scholarly Communication in the Context of Open Science

Jadranka Stojanovski [1,*] **and Iva Grabarić Andonovski** [2]

1   Department of Information Science, University of Zadar, Mihovila Pavlinovića 1, 23000 Zadar, Croatia
2   Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, 10000 Zagreb, Croatia; iva.grabaric.andonovski@pbf.unizg.hr
*   Correspondence: jadranka.stojanovski@irb.hr

Open access has emerged from the need to make scholarly communication freely available to the scientific community and not hidden behind a paywall. One significant catalyst of the initiative was Plan S, which mandates open access to publications resulting from projects funded by major European financial institutions within cOAlition S [1]. Although initiatives promoting open access have increased the accessibility of scientific literature, progress in scientific communication as a whole has been limited, with most other research outputs remaining inaccessible. The European Union Council has called on member states to support policies towards a non-profit, open access scholarly publishing model, encouraging the development of national open access policies which will promote equitable and sustainable scholarly publishing under open licences and apply FAIR (Findable, Accessible, Interoperable, and Reproducible) principles to research data [2]. The development of the European Open Science Cloud (EOSC) created a perfect environment for hosting and processing research data, where researchers can publish, find, and re-use data, tools, and services [3]. Open science initiatives also focus on involving the public in scientific processes through citizen science, reforming evaluation and reward systems, and establishing appropriate research infrastructures and related services. Sharing all outputs in the earliest stages of research, including research data, software, protocols, methods, etc., is the basis for the accelerated development of responsible science.

Recognizing the importance of open access early on, Croatia joined the initiative by developing the freely accessible Croatian scientific bibliography, CROSBI, in 1997, followed by several other projects, including the establishment of the HRČAK repository, which houses more than 450 Croatian OA journals, with over 90% in diamond open access, and the DABAR national infrastructure with over 120 institutional repositories [4]. Although national open science policies have not yet been adopted, the Croatian Open Science Cloud Initiative (HR-OOZ) was launched as a part of EOSC in 2021 to develop a modern scholarly system in Croatia based on open science that aligns with European research standards and initiatives [5]. Nine years ago, we launched an international conference on scholarly communication in the context of open science called PUBMET, which has continued to be held at the University of Zadar, becoming one of the major conferences on open acess that brings together researchers, editors, publishers, librarians, policymakers, experts in scientific communication, and educators.

The PUBMET2022 conference[1] gathered 120 in-person and 300 online participants over the course of three days and provided a platform for the discussion of the most current topics in open science and the trends shaping the development of scientific communication. The conference featured six invited lectures, sixteen short talks, ten workshops, a poster session, and three panel discussions on various topics related to open science. From the conference, ten submissions were carefully selected and peer reviewed for publication in

this Special Issue of *Publications*. They cover various topics, such as the landscape of scholarly book publishing, crowdfunding initiatives for publishing open access monographs by European universities, the role of universities and public libraries in promoting citizen science in Europe, open tools (for self-archiving and assessing the potential of green open access, automatic XML extraction, and e-book formatting), the promotion of open science within NI4OS-Europe and NOSCIs initiatives, the GoTriple open science platform for social sciences and humanities, the promotion of openness and transparency guidelines across journals, and mechanisms of information diffusion within academic institutions using the information space model.

These papers provide valuable insight into the initiatives and tools that promote open science, which will help readers to understand the complexity of open access and its wide impact on different aspects, not only on the openess of research data and tools, but also on the increase in the quality and transparency of scientific communication, the development of governmental policies, the inclusion of the community, and the promotion of equity, diversity and sustainability. The information presented in this Special Issue will be also of great help to institutions, organizations and policy makers, giving them an overview of successfully implemented guidelines, platforms, and tools for promoting open science, which can be of use in the development of their own open science infrastructure and national open access policies.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1]   PUBMET2022 was organized by the University of Zadar Department of Information Sciences, Ruđer Bošković Institute, Croatian Association for Scholarly Communication, University of Zagreb School of Medicine, Faculty of Food Technology and Biotechnology and Faculty of Humanities and Social Sciences, and University of Rijeka Faculty of Medicine, under the auspices of the Croatian Ministry of Science and Education, NI4OS-Europe, OpenAIRE, EASE, OPERAS and SPARC Europe.

References

1.   cOAlition, S. Plan S: Making Full and Immediate Open Access a Reality. Available online: https://www.coalition-s.org/about/ (accessed on 6 March 2023).
2.   Council of the European Union. High-Quality, Transparent, Open, Trustworthy and Equitable Scholarly Publishing—Council Conclusions (Approved on 23 May 2023). Available online: https://data.consilium.europa.eu/doc/document/ST-9616-2023-INIT/en/pdf (accessed on 7 July 2023).
3.   European Commission. European Open Science Cloud (EOSC). Available online: https://eosc-portal.eu/ (accessed on 7 July 2023).
4.   Macan, B.; Stojanovski, J. National Open Access Desk Croatia. 2022. Available online: https://www.openaire.eu/os-croatia (accessed on 6 March 2023).
5.   Celjak, D. Croatian Open Science Cloud Initiative (HR-OOZ Initiative). In *Digital Art History—Methods, Practices, Epistemologies III, Proceedings of the International Conference, Online, 12–13 October 2021*; University of Zagreb: Zagreb, Croatia, 2021. Available online: https://repozitorij.srce.unizg.hr/islandora/object/srce:494 (accessed on 25 April 2023).

MDPI

Article

Building National Open Science Cloud Initiatives (NOSCIs) in Southeast Europe: Supporting Research and Scholarly Communication

Milica Ševkušić [1,*], Eleni Toli [2,*], Katerina Lenaki [3], Kalliopi Kanavou [2], Electra Sifakaki [2], Biljana Kosanović [4], Ilias Papastamatiou [5] and Elli Papadopoulou [2]

[1] Institute of Technical Sciences of SASA (Serbian Academy of Sciences and Arts), 11000 Belgrade, Serbia
[2] Athena Research & Innovation Centre, 15125 Marousi, Greece
[3] Ministry of Education and Religious Affairs, 15122 Marousi, Greece
[4] University of Belgrade Computer Centre, 11000 Belgrade, Serbia
[5] National Infrastructures for Research and Technology—GRNET S.A., 11523 Athens, Greece
[*] Correspondence: biblioteka@itn.sanu.ac.rs (M.Š.); elto@athenarc.gr (E.T.)

Abstract: The Horizon 2020 project National Initiatives for Open Science in Europe—NI4OS Europe supports the development of the European Open Science Cloud (EOSC) by integrating 15 countries in Southeast Europe into the governance structure of this new pan-European research environment. Through a qualitative secondary analysis of the data collected during the project, the paper focuses on the main instrument developed by the project with the aim of enabling the integration of the partner countries in the EOSC—a network of national Open Science Cloud Initiatives (NOSCIs)—and explains how the concept of NOSCI and a wide range of related activities, tools, services, and resources foster research and open scholarly communication. The paper has three main sections: the first identifies challenges to scholarly communication in Southeast Europe, the second describes the methodology used to deal with these challenges revolving around the concept of NOSCI, whereas the third presents a set of indicators to track the change generated by project actions and discusses the impact of this methodology and project outputs in the area of scholarly communication.

Keywords: European Open Science Cloud; NI4OS-Europe; National Open Science Cloud Initiatives; Open Science; national policies; Southeast Europe; scholarly communication

1. Introduction

National Open Science Cloud Initiatives (NOSCIs) are national-level coalitions of Open Science stakeholders that seek to facilitate the integration of EU Member States and Associated Countries in the European Open Science Cloud (EOSC). Their establishment is associated with the acceleration of Open Science (OS) transformation and consequent scholarly communication optimisation worldwide. The purpose of this paper is to shed light on the contribution of the "National Initiatives for Open Science in Europe—NI4OS Europe" project in building NOSCIs [1] in 15 countries of Southeast Europe (Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Croatia, Cypris, Georgia, Greece, Hungary, Moldova, Montenegro, North Macedonia, Romania, Serbia, Slovenia, https://ni4os.eu/partners/, accessed on 2 September 2022) and its impact on all aspects of scholarly communication.

We consider scholarly communication to be defined as the process in which scientists share views and findings regarding their subject of research, contributing thus to the progress of peer-reviewed scientific knowledge worldwide. Closely linked to the research cycle, it is generally considered to include three distinct stages. According to Graham [2], it is a system that involves (a) the informal communication within scientific

networks—where research and idea flows are generated, (b) the initial public dissemination of research results—such as in preprints and conferences and finally (c) the formal publication of research output so that it is available to scientists and the public—mainly in scientific journals.

Technological developments of the last two decades, especially high capacity, linked networks and advanced electronic tools and services have brought major changes in all stages of scholarly communication and increased expectations for better and more open science. Research workflows nowadays reflect the great variety of sources and means, including distinct yet interdependent phases (discovery, analysis, writing, publication, outreach, assessment) [3], and scientists mostly publish in primary sources online (e.g., in working papers, preprints, reports, theses and dissertations, journals, monographs, conference proceedings and patents according to Das, 2015 [4]). Scholarly communities as a whole, i.e., scientists, universities and their libraries, publishers, government bodies, research councils, funders, as well as readers [2], have a vital impact on the scholarly communication system regarding new alternatives in access, publishing and evaluation of research content. They are responsible for creating the conditions for bibliodiversity, that is, the diversity in services, funding and evaluation mechanisms [5]. Taking this into account, their role is crucial in order to achieve the transition to open access and open science and the "accommodation of the different workflows, languages, publication outputs, and research topics that support the needs and epistemic pluralism of different research communities" [5]. The diversification of the concept of scholarly communication and its high importance is also reflected in the EOSC Portal data model, where Scholarly Communication is a subcategory of Sharing and Discovery and is further divided into Analysis, Assessment, Discovery, Outreach, Preparation, Publication, Writing and Other [6].

The relationship between EOSC and scholarly communication is reciprocal and multidimensional. EOSC should facilitate a shift from the "current state of the art towards an Open Science scholarly communication ecosystem that is based on incentives and facilitates Open Science principles and practices in performing and sharing science" [7]. At the same time, scholarly communication on the EOSC vision and services can support the further widening of OS in Europe, the engagement of research communities in EOSC and the broader use of its services. It is therefore important to highlight even more the existing links between EOSC and scholarly communication, by addressing the whole spectrum of EOSC priorities and aims at all levels (policy, governance, users, technical issues). The complexity of the EOSC universe often does not allow scholarly communities to immediately observe and understand the benefits of EOSC. Their further engagement in the EOSC discourse will certainly enrich the OS discussion in Europe.

As already mentioned, the project National Initiatives for Open Science in Europe—NI4OS Europe works towards widening OS and EOSC in Southeast Europe. It is one of the four EOSC Regional Projects and is part of the INFRAEOSC-05 Collaboration (https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/infraeosc-05-2018-2019). The project was launched in September 2019 and will end in February 2023 (more details: https://cordis.europa.eu/project/id/857645; project website: https://ni4os.eu – accessed on 6 September 2022). It supports the development of the EOSC by contributing to its portfolio of services and the establishment of NOSCIs, engaging national and regional research communities of Southeast Europe in the EOSC governance, strengthening Open Science (OS) practices and promoting the FAIR principles [8].

This paper highlights the scholarly communication perspective and demonstrates how approaches, instruments, services and tools developed during the project help to establish sustainable and networked local environments for research and open scholarly communication in terms understandable to its main target audience—the research community. The presented analysis may as well be useful to policy-makers in other regions and especially in developing countries. The region of Southeast Europe is marked by a great diversity of local contexts, due to which the project team had to devise flexible solutions that could

be replicated in various environments. This gives to the project findings (challenges and solutions) a higher value, as wider OS and EOSC communities can benefit from them.

The paper has three main sections: the first describes the data and methodology used to deal with the challenges revolving around the concept of NOSCI in the target area, the second identifies challenges to scholarly communication in Southeast Europe at the regional level, whereas the third presents a set of indicators to track the change generated by project actions and discusses the impact of this methodology and project outputs in the area of scholarly communication.

2. Data and Methodology

The paper uses a qualitative secondary analysis to show whether the solutions offered by the project have an impact on scholarly communication. It does not involve original research. The following data collected during the project, using various methods, are reused in this paper:

- Information about Open Science stakeholders in the region, collected during the NI4OS-Europe landscaping activity (expert survey) [9];
- Country sheets analysis: report from the EOSC Executive Board Working Group (WG) Landscape (expert survey) [10];
- Survey data collected in the autumn of 2019 for the NI4OS-Europe landscaping study (survey) [11];
- Analysis of Open Science policies in the region and across Europe (desk analysis) [1];
- Project partners' reports about NOSCI development (expert survey).

Publicly available statistical data provided by EUROSTAT and the World Bank are also used.

Based on these sources of information, and primarily the stakeholder information and the survey data, direct and indirect challenges to scholarly communication were identified. Desk analysis is used to explain the project's response to these challenges provided through the overarching concept of NOSCI. The analysis is entirely focused on the regional level and is limited to the challenges that are specific to the region of Southeast Europe and those that are shared by the countries involved in the project. Challenges present in individual countries go beyond the scope of this paper. The reason for this lies in the fact that the project has not dealt with each country as an isolated case study but has rather sought to devise more general solutions that are locally applicable.

To measure the impact of the response, we introduce a set of indicators derived from the identified challenges. In the process of defining the indicators for this study, the indicators for monitoring EOSC readiness [12] and NOSCI establishment [1,13] were analysed. Keeping in mind that these indicators are intended for monitoring at the national level and do not always apply to scholarly communication, it was necessary to select those that are relevant for scholarly communication and adjust them to regional-level monitoring. The mapping between the indicators defined for this study and the NI4OS-Europe Blueprint metrics [13] and EOSC readiness indicators [12] is shown in Table 5.

The analysis follows the challenge—response—impact matrix, which is reflected in the structure of the paper.

3. Landscaping

3.1. Remarks on the Landscaping Survey and Regional Challenges

The challenges to scholarly communication in the NI4OS-Europe partner countries were identified at the regional level and are based on the data collected during the landscaping activity at the beginning of the project. The data include the literature data, information provided by national experts, and, most importantly, the data collected in a survey conducted in the autumn of 2019, which also provided an input to the EOSC Secretariat's Landscape Activity The five INFRAEOSC-5b projects conducted landscaping activities in a coordinated manner and their inputs were eventually aggregated and analysed in a study commissioned by the EOSC Secretariat [14]. This initial mapping of the existing

Open Science (OS) initiatives, infrastructures, services, policies, stakeholders and topics in each of the partner countries at the beginning of the project [11], which helped to shape further actions, serves as the starting point for the identification of challenges to scholarly communication in our analysis. The relevance of information collected on this occasion for the present analysis is best illustrated by the fact that 61.2% (30 out of 49) of the questions in the landscaping questionnaire for stakeholders performing research. The survey included five questionnaires—one for each stakeholder group. In this paper, we are focusing on the main actors of scholarly communications—those who perform research. More information about the stakeholder groups and the structure of the survey: [11] were directly related to scholarly communication. Table 1 summarises the topics covered by the questionnaire and indicates their relation to scholarly communication.

Table 1. Topics covered in the questionnaire for research performing stakeholders: relation to scholarly communication.

Main Topic	Specific Topics	Relation to Scholarly Communication	No. of Questions
General			4 + 1
Profile of the organization	domain and size		3
	content and rights	Direct/Publication	1
Funding	criteria	Indirect	2
	user support (for services)	Indirect	1
Policies	institutional	Direct/Discovery, Outreach, Publication, Assessment	1
	infrastructure roadmap	Indirect	1 + 1
	OS compliance	Direct/Discovery, Outreach, Publication, Assessment	1
	research assessment	Direct/Assessment	1 + 2
Infrastructure	for OS	Direct/Discovery, Outreach, Publication	1
	needs and preferences	Indirect	1
Services	needs and preferences	Indirect	1
Open Science	training	Direct/Skills and competencies	3
	publication repositories	Direct/Discovery, Outreach, Publication	1 + 6
	data repositories	Direct/Discovery, Outreach, Publication	1 + 7
FAIR	awareness	Direct/Analysis, Preparation, Writing, Discovery, Outreach, Publication	1
	implementation	Direct/Analysis, Preparation, Writing, Discovery, Outreach, Publication	3
	support	Direct/Skills and competences	1
EOSC	awareness	Indirect	2 + 2

The "direct" relation means that a question refers to one or more stages of scholarly communication, policies, infrastructure, services, skills and competencies for scholarly communication, or the assessment of research outputs. For example, the investment in a FAIR-compliant repository creates a channel for the dissemination of research outputs. The subcategories of scholarly communication from the EOSC portal [6] are used in the table to describe more precisely the direct relation. The "indirect" relation means that the subject of a question has an indirect impact on scholarly communication, e.g., the main purpose of the user support for services is not necessarily to improve scholarly communication; however, it may not only have such an effect, but it is very likely that researchers will see it as being in the service of scholarly communication. This "scholcomm-centric" perspective among researchers is reflected in the responses to the question "What do you expect from EOSC?", which predominantly referred to concepts directly related to scholarly communication (Table 2) This open-ended question was answered by less than 40% of the respondents. The respondents could mention as many concepts as they wished, and some even provided

descriptive answers. Their answers were analysed to extract distinct concepts, which were further normalised. Only the concepts mentioned more than five times are shown in the table.

Table 2. Researchers' expectations from EOSC: survey results.

Concept	Mentions
access	62
data	44
services	27
support	16
training	15
computing resources and services	15
sharing	13
collaboration	10
Open Science	10
research outputs	10
Open Access	9
repositories	9
information	8
infrastructure	7
integration	6
networking	6
publications	6
resources	6
guidance	5
Open Data	5
research support	5
standardisation	5

Southeast Europe is a highly diversified region in political, social and economic terms, which determine the research environment and also have an indirect impact on scholarly communication.

Accordingly, the NI4OS-Europe project team had to take into account some overall challenges related to this broader social and political context. Over the past three decades, the region has witnessed turbulent changes. Out of the 15 partner countries, 9 were part of 2 federations that no longer exist—USSR (Armenia, Georgia, Moldova) and the former Yugoslavia (Bosnia and Herzegovina, Croatia, Montenegro, North Macedonia, Serbia, Slovenia). Eight countries (Albania, Bulgaria, Croatia, Hungary, Montenegro, North Macedonia, Romania, Slovenia) became members of the North Atlantic Treaty Organisation (NATO), four of them (Albania, Bulgaria, Hungary, Romania) had previously belonged to the opposing military alliance (Treaty of Warsaw). Thirteen partner countries have experienced major political transformations as a result of the collapse of communism. Three decades ago, only Greece was a member of the European Union. In the meantime, the EU has been extended to include six more partner countries (Bulgaria, Croatia, Cyprus, Hungary, Romania, Slovenia) [15]. Finally, the political and military conflicts that accompanied the break-up of Yugoslavia disrupted the research environment in the Balkans and it took 15 years to recover research collaboration to the level before the conflicts [16]. This broader socio-political and historical context will not be discussed further, but it has to be taken into consideration, as it largely explains infrastructural fragmentation and the lack of collaboration towards developing, for example, regionally coordinated publishing services or aggregators.

In the following sections we provide some additional insights on the challenges identified through our landscaping activity that are of particular interest for scholarly communities.

3.2. The Place of Open Science in the Research Governance Systems

Among other things, the success of initiatives and policies is related to the position of the authority or body behind them to impose relevant regulations and mobilise various stakeholders and research communities. In most countries in the region, regulations relating to OS were adopted by ministries responsible for education, science and research. OS may also be included in a national digital agenda. In Hungary, the Ministry of Innovation and Technology is responsible for OS, while in Georgia, OS policy development is coordinated by the Shota Rustaveli National Science Foundation. Since 2019, research activities in Greece are under the responsibility of the Ministry of Development and Investment while at the same time a general framework for open science is included in the Digital Transformation Strategy 2020–2025 of the Ministry of Digital Governance. It is noteworthy that in Cyprus OS policy-related activities are supported by OpenAIRE, whereas in Armenia the policy is developed within an Erasmus+ project [8].

The question inevitably arises whether individual ministries or funders are well-positioned to cover all relevant topics in policies and ensure their wide implementation. This is also a challenge for scholarly communication because policies and funding relating to Open Science developed by a single ministry may not address all relevant aspects of scholarly communication. On the other hand, involving all relevant ministries, funders and stakeholders in the policy development and implementation process may be difficult to coordinate.

3.3. Different Policy Traditions

Although the landscaping survey addressed various levels (national, institutional) and aspects of policies (open access to publications and research data, sharing software under free licences, the preservation of scientific information, information and data security, rules regarding repositories, publishing platforms, FAIR principles, intellectual property rights, access to services and terms of use), the responses were inconsistent, especially as regards institutional policies [11]. Additional efforts were made to refine the data through an expert survey [1,8].

In the autumn of 2019, most countries did not have a national OS policy. The existing policies mostly addressed publications. Some also mentioned research data, but very few discussed the FAIR dimension of them. This suggests a rather conservative perception of scholarly communication as being limited to communicating research through publications only.

In most countries it was possible to find some documents addressing OS explicitly or individual aspects of OS implicitly. These documents took various formats (platform, strategy, agenda, plan) and approaches (mandating vs. recommending), while the documents were adopted by different bodies. Due to this, devising a single model for the alignment of policies with OS principles would be ineffective.

3.4. Non-EU Countries Are Less Integrated in European Open Science Networks

Throughout the NI4OS-Europe project, participation of the partner countries in the so-called "EOSC pillars" (consortia, associations or networks contributing to the development of EOSC) has been monitored as an indicator of progress towards EOSC integration. The list of the "pillars" is as follows: OpenAIRE (OpenAIRE—Open Access Infrastructure for Research in Europe) is a European infrastructure for Open Science: https://www.openaire.eu, accessed on 20 July 2022); European Grid Infrastructure (EGI, which seeks to provide access to high-throughput computing resources in Europe using grid computing techniques: https://www.egi.eu, accessed on 20 July 2022); Research Data Alliance (RDA, a research community organisation striving to facilitate open data sharing at a global level, founded in 2013 by the European Commission, the American National Science Foundation and National Institute of Standards and Technology, and the Australian Department of Innovation; it relies on a network of national RDA nodes: https://www.rd-alliance.org/, accessed on 20 July 2022); and GÉANT (Gigabit Research and Education Network, which is

a pan-European network connecting national research and education networks (NRENs) across Europe: https://geant.org/, accessed on 20 July 2022). The list has been limited to the "ones that are relevant for the majority of the NI4OS-Europe countries" [1]. In line with the focus of this study, we will use the term "Open Science networks" and extend the list to include a number of organisations, initiatives or European Research Infrastructure Consortia (ERICs) involved in EOSC-related projects and the development of infrastructure, services, tools, guidelines and best practice relevant for open scholarly communication. These are EUDAT (European Collaborative Data Infrastructure, which is a European infrastructure that integrates data services and resources supporting research: https://www.eudat.eu/ accessed on 20 July 2022); National Research Infrastructure Roadmap (https://www.esfri.eu/national-roadmaps, accessed on 20 July 2022); CESSDA (Consortium of European Social Science Data Archives, which is a European Research Infrastructure Consortium that offers data services to the social sciences by bringing together social science data archives across Europe: https://www.cessda.eu/, accessed on 20 July 2022); DARIAH (Digital Research Infrastructure for the Arts and Humanities, which is a European Research Infrastructure Consortium that supports digitally-enabled research and teaching across the arts and humanities: https://www.dariah.eu/, accessed on 20 July 2022); EIFL (Electronic Information for Libraries, which is a not-for-profit organization supporting libraries in developing and transition economy countries to gain access to knowledge: https://www.eifl.net/, accessed on 20 July 2022); CLARIN (Common LAnguage Resources and Technology INfrastructure, which is an ERIC offering language data, language technology data processing and expertise to the research community: https://www.clarin.eu/, accessed on 20 July 2022); and OPERAS (a European Research Infrastructure dedicated to open scholarly communication in the social sciences and humanities: https://www.operas-eu.org/ accessed on 20 July 2022). Involvement in such networks offers to participants various opportunities (e.g., to participate in projects and training; qualify for technical support; have access to infrastructure, tools and services, data and guidelines; and exchange information, etc.).

Figure 1 shows the NI4OS-Europe partner countries represented in each network at the beginning of the project, in the autumn of 2019 (light grey indicates pending initiatives to join a particular network. In all four cases, the initiatives were successful). It is apparent that the non-EU countries are considerably less involved in the activities of the selected networks. We did not investigate the reasons for this, and we can merely speculate that they may include the lack of familiarity with the networks, the lack of interest among relevant stakeholders at the national level, the lack of funds to pay participation fees (where required), etc.

	Country	Open AIRE	EUDAT	EGI	GEANT	National RI Roadmaps	RDA NODE	CESSDA	DARIAH	EIFL	CLARIN	OPERAS
EU	Bulgaria	■		■	■	■	■	■	■		■	
	Croatia	■	■	■	■	■	■	■		■	■	■
	Cyprus	■	■			■		■	■	■	■	
	Greece	■	■	■	■	■	■	■	■	■	■	■
	Hungary	■		■	■		■					
	Romania	■			■	■						
	Slovenia	■		■	■	■	■	■	■	■	■	■
Non-EU	Albania				▓					▓		
	Armenia	▓										
	Bosnia & Herzegovina	░						▓				
	Georgia	░			▓							
	Montenegro	░						▓				
	Moldova				▓					▓		
	North Macedonia			▓	▓					▓		
	Serbia	▓										░

Figure 1. Participation of the NI4OS-Europe partner countries in European Open Science networks.

Poor involvement with international OS networks bears risks, such as infrastructure obsolescence or investment in unsustainable tools and services that are not compliant with standards. This may have negative effects on scholarly communication, especially in the long term. Local stakeholders may stay behind major developments in this area, such as the development of Open Access book publishing platforms, efforts to create multilingual controlled vocabularies, discussions about research evaluation, etc.

3.5. Varying Structure of OS Stakeholders across Countries

The mapping of OS stakeholders in the partner countries was part of the landscaping action at the beginning of the project [11]. The data were provided by the project partners after the stakeholder groups had been defined. This was done in two steps: the project partners provided either a preliminary list of the stakeholders or just the numbers per group; the project partners were incited to revise the preliminary lists or to provide a list with contact information (if they had previously provided only numbers). The stakeholder map was made based on the data collected in the second step. The main purpose of this action was identifying institutions, infrastructures and services to be targeted by project activities and, in particular, providing input for the model of national OS initiatives and the EOSC Landscape Activity [14].

Five stakeholder groups were defined:

1. Funders and policymakers—FUND (the actors who fund research and, most commonly, shape research-related policies;
2. Those who perform research—CREATE (research performing organisations and researchers);
3. Those who perform research—SUPPORT (repositories, research infrastructures, e-infrastructures, service providers, libraries);
4. Those who "consume" research—CONSUME (SMEs and citizens);
5. OS facilitators (including OS initiatives)—FACILITATE (international nodes, coordinators).

Table 3 shows the number of stakeholders per group identified in the partner countries. The completeness and reliability of the data may be disputed, as it is possible that the project partners were more familiar with a particular group of stakeholders or failed to identify others. In addition, there are some differences between the initial input and the final map, e.g., there are no members of the FUND group for Croatia in the final map, whereas during the initial data collection there were two. Similarly, there are no SUPPORT stakeholders for Montenegro in the final map, whereas there were 10 during the initial data collection. We assume that the data in the map are more accurate because during the initial data collection some partners provided just numbers. However, all project partners are organisations that are expected to have a good insight into the local situation and it is reasonable to assume that the data provided by them do not significantly deviate from the actual situation.

It is interesting that Cyprus and Greece have considerably more research funders and policy makers than the other countries, which may suggest less centralised research systems. Slovenia has the greatest number of EOSC facilitators (EOSC Working Group representatives, CESSDA representatives, etc.), which reflects its presence in various European networks. Apart from researchers, the SUPPORT group is the most directly relevant for scholarly communication, as it includes stakeholders who provide and maintain infrastructure (repositories, publishing platforms, e-infrastructure) and provide relevant services (metadata creation, PID assignment, dissemination, etc.), as well as training. If this group is insufficiently represented in a country, there is a risk that infrastructure development and skills will lag even if relevant policies are in place and funding is sufficient.

Table 3. Open Science stakeholders in the NI4OS-Europe partner countries.

Country	Fund	Create	Support	Consume	Facilitate
Albania	2	14	3	0	0
Armenia	4	23	0	1	0
Bosnia & Herzegovina	3	11	10	6	1
Bulgaria	3	17	8	5	3
Croatia	0	54	15	3	0
Cyprus	20	32	13	7	0
Georgia	1	14	4	1	0
Greece	15	58	48	11	8
Hungary	6	52	28	4	8
Moldova	3	46	30	2	5
Montenegro	1	5	0	6	0
N. Macedonia	4	21	4	3	0
Romania	6	27	33	24	2
Serbia	4	106	17	10	3
Slovenia	6	63	43	4	16

3.6. Varying Levels of Available Funding

Investment in research development in the NI4OS-Europe partner countries is generally lower than in the rest of Europe, and this is primarily due to their lower economic performance. While most EU members among the partner countries belong to High Income Economies according to the World Bank ranking [17], the associated countries mostly fall into the group of Upper-Middle-Income Economies. Table 4 presents the research and development expenditure in the countries in the region, which is considerably lower than in the EU: in all partner countries it is below the EU average (2.2% of GDP) and only in Slovenia is it close to the EU average [18]. EU members among the partner countries were more actively involved in Horizon 2020 projects and they, accordingly, received more funds [19].

Table 4. NI4OS-Europe partner countries in 2019: World Bank Country Ranking, research and development expenditure (percent of GDP) and the net contribution in Horizon 2020.

Category	Country	World Bank Country Ranking Income	Research and Development Expenditure (% of GDP)	Net EU Contribution in Horizon 2020 (Million EUR)
EU members	Bulgaria	UM	0.83236998	161.25
	Croatia	H	1.08105004	137.97
	Cyprus	H	0.71463001	319.47
	Greece	H	1.27566004	1710.82
	Hungary	H	1.47736001	371.13
	Romania	H	0.47832	300.69
	Slovenia	H	2.04703999	379.85
Non-EU members	Albania	UM	no data	5.79
	Armenia	UM	0.17854001	4.20
	Georgia	UM	0.28468001	8.71
	Moldova	LM	0.23672	7.36
	Bosnia & Herzegovina	UM	0.19264001	8.72
	Montenegro	UM	0.36328	4.62
	N. Macedonia	UM	0.36783001	14.80
	Serbia	UM	0.88666999	134.85

Sources: World Bank; Horizon Dashboard. LM—Low-middle-income. UM—Upper-middle-income. H—High income. GDP—Gross Domestic Product. EU—European Union.

The size and structure of funding for scholarly communication in individual countries are difficult to estimate. This topic was beyond the scope of the NI4OS-Europe landscaping

activity. On the qualitative level, it is important to mention that the survey data reveal that in all countries in the region publicly funded institutions provide funding and technical support for repositories and journal publishing platforms. In Bulgaria, Croatia, Romania, Serbia and Slovenia, national funders provide subsidies for scholarly journals, but only in Croatia and Slovenia is support provided only to Open Access journals [20].

3.7. The Lack of Information across the Region

At the beginning of the project, there were no curated national or regional registries offering information about relevant institutional stakeholders, nor were there any standard-ised and curated national or regional registries of services relevant for OS. The information provided by international registries, such as OpenDOAR, re3data, FAIRSharing, etc. was incomplete. The most reliable source of information about OS policies were the OpenAIRE country pages, for the countries involved in OpenAIRE. Due to this, it was impossible to verify the information about infrastructure and policies collected in the landscaping survey [11].

The lack of information is also a serious challenge for scholarly communication. It makes it difficult to identify potential partners, which may discourage collaboration. Additionally, there is a risk of effort duplication, instead of replicating or sharing sustainable solutions developed within the region.

3.8. The Lack of Incentives for Open Science

Two sources of information about incentives for OS were used: the landscaping survey data (questions about institutions' internal rules and the aspects taken into account when evaluating researchers) and an expert survey presenting four case studies—Croatia, Greece, Serbia and Slovenia [8].

The responses to the question about internal rules for OS-related topics (intellectual property rights, open-source software, publishing platforms, article and processing charges) suggest that in most institutions across the region, OS practices were encouraged. However, the data are inconsistent with the findings of the expert survey and we will not analyse them further. As far as research evaluation is concerned, the survey data show that publications were identified as the most important parameter for researcher evaluation in the NI4OS-Europe partner countries. On the other hand, data sharing, software, OS and OA, social outreach, knowledge transfer, and citizen science were recognised as little to moderately important. The radar diagram of Figure 2 is based on the responses to the question "Which of these aspects are taken into account most when evaluating researchers?" provided by researchers and representatives of research performing organisations (CREATE). The respondents were expected to assess 14 options (13 options presented in the diagram and "Other") as not important, little important, moderately important, important, or very important. The responses were translated into numerical values and an average value was calculated.

In brief, throughout the region, there were hardly any incentives for Open Science ac-tivities beyond policy encouragement. This lack of incentives did not encourage innovation in scholarly communication and it can be argued that it undermined the development and sustainability of emerging platforms and initiatives.
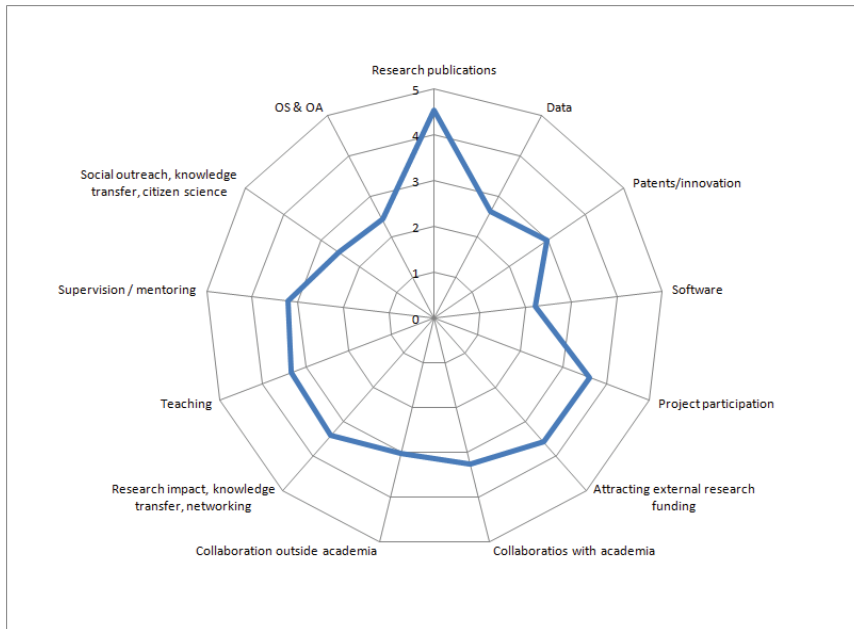
Figure 2. The importance of various parameters in research evaluation (survey data).

3.9. Uneven Infrastructure Development

The project did not deal specifically with the infrastructure for scholarly communication. It covered generic (cloud computing, data archiving and discovery services, etc.) and thematic services, as well as repositories [21]. In line with the inclusive definition of scholarly communication, all these services support various phases of scholarly communication (discovery, analysis, and partly publication). However, publishing platforms and services for writing, outreach, and assessment were beyond the scope of the project, though some of them (journal publishing platforms and e-learning resources) were captured in the landscaping analysis. As the comprehensive service catalogues are yet to be developed and the survey data are often inconsistent—which may reflect poor awareness and a lack of interest—we do not have sufficient data for a detailed analysis of the infrastructure in the region. Still, some immediately observable facts clearly demonstrate that access to the infrastructure relevant for scholarly communication varies across countries. Although it is apparent that, for example, thematic services for archaeology and heritage research are more developed in Greece and Cyprus than in other countries in the region, we will not try to make any conclusions based on the distribution of thematic services mainly because it was impossible to assess their availability to various research communities in a country and, accordingly, their impact.

The NI4OS-Europe partner countries that are members of the European Union and are involved in various European networks have better access to shared, international infrastructures than the Associated Countries, though this gap has been mitigated through projects aiming to establish pan-European infrastructures (Grid computing infrastructure, OpenAIRE services, etc.) [1].

The integration of repositories with OpenAIRE could be another indicator of infrastructure development. In order to be harvested by OpenAIRE, repositories have to meet certain technical requirements. A small number of repositories from a country may suggest either that there are no repositories, that the existing repositories are not interoperable, or that there is a lack of awareness about the importance of interoperability with infrastructures, such as OpenAIRE. Croatia and Serbia had the largest number of publication repositories

integrated with OpenAIRE. At the same time, there were five countries (Albania, Bosnia and Herzegovina, Georgia, Montenegro, and Romania) that had no repositories harvested by OpenAIRE. It is noteworthy that Croatia had the greatest number of repositories thanks to a publicly funded national repository infrastructure [22].

In the landscaping analysis, data repositories were discussed separately from publication repositories. A small number of data repositories were identified in the survey (14), five of which were in Greece and three were in Slovenia. The data repository registry re3data (https://www.re3data.org/, accessed on 26 July 2022) listed data repositories from Bosnia and Herzegovina, Croatia, Greece, Hungary, Romania and Slovenia. However, one of the repositories was incorrectly associated with Bosnia and Herzegovina, instead of Serbia [11].

Poorly developed infrastructure can not only limit the visibility of local research but can also discourage research communities from adopting open practices.

3.10. Undeveloped Training for Open Science

The question about the training and support provided by institutions was strongly focused on various aspects of scholarly communication (repositories, research data, publishing platforms, persistent identifiers, licences, article and book processing charges, intellectual property rights, open-source software, open educational resources and open practices). The survey results have so far been analysed in various contexts using the full range of data for all countries and all stakeholder groups [8,23]. It was concluded that institutions in the NI4OS-Europe partner countries mostly provided training in intellectual property rights and copyright (47%) and repositories (40%), while only 26% provided training in research data management—publishing of open data, FAIR, RDM plans, data protection, data curation, and long-term preservation, as shown in Figure 3. According to the survey data, 22–38% (depending on the topic) of the respondents did not plan to provide support or training [8]. The number of responses collected per country varied and doing the analysis for individual countries did not make sense.

	Repositories	Research data	Publishing platforms	PIDs	Licenses	IPR	APC/BPC	Open-source software	OER	Open practices
yes	62.77%	36.17%	39.36%	37.23%	45.74%	54.26%	22.34%	40.43%	38.30%	43.62%
planned	18.09%	44.68%	27.66%	29.79%	20.21%	13.83%	22.34%	27.66%	35.11%	28.72%
not planned	8.51%	9.57%	19.15%	17.02%	18.09%	19.15%	37.23%	18.09%	14.89%	15.96%
don't know	8.51%	5.32%	10.64%	12.77%	11.70%	8.51%	14.89%	8.51%	5.32%	8.51%
other	2.13%	4.26%	3.19%	3.19%	4.26%	4.26%	3.19%	5.32%	6.38%	3.19%

Figure 3. Training and support provided by libraries, e-infrastructures, research infrastructures and service providers (survey data).

If we limit the analysis to the SUPPORT group, which includes stakeholders whose mission is related to support and training (libraries, e-infrastructures, research infrastructures and service providers), the share of those who did not plan to organise training is somewhat smaller (8–37%). However, the percentage of those who did not know whether their institutions offered training was not insignificant (5–13%). It is also alarming that for most topics less than half of the respondents from the SUPPORT group provided training.

Unfortunately, we may only speculate about the reasons for the fairly low training and support offer because this issue was not covered by the survey.

3.11. Linguistic Diversity

More than 10 different languages, none of which is considered a major European language, are spoken in the NI4OS-Europe partner countries. Five different alphabets are used, which made it difficult to find or identify information in the local languages on institutional websites and analyse policies and services.

In most countries in the region, English is not so widely used in research communities as in other regions of Europe. In practice, this means that the efficiency of training provided in English may be limited. The reusability of materials in local languages is also limited and can be achieved only in some clusters (Greece and Cyprus, Romania and Moldova, among Slavic-speaking partner countries). This is also a challenge in scholarly communication, especially in those disciplines where local languages are predominantly used.

4. National Open Science Cloud Initiatives

EOSC constitutes a major ambition in the European Open Science policy, being a federated ecosystem of research infrastructures, e-infrastructures and services that allow the scientific community to share and process publicly funded research results and data across borders and scientific domains. Having the researchers in its core, early enough it was understood that efficient ways needed to be found to engage with researchers and scholarly communities, and clearly convey the message of the whole new possibilities provided by EOSC initiative to the production of research and innovation. Within its scope and ambitions, the EOSC reinforces Open Science, Open Innovation and Open to the world policies and fosters best practices of global data findability and accessibility (FAIR data); helps researchers to get their data skills recognised and rewarded; helps to address issues of access and copyright (IPR) and data subject privacy; allows easier replicability of results and limit data wastage; and it contributes to clarification of the funding model for data generation and preservation, reducing rent-seeking and priming the market for innovative research services.

EOSC stakeholders and related projects had to answer the question of how to promote and support the implementation of the above not only at an overall European level but within their countries and their national research ecosystems. For NI4OS-Europe, an additional difficulty has been posed by the region itself. Acting in an area with high diversity and various OS maturity levels, the implementation of the above cannot take place in a homogenised way. Out of this need, the concept of the National Open Science Cloud Initiatives (NOSCIs) has been developed in response to the specific traits and challenges in the targeted region, based on complex and multilayered analyses of stakeholders, policies and local contexts [1]. NOSCIs can be considered as a coalition of national organisations that have a prominent role and interest in the EOSC and have as their main aim the promotion of synergies at the national level, and the optimisation/articulation of their participation in European and global challenges in this field of Open Science. Similar to scholarly communication, NOSCIs are inclusive and require the involvement of stakeholders from across the research lifecycle. Connecting them at the national level provides not only a testbed for the formulation of OS policies but also a forum for knowledge dissemination and sharing.

To support the establishment of the NOSCIs, the NI4OS-Europe project proposed the Blueprint [1]. This is a holistic framework, inspired by Open Science models and guidelines. It includes modular workflows, a set of indicators and operational aspects for facilitating the establishment, governance and operation of the national initiatives. It adopts an agile approach, as national initiatives can have different formats of organisation and levels of maturity—the NI4OS-Europe Blueprint can be seen as a general "best-case scenario" guideline that gives to countries or to national initiatives maximum flexibility, while making sure that all aspects important to them are addressed.

As part of the Blueprint, a simplified five-step methodology for the establishment of the NOSCIs was introduced, presented in Figure 4. For the purposes of this paper, it may be interesting to highlight its structured yet inclusive approach, which proportionally addresses several aspects of the research workflow phases in scholarly communication: discovery, analysis, writing, publication, outreach and assessment.

Blueprint
Workflows | The five-steps methodology

Figure 4. A blueprint of the five-step set-up methodology for the NOSCIs.

As in any solid start of an endeavour of this size involving various actors with diverse capacities, an essential first step is to identify local EOSC & OS stakeholders and their roles and in turn design and establish proper communication workflows between them. A landscape review on available (e-)infrastructures and training should follow to create a deep understanding of the current status and bring all stakeholders up to date and on the same path. Communication among different parties is important, therefore regular meetings should be planned. To ensure the sustainability of the whole endeavour, communication and engagement of relevant government officials are crucial elements, which should be confirmed right from the beginning. Last but not least, reaching out to the wider public to communicate the status and goals of Open Science Cloud in the country; informing the public of EOSC updates will advance synergies at the national level and strengthen links to the EOSC. Organising events targeted at the wider public will introduce the NOSCI (even if still under formation) to all Open Science Cloud-related national communities (users, developers, infrastructure providers, funding agencies, related public bodies, industry, etc.).

In addition to the organisational, operational and governance aspects, the Blueprint stresses the important role that policies play in the sharing and promotion of research outputs. In fact, it is important that OS policies support a free flow of knowledge and data and overall access via the Internet, starting from the three main outputs of research: literature, data and software. This approach to the various research outputs, together with the way infrastructures and services are offered and the framework on research assessment and capacity building through skills and training, are or should be important elements in any national OS strategy.

5. Impact

The various landscaping activities in relation to EOSC have revealed the different levels of Open Science and EOSC readiness across European Countries. These efforts have been carried out at two levels, by the INFRAEOSC-5 projects through the dedicated thematic task force and by the Landscaping Task Force of the EOSC Executive Board (which was part of the previous governance phase of EOSC before the establishment of the Association. Landscape Working Group | EOSCSecretariat. Retrieved 22 October 2020, from https://www.eoscsecretariat.eu/working-groups/landscape-working-group).

The landscaping has been one of the very first activities for the projects and significantly contributed to the creation of strong collaborative links among them. The outcomes of the landscaping activities soon sparked a discussion among EOSC stakeholders and EOSC supporting projects about the necessity of having an accurate understanding of the status of EOSC readiness in each country and about the methodology and steps that are needed to measure it. This is important not only to understand the starting point for setting up national initiatives supporting EOSC but also to monitor within each Member State both the progress of EOSC as a whole, as well as particular aspects of it, with scholarly communication being one of them.

It is in this frame that the NI4OS-Europe project presented its Blueprint, including an indicative set of metrics that may be used for the assessment of the status and progress of the NOSCI in the region, which is in line with the EOSC Readiness Indicators identified by the former Landscaping Task Force. The NI4OS-Europe indicators have not been created to specifically measure aspects of scholarly communication but rather have a broader scope and can be predominantly used as a guide to complement the establishment and operation of a NOSCI.

Building on this previous work on indicators, the current paper delivers a contribution in relation to scholarly communication at three levels. First, it highlights a subset of indicators of interest for measuring activities in scholarly communication. They are derived from the NI4OS Blueprint metrics and are listed in the "Indicator" column in Table 5. Then, it does a parallel mapping between the scholcomm-related indicators and the NI4OS-Europe Blueprint and the EOSC Readiness Indicators. Finally, it establishes a relation between the proposed indicators relevant for scholarly communication and a number of OS-related challenges identified in the region. To the best of our knowledge, Table 5 provides the first attempt to map and adjust existing indicators in relation to challenges in scholarly communication. In the creation of the blueprint metrics, their reusability potential has been a major criterion from the outset. Now, this selection process in relation to scholarly communication further increases their reproducibility.

Table 5. The challenges to scholarly communication in the NI4OS-Europe countries and the mapping between the impact indicators for scholarly communication and the NI4OS-Europe Blueprint metrics and EOSC readiness indicators.

Challenge	Indicator	Mapping with NI4OS-Europe Blueprint Metrics	Mapping with EOSC Readiness Indicators
Different position of OS in various research governance systems	NOSCIs established	1.a.1. NOSCI established 1.b.1. Form of organisation	National initiative in place/planned
Different policy traditions	Open Science policies are in place	1.d.3. National/Institutional policies around Open Science	OS/FAIR policies supported/ monitored/ planned
Non-EU countries are less integrated in European OS networks	Membership in EOSC and major European OS networks	4b. Metrics concerning membership in international bodies/fora: 4.b.1. EOSC Association participation 4.b.2. EOSC pillars participation 4.b.3.Other	Integration in the EU bodies
Varying structure of OS stakeholders across countries	Multistakeholder representation in the NOSCIs	1.c.1 Membership: number of organizations 1.c.2 Membership: type of organisations	National initiative in place/planned Stakeholders involved

Table 5. *Cont.*

Challenge	Indicator	Mapping with NI4OS-Europe Blueprint Metrics	Mapping with EOSC Readiness Indicators
Varying levels of available funding	Funding for scholcomm infrastructure and activities is addressed in OS policies	4.a.1. National Fund for OS/OSC in place/ planned 4.a.2. Funding: national OSC project 4.a.3. Funding: International and European OSC Projects	National initiative in place/planned Funding Specific funding programmes
Lack of information across the region	New sources of information created	2.c.3. National Open Science portal	N/A
Lack of incentives for Open Science	Incentives for OS are included in policies	1.d.3. National/Institutional policies around Open Science	OS/FAIR policies supported/monitored/ planned—Incentives
Uneven infrastructure development	Improved infrastructure	2.a.1. Number and Type of infrastructures 2.a.2. Access policies in place	Architecture indicators National (regional) registry or other federation mechanisms for data in place/planned National (regional) dataset catalogue(s) in place/planned National PID Policy User satisfaction Data usage Interoperability API usage
	Improved access to infrastructure	2.b.1. Number of services offered 2.b.2. Enrolled services	
	Regional registry	2.c.1. National (regional) registry or other federation mechanisms for data in place/planned 2.b.6. FAIR-enabling services	
Undeveloped training for Open Science (landscaping data)	Training and skills resources created	C. Training & Skills 3.a. Community Metrics 3.a.1 National/ regional curricula in place/planned	National/regional curricula in place/planned
	Training and skills and programmes implemented	3.a.2. Basic training available for researchers and research support staff (e.g., National Competence centres 3.a.3. Number of trained people per year 3.a.4. Number of training material	Basic training available for researchers and research support staff People trained per year
Linguistic diversity	Relevant information is made available in local languages	N/A because NOSCI indicators are applied at the national level, while this is a regional challenge	N/A

The following passages briefly summarize the status of the scholcomm-related indicators in SEE. Although a progress from the initial stage is obvious for all identified challenges, some issues cannot be resolved at the regional level and the final success depends on the actions taken by national initiatives [24].

The establishment of 15 NOSCIs in SEE, one of NI4OS-Europe project outputs, is characterized by multistakeholder governance models and forms, such as task forces, consortia, and national projects. Their role is considered prominent concerning the development of open science ecosystems, especially the EOSC vision, and they consequently have a remarkable impact in the research environment and scholarly communication.

In this context, the formation of national and institutional OS policies becomes a priority. Interested stakeholders join forces to manage OS as a national issue and agree on a common framework by signing a Memorandum of Understanding (MoU). The newly developed national and institutional OS policies explicitly address various aspects of OS, also encompassing incentives for open science activities.

The non-EU countries are encouraged and decisively supported to extend their participation in OS initiatives and networks. The adoption of best practices recommended by NI4OS-Europe improves their scholarly environment and makes them achieve compliance

with the EOSC Rules of Participation. In particular, the partner countries are encouraged to join the EOSC Association.

Furthermore, in order to promote capacity building for scholarly communication, OS policies address the issue of raising funds from sources of various levels. The ultimate goal is thus to absorb a share from the national budget and to take advantage of all possible opportunities in European and international funding programs. It is important to note that partner countries, including non-EU members, are now included in Europe's strategic plans and Horizon Europe's Working Programs (e.g., European Commission, Directorate-General for Research and Innovation, Strategic foresight in the Western Balkans: recovery on the horizon, Publications Office of the European Union, 2021, https://data.europa.eu /doi/10.2777/202437 and European Commission, Directorate-General for Research and Innovation, Commission Implementing Decision C (2022)2975, Horizon Europe Work Programme 2021–2022. Widening participation and strengthening the European Research Area, 10 May 2022. https://eur-lex.europa.eu/resource.html?uri=cellar:c1f95e49-d11b-11ec-a95f-01aa75ed71a1.0001.02/DOC_12&format=PDF, accessed on 30 June 2022) in the area of Research and Development, aiming specifically at promoting the transition to a new research framework for scholarly communication and the adoption of new innovative practices including OS.

The local and regional infrastructure and access to it have been improved through the effort to prepare services in the NI4OS-Europe partner countries for onboarding to the EOSC Catalogue of Services and Marketplace. As an intermediate step in this process, the NI4OS-Europe Service Catalogue (https://catalogue.ni4os.eu/, accessed on 30 August 2022) has been established [24]. It provides information about selected repositories, thematic, generic and core services in the partner countries, as well as about service policies. User manuals and training materials are provided, as well as a helpdesk. The services are monitored, and the monitoring information is publicly available [25]. A set of tools have been developed to support FAIR and Open Research Data Management and inclusion to EOSC: the Licence Clearance Tool (LCT) to automate license clearance; the EOSC Rules of Participation Tool (RoLECT) specifically addressing legal and ethical compliance for EOSC; the Repository Policy Generator (RePol), a tool for drafting repository and privacy policies. The project has created robust guidelines and a network of experts to support the process of establishing national service catalogues. The visibility of local and regional services has improved significantly. At the same time, the Catalogue makes it easier for research communities to find reliable tools and services.

Progress has also been made in terms of the limited availability of information on OS policy and activities across the region, through the launch of new informative resources and the creation of new materials. All interested parties may consult stakeholder registries, such as the Stakeholder Map in NI4OS-Europe website to trace local and regional actors. In addition, they may access service registries—among them the NI4OS-Europe Service Catalogue to find more about services in the region and their policies. Moreover, they now have at their disposal policy registries so as to become aware of what applies in each country together with information about infrastructure. The latter are hosted mainly in the NOSCIs' pages and the emerging NOSCIs' portals.

At the same time, training for researchers and research support staff is organized, specialized material and skills resources are produced, covering OS practices in several aspects of scholarly communication. The training materials created during the project are available on the NI4OS-Europe Training Platform [23]. This action increases consciousness regarding the potential of OS and promotes the use of the available resources for the benefit of research and scientific knowledge in the countries and worldwide.

Finally, it is worth mentioning that there is an overall approach to overcome limitations due to language diversity, by providing informative and training material not only in English but also translated in local languages of the partner countries.

The NOSCIs' approach has been very successful and already 10 NOSCIs have been established while 5 more are on the course to be established. The concept of NOSCI is

flexible and it allows for various governance models and stakeholder involvement. In the NI4OS-Europe countries where NOSCIs are established, they already play a prominent role in the facilitation of the EOSC governance and also as enablers of EOSC inclusion and Open Science widening at the national level. They are thus directly involved or are even leading scholarly communication activities in their countries. Relevant examples include the contributions to the drafting of national OS plans, the successful implementation of institutional OS strategies, the organisation of dissemination events on OS, the delivery of training events and material in relation to Open Research Data Management (ORDM).

6. Conclusions

The NI4OS-Europe project supports the development of OS and FAIR policies in 15 Member States and Associated Countries as well as the development and inclusion of the national Open Science Cloud Initiatives (NOSCIs) in the overall scheme of EOSC governance. By doing so, it increases engagement within a trusted, federated environment that allows researchers to search, reuse and publish data and services and builds OS capacity in the region, contributing directly to fundamental priorities for scholarly communities.

To amplify the reliability of the approach, we considered it important to provide an evaluation methodology for the process of establishing, operating and monitoring the results of the NOSCIs. The NOSCI indicators have been re-examined in relation to their suitability in the scholarly communication context, and a set of relevant metrics has been derived, along with the challenges to which they respond.

Considering the higher degree of complexity in our region (political, historical, OS policies) we believe that the set of indicators for scholarly communication in relation to EOSC/OS activities has a high reproducibility potential. The solutions offered by the NI4OS-Europe project are flexible and adaptable. In this respect, this analysis may be useful not only to Open Science stakeholders in Europe but also in other parts of the world, particularly in developing countries.

Author Contributions: Conceptualization, M.Š. and E.T.; methodology, M.Š. and E.T.; validation, E.S., K.K. and E.T.; formal analysis, E.S., K.K. and M.Š.; investigation, B.K., E.P., E.S., K.K., I.P. and K.L.; data curation, E.S. and K.K.; writing—original draft preparation, M.Š., E.T., K.L. and I.P.; writing—review and editing, E.S., K.K. and E.T.; visualization, M.Š., I.P., E.S. and K.K.; supervision, E.T.; project administration, E.T.; funding acquisition, E.T. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: No original, raw data has been generated for this manuscript. The used data sources are listed under Data and Methodology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Toli, E.; Papadopoulou, E.; Liatas, C.; Sifakaki, E.; Papastamatiou, I.; Prnjat, O. *NI4OS-Europe National OSC Initiatives Models*; Zenodo: Meyrin, Switzerland, 2020. [CrossRef]
2. Graham, T. Scholarly Communication. *Serials* **2000**, *13*, 3–11. [CrossRef]
3. Kramer, B.; Bosman, J. 101 Innovations in Scholarly Communication—The Changing Research Workflow, figshare. 2015. Available online: http://eprints.lse.ac.uk/70920/ (accessed on 5 July 2022).
4. Das, A.K. *Scholarly Communications*; Unesco: Paris, France, 2015; ISBN 978-92-3-100078-2.
5. Shearer, K.; Chan, L.; Kuchma, I.; Mounier, P. *Fostering Bibliodiversity in Scholarly Communications: A Call for Action*; Zenodo: Meyrin, Switzerland, 2020.
6. Sanchez, P.J.-A. *EOSC Portal Profiles v4.00*; Zenodo: Meyrin, Switzerland, 2021. [CrossRef]
7. *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)*; European Commission; Directorate General for Research and Innovation: Brussels, Belgium, 2022; ISBN 978-92-76-28941-8.

8. Macan, B.; Stojanovski, J.; Vodopijevec, A. *NI4OS-Europe Incentives for Supporting ORDM and FAIR*; Zenodo: Meyrin, Switzerland, 2020. [CrossRef]
9. Kosanović, B.; Ševkušić, M.; Otašević, V. *Open Science Stakeholders in Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Georgia, Greece, Hungary, Moldova, Montenegro, North Macedonia, Romania, Serbia and Slovenia [Dataset]*; Zenodo: Meyrin, Switzerland, 2020. [CrossRef]
10. European Commission. *Directorate General for Research and Innovation*; Country Sheets Analysis: Report from the EOSC Executive Board Working Group (WG) Landscape; Publications Office: Luxembourg, 2020; ISBN 978-92-76-25379-2.
11. Kosanović, B.; Ševkušić, M. *NI4OS-Europe Stakeholder Map, Inventory, Policy Matrix*; Zenodo: Meyrin, Switzerland, 2019. [CrossRef]
12. Arvola, M.; Beckmann, V.; Budroni, P.; Castelli, D.; Cavalli, V.; Davidson, J.; Di Giorgio, S.; Dovey, M.; Fischer, L.; Hönegger, L.; et al. *Second Working Proposal for Living Indicators to Monitor MS Progresses Towards EOSC Readiness*; Zenodo: Meyrin, Switzerland, 2021.
13. Toli, E.; Papadopoulou, E.; Liatas, C.; Sifakaki, E.; Papastamatiou, I.; Koumantaros, K. *NI4OS-Europe Indicators for the NOSCI's Establishment Checklist*; Zenodo: Meyrin, Switzerland, 2021. [CrossRef]
14. Forlenza, D.; Tregua, A.; Tanlongo, F.; Kosanović, B.; Toli, E.; Cavalli, V.; Matyska, L.; Hammargren, P.-O. *EOSC: Landscaping Analysis*; Zenodo: Meyrin, Switzerland, 2021.
15. Ramet, S.P. *Central and Southeast European Politics since 1989*; Cambridge University Press: Cambridge, UK, 2010; ISBN 978-1-139-48750-4.
16. Ivanović, D.; Jovanović, M.; Fritsche, F. Analysis of Scientific Productivity and Cooperation in the Republics of Former Yugoslavia before, during and after the Yugoslav Wars. *Scientometrics* **2016**, *107*, 499–519. [CrossRef]
17. World Bank Country and Lending Groups—World Bank Data Help Desk. Available online: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (accessed on 14 August 2022).
18. Research and Development Expenditure (% of GDP). Available online: https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS (accessed on 13 August 2022).
19. R&I Country Profile—Participation in Programmes. Available online: https://webgate.ec.europa.eu/dashboard/sense/app/a976d168-2023-41d8-acec-e77640154726/sheet/d23bba31-e385-4cc0-975e-a67059972142/state/analysis (accessed on 14 August 2022).
20. Laakso, M.; Multas, A.-M. *European Scholarly Journals from Small- and Mid-Size Publishers in Times of Open Access: Mapping Journals and Public Funding Mechanisms*; Zenodo: Meyrin, Switzerland, 2022.
21. Jakimovski, B.; Mishev, A.; Athenodorou, A.; Klemenčič, B.; Vudragovic, D.; Koumantaros, K.; Radonjić, M.; Prnjat, O. *NI4OS-Europe Provider Landscape Analysis and Provider Categorization*; Zenodo: Meyrin, Switzerland, 2020. [CrossRef]
22. Celjak, D.; Bekić, Z.; Cundeković, M.; Jertec, L.; Milinović, M.; Zubić, A. DABAR—The National Infrastructure for Digital Repositories. In Proceedings of the EUNIS 23rd Annual Congress: Shaping the Digital Future of Universities: Book of Proceedings, Münster, Germany, 7–9 June 2017; pp. 16–24.
23. Filiposka, S.; Athenodorou, A. *NI4OS-Europe Training Plan*; Zenodo: Meyrin, Switzerland, 2020. [CrossRef]
24. Kosanović, B.; Ševkušić, M. *NI4OS-Europe Stakeholder Map, Inventory, Policy Matrix Update*; Zenodo: Meyrin, Switzerland, 2022. [CrossRef]
25. Mishev, A.; Yordanov, S.; Vudragovic, D.; Imamagic, E.; Zamani, T.; Prnjat, O.; Radonjic, M. *NI4OS-Europe Best Practices for on-Boarding and Related Policies 2nd Version*; Zenodo: Meyrin, Switzerland, 2021. [CrossRef]

Article

Challenges of Promoting Open Science within the NI4OS-Europe Project in Hungary

Ákos Lencsés [1,2,*] and Péter Sütő [3,*]

1 Governmental Agency for IT Development, 1134 Budapest, Hungary
2 Department of Cultural Studies, Kodolányi János University, 1139 Budapest, Hungary
3 University Library, University of Public Service, 1083 Budapest, Hungary
* Correspondence: lencses.akos@kifu.gov.hu (Á.L.); suto.peter@uni-nke.hu (P.S.)

Abstract: National Initiatives for Open Science in Europe (NI4OS-Europe) is a Horizon 2020 project related to the European Open Science Cloud (EOSC). One of the project objectives is promoting EOSC and open science in 15 Central and East European EU states and EU-associated countries. This paper describes the variety of promoting activities carried out in Hungary as part of the NI4OS-Europe project by the Governmental Agency for IT Development (KIFÜ). Identifying good practices will give us the chance to find the best communication channels and methods to promote open science and to manage expectations of funders, researchers and librarians. The audience diversity of organized NI4OS events was analyzed in this study. The anonymized dataset based on registration forms was filtered by profession. Results suggest that events are generally visited by more librarians than researchers. The only exception is the third forum where the main Hungarian research fund as co-organizer might have attracted researchers' attention. This suggests that librarians are considered to be in charge of open science issues in general. Usage data of the open science news feed were also studied. The 130 posts between May 2021 and April 2022 and 2500 visitors until the end of June 2022 give us the chance to learn about the characteristics of the most visited posts. We can conclude that the focus of communication is on open and FAIR data management, while other areas receive less attention. The results show that despite more international posts being published, the target group is more interested in local information.

Keywords: Hungary; NI4OS-Europe; EOSC; open science; science communication

1. Introduction

National Initiatives for Open Science in Europe (NI4OS-Europe) is a Horizon 2020 project related to the European Open Science Cloud (EOSC) that runs between 1 September 2019 and 28 February 2023. One of the project objectives is promoting EOSC and open science in 15 Central and East European EU states and EU-associated countries. In the case of Hungary, two actively cooperating institutions take part in the consortium. The University of Debrecen University and National Library (DEENK) is the central library of one of the largest higher education institutions of the country. With a history reaching back to 500 years, 14 faculties and 30,000 FTE, University of Debrecen is among the top universities in Hungary. The other consortium member, the Governmental Agency for IT Development (KIFÜ) is the Hungarian national research and education network (NREN) provider. KIFÜ serves digitalization in Hungary, having 6400 customers and 2.5 million users; it offers a wide range of IT services for research and higher education.

This paper describes the variety of promoting activities carried out in Hungary as part of the NI4OS-Europe project by the Governmental Agency for IT Development (KIFÜ) in 2021 and H1 2022. An overview of these activities and identifying good practices will give us the chance to find the best communication channels and methods to promote open science and to manage expectations of funders, researchers and librarians in Hungary.

2. Materials and Methods

EOSC, starting its operation after many years of discussions in 2018, was originally aimed at managing European research data via European infrastructure [1]. Today, EOSC promotes not only FAIR principles but all other aspects of Open Science and has become a major player of the new, open research paradigm in Europe.

Several papers already highlighted the gap between the consensus of the importance of open research culture and the daily routine of researchers [2–5]. However, a relatively small amount of literature is available on best practices in promoting open science among researchers or analyzing best drivers that could help researchers embrace open science.

A detailed paper has been published recently by Robson et al. [6] that discusses mostly the psychological aspects of promoting open science. This publication analyzes the needs of different target groups (e.g., stakeholder groups, individual researchers) and the approaches that help us to understand what various practices are needed to reach out to these groups. In this paper, we also attempt to group NI4OS-Europe promoting activities by target groups.

Examining the conditions in Hungary, the engagement to open science goes back to the Budapest Open Access Initiative in 2001. In the next 1.5 decades, development of green open access was given priority, which led to the appearance of about 40 institutional repositories, some institutional OA policies and the creation of the Hungarian Open Repositories consortium (HUNOR). During this period, the open science communication focused on promoting green open access and on creating the related infrastructures and strategies [7–9]. Unsurprisingly, libraries focus mainly on infrastructure and databases/repositories as they are among the main focuses of the library sector in Hungary [10].

Significant progress was made between 2018 and 2020, when the Hungarian Electronic Information Service National Program (EISZ) concluded transformative open access agreements with major publishers [11]. As a result, communication related to open science also shifted to the promotion of gold open access publishing. At the same time, the broader interpretation of open science appeared first through international conferences such as Focus on Open Science, Budapest series [12], and later an increasing number of local events were dedicated to this topic as well.

In the period between 2019 and 2022, the open science focus on the strategic level shifted towards open and FAIR data management in Hungary. It seems as though we are at the beginning of a similar journey, on which we have already made significant progress in relation to green open access. The Hungarian research community focuses on the infrastructure, building the first data repositories, analyzing the arguments, creating strategies and trying to reach out and engage the researchers [13].

We can state that open science is receiving increasing attention from researchers, information specialists, strategy- and decision-makers and the public in Hungary. The National Position Paper on Open Science was signed and developed by the National Research, Development and Innovation Office (NKFIH) and other research related institutions and organizations in 2021. As we see, during these two decades, several papers discussed the current position and the benefits of open science on a strategic level as well as the difficulties of engaging the researchers in practice. Still, we can find few studies on best practices of promoting open science in the Hungarian research environment. These papers [14,15] support our assumption that researchers' achievement and commitment to open science is only possible with well-planned and well-prepared communication, where the practical implementation with the local environment and for the local researcher community must be emphasized in order to achieve results.

We analyzed the audience diversity of four online Hungarian Open Science Forum events organized as part of the NI4OS-Europe project. The anonymized dataset based on registration forms was filtered by profession. We attempted to identify the main differences of events attracting mostly librarians and those where the majority of the audience were researchers.

An open science newsfeed was also introduced by KIFÜ as part of NI4OS-Europe. The growing interest toward this newsfeed gave us a chance to analyze usage data of the 133 most-visited posts published in the period between May 2021 and April 2022, with the number of visitors until 30 June 2022. This gives us the opportunity to learn about the characteristics of the most visited posts. Data on the newsfeed were collected from KIFÜ's newsfeed log using Matomo Analytics. The newsfeed was analyzed through simple descriptive statistics, and the possible significant differences among variables were examined to validate hypotheses that the views of the posts depend on the selected criteria. The possible significant differences among variables were examined through the independent *T*-test. Where the sample sizes and variances were unequal between the groups (Leven's test $p < 0.05$), we used Welsch's *t*-test. The statistical analyses were carried out by using the software SPSS 21.0. For deeper analysis, a post classification was processed along the following groups: open science in general, open access, open and FAIR data, open methods, citizen science, open science infrastructure and financing open science (see Figure 1). Due to the significant overlap amongst the open science fields in the content of newsfeed items, they cannot be used as independent variables. In addition, group size after distribution is too small in many cases. Therefore, the comparison of usage data for these groups using mathematical statistical methods cannot provide reliable results.
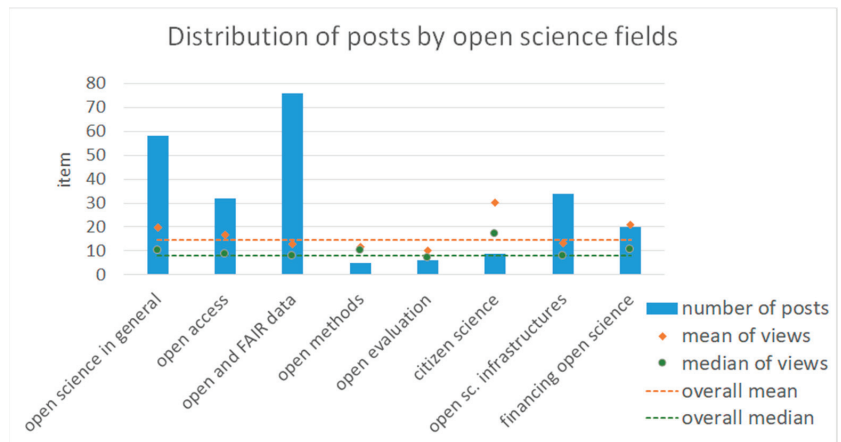
Figure 1. Distribution of posts by open science fields.

Due to the very small sample size, all the results of these studies need to be handled cautiously. Possible bias both in online traffic and event participations cannot be entirely precluded.

3. Results

As described above, one of the main objectives of the NI4OS-Europe project is to promote open science in Central and East European countries. The original concept counted on real-life activities, such as seminars and other on-site events. The rapidly spreading COVID pandemic and the restrictions it has caused made it impossible to conduct on-site events in 2021 and H1 2022, and all promoting efforts had to focus on online activities. This caused major changes compared to the original plans; however, being online only might have helped to reach out to a wider audience. Having all activities online also made it easier to archive all the materials and lectures and made them available for any possible later re-use.

All the promotion activities carried out as part of the NI4OS-Europe project were aimed at different, sometimes overlapping groups in Hungary, e.g.,

- publishing researcher interviews on research data management and open science practices for early career researchers;[1]
- open science news feed for open science practitioners;[2]
- publishing an e-learning course on EOSC and open science for graduate and PhD students;[3]
- testing RDM tools for the Hungarian research community;
- organizing various events, including the Hungarian Open Science Forum targeting senior researchers and stakeholders;
- an EOSC Champion program at three major Hungarian universities.

3.1. Researcher Interviews

The idea of publishing interviews with prominent researchers is considered as a bottom-up approach. Having no pressure or expectations, researchers could freely talk about their experience and practices regarding research data management (RDM), also giving voice to their concerns regarding open science. Interviewees were chosen from the widest possible range of research areas to show that RDM cannot be narrowed only to science, technology, engineering and mathematics (STEM) fields. All the interviews were published on Videotorium, which is the main Hungarian online video sharing platform of research and educational videos run by KIFÜ. A channel 'EOSC and Open Science' was launched to accommodate all the interviews, freely available to all.

Altogether, seven videos were published, and 396 views were recorded until 30 June 2022 (see Table 1).

Table 1. Video interviews published as part of the Hungarian NI4OS-Europe activity until 30 June 2022.

Interviewee	ORCID	Research Field	Interview Publication Date	Number of Views until 30 June 2022
Zoltán Kmetty	0000-0002-6775-8938	sociology	17 March 2022	94
Gábor Palkó	0000-0002-4394-8577	literary history	17 June 2021	80
Tamás Ferenci	0000-0001-6791-3080	biostatistics	17 June 2021	79
András Perczel	0000-0003-1252-6416	biochemistry	16 June 2021	70
Zoltán Kis and László Szentmiklósi (joint interview)	0000-0002-8365-8507; 0000-0001-7747-8545	physics and chemical engineering	16 March 2022	34
Miriam Szőcs	NA	art history	17 March 2022	26
György Eigner	0000-0001-8038-2210	system engineering	27 September 2021	13

It might be tempting to group views according to research fields. However, the relatively small sample size, both of videos and views, prevent us from rushing to any conclusion. It might be worth considering whether alternative video sharing platforms in addition to Videotorium could help generate more views, especially noting that YouTube has become one of the largest search engines in the world [16].

3.2. Open Science Newsfeed

The open science newsfeed of KIFÜ was launched for the test phase in April 2021, and the live, daily–weekly updates were started later[4]. According to KIFÜ's role in open science in Hungary, this online newsfeed is an important source of information on open science for the Hungarian community. Therefore, analyzing the KIFÜ's newsfeed might show us what kind of open science information the Hungarian open science representatives are trying to convey to the community. On the other hand, we can see how this attempt meets the audience's interest.

The newsfeed informed researchers and stakeholders about international and local open science trends and events. Not all the posts were related directly to NI4OS-Europe; rather, the aim was the widest possible range of information. Most of the events promoted via the newsfeed were organized by different European associations or institutions. The main focus was on NI4OS-Europe and EOSC-related news, while many posts called the attention to press releases, policy papers, research articles and other publications regarding all aspects of open science.

The analysis covers the 133 most-visited posts that were published in the open science newsfeed of KIFÜ between 5 May 2021 and 7 April 2022. Online traffic until 30 June 2022 was recorded by Matomo Analytics. We collected the number of individual views of each item, filtering out multiple viewings by the same users. Then, we categorized the posts according to different aspects such as type, topic focus (international or local focus) and open science field. During the analysis, the distribution of items by category can be considered as a representation of the communication goals, while the number of views is a representation of the needs of the target group. Considering the shortcoming that the target group could only choose from what was presented in the newsfeed, analyses were carried out through simple descriptive statistics and examining possible significant differences among variables [17] (Falus and Ollé, 2008) by using the software SPSS 21.0.

For the first analysis, posts were grouped into two categories: 'event' and 'other'. Posts advertising some event (invitation to a future event or report of a past event) were categorized as 'event'. 'Other' posts discussed press releases, open science trends, EOSC news, reports, etc. Data show that 81 posts (61% of all posts analyzed) advertised an event, while 52 posts (39%) were categorized as 'other'. On one hand, this rate is a facility, and on the other hand, it can be a conscious communication strategy to achieve a more lasting impact and a greater commitment through an event. On the other hand, usage statistics show a slightly different ratio: while events gained 71% of total page views, other posts gained 29% only (see Figure 2). Carrying out an independent-samples T–test, we found that the mean of views for event and non-event type ($p = 0.07$, t = -1.826, F-test $p = 0.083$) are not significantly different.

Second analysis also grouped posts into two categories based on 'internationality'. Posts related to reports, press releases, events, etc. of institutions outside Hungary were labeled as 'international'. Events organized by Hungarian institutions, and posts related to Hungarian trends were labeled as 'Hungary'. The internationality of the items in the newsfeed shows that 98 posts (74%) were international, and only 35 of them (26%) presented local content. This means that the majority of information on open science is still coming from abroad. However, if we compare the views of the posts in distribution of international and local content, we see an opposite result: the local content was visited more often (See Figure 3). The independent-samples *T*-test with Welsch's test also validates that the mean number of views for international posts and for local posts is significantly different (Welsch's test: $p < 0.001$, t = 4.050; F-test $p < 0.001$). The average views for local posts were 20.473 more than the average views for international posts. This result shows that despite the fact that more international posts are available, the target group is more interested in local information.
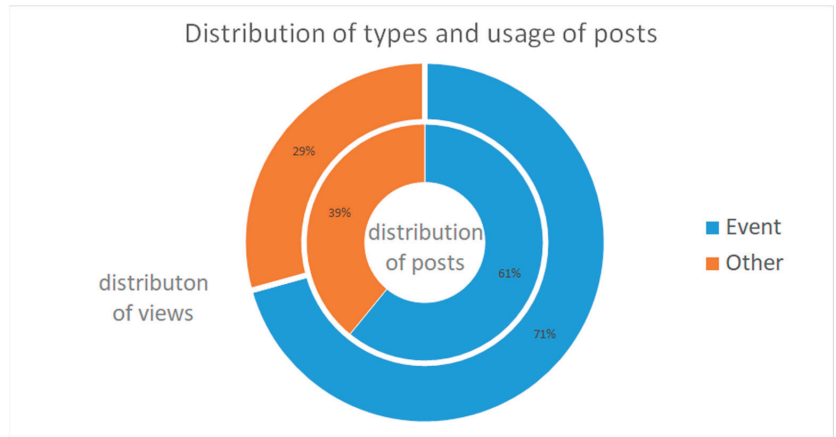
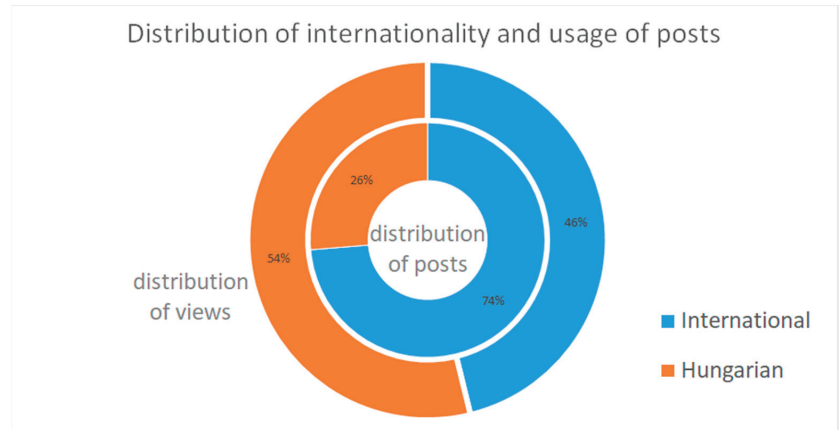**Figure 2.** Distribution of types and usage of posts.

Figure 3. Distribution of internationality and usage of posts.

Since the events had greater presence in the posts, we decided to analyze it in more detail. The distribution of the 81 events related posts by internationality shows a very similar picture to what we saw when examining the internationality of the entire sample (see Figure 4). The presence of international events was much higher amongst the posts: 58 (72%) were labeled as 'international', while 23 of them (28%) were Hungarian. The usage statistics of posts was, however, lower compared to the local events: international events gained only 44%, while Hungarian events gained 56% of all page views. The Welsch's test validates that the mean number of views for posts on international events and for posts on local events is significantly different (Welsch's test: $p = 0.003$, t = 3.323; F-test $p < 0.001$). The average views for posts on local events was 22.629 more than the average views for posts on international events. This result shows that despite the fact that more international events are available on open science, the target group is more interested in local events. This is particularly an interesting result in light of the fact that, due to the pandemic situation, events were mostly organized online, so participation in international open science events did not involve more efforts than participation in local events. Various reasons can stay behind this phenomenon. One of these factors could be the language barrier, when a librarian or a researcher feels no ease at joining an English language event. While language barrier can be easily one of the main drivers of the unequal distribution

of local and international event post visitor numbers (see Figure 4), this clearly cannot be the only reason why posts on local trends gain more visitors than international ones (see Figure 3).
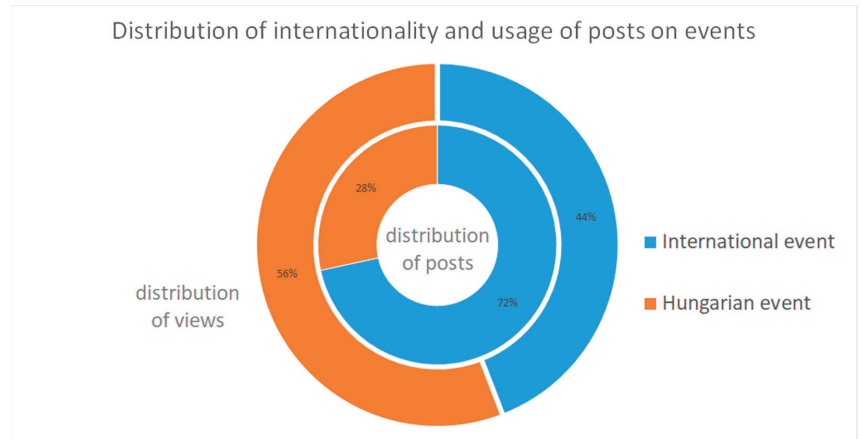

Figure 4. Distribution of internationality and usage of posts on events.

Grouping the newsfeed items by open science fields adequately shows the recent focus of the open science representatives in Hungary (see Figure 1). The hottest topic during this period was the open and FAIR data, as they were related to 57% percent of the newsfeed items, while the second and third specific fields (open science infrastructures and open access) were related to 25% and 24% only. Looking to the end of the list, citizen science (7%), open evaluation (5%) and open methods (4%) seem less popular or less important in the communication and open science promotion in Hungary.

Using descriptive statistics to analyze the covered open science fields, we see that the means of views as well as the medians of views are quite close to the overall, except for field citizen science (see Figure 1). This means that the interest of the newsfeed users did not differ from the received content distribution. The only exception was the citizen-science-related posts where users showed more activity. We can conclude that the focus of communication is on open and FAIR data management, while other areas receive less attention. This result can help in shaping the open science communication strategy in Hungary.

3.3. E-Learning Course on EOSC

Another open science promotion activity was the development of a Moodle-based e-learning course 'Open Science and EOSC in practice' by the KIFÜ team on H1 2021. The course was launched early July 2021, while an online workshop promoting the course and discussing possible developments took place on 13 July 2021. The course began with four modules, while an additional fifth module was added in January 2022. As the main target was student groups, the courses were launched in Hungarian only, and no prior knowledge was required. Modules are richly illustrated with short videos, diagrams, quizzes, and other interactive tools to make it easy to integrate them with any higher education course.

The five modules of the course can be handled individually, discussing different aspects of EOSC and open science (see Table 2).

Table 2. Modules of the e-learning course 'Open Science and EOSC in practice'.

Module	Title	Key Topics
Module 1	A practical guide on how open science can support researchers	Replication crisis, OS tools, OS community building, OS requirements of Horizon Europe calls
Module 2	FAIR data management	FAIR principles, DMP tools, financial aspects of data management
Module 3	Changes driven by EOSC	EOSC development phases, researcher contribution to EOSC, EOSC working groups
Module 4	EOSC services	EOSC portal and marketplace, EOSC on-boarding process
Module 5	Trends of the research systems	OS metrics, research evaluation

We collected feedback on the testing of the educational module developed for teaching open science, for which we used a workshop as the most effective method. This direct communication ensured the correct and accurate interpretation of the information received from the testers. The modules were tested in advance by 40 people, mostly librarians, of 15 research and higher education institutions. Results of the tests and views of the people involved in the testing were discussed during the workshop organized on 13 July 2021. The feedback praised the interactivity of the course and the rich illustration materials. One attendee also added that 'it would be awesome if science could work in this [Open Science] way'.

It also became clear during the workshop that integrating the course with the university curricula had certain barriers. Open science usually is not considered as individual courses, and some of the Hungarian universities lack even obligatory general courses regarding research support or research methodology. Open science and research methodology are considered often as 'library businesses', while librarians have no full-semester courses and are invited to contribute existing courses only occasionally.

Based on these responses, the e-learning course was opened and recommended for the EOSC Champion program to gain more feedback.

3.4. EOSC Champion Program

The EOSC Champion program was devised as a series of nine events at major Hungarian universities. While mentoring is considered to be more effective in a multi-year connection [18], these one-year-long champion programs might also have a positive effect on the early career researchers' career path by showing an alternative research methodology.

To run this program, cooperation was built between KIFÜ and university professors who promoted EOSC and open science among their fellow researchers and PhD students. Altogether, three universities took part in the program: Eötvös Loránd University, Óbuda University and University of Szeged.

About one third of the program focused on research data management issues, while two events specifically discussed EOSC as follows:

1. Introduction
2. Open Science
3. Scientific Publications
4. Research Data Life Cycle and Management
5. Data Management Plan
6. FAIR Principles
7. EOSC—overview
8. EOSC—services
9. Future Trends in Scientific Research Systems and Careers

These were held as a monthly series, where KIFÜ provided help for the professors for each event, including PowerPoint slides, a list of questions possible to drive open discussions and further publications on the recent development of the certain topic. Events were held mostly as in-person seminars, occasionally changed to online. Monthly project meetings were also conducted for the champions to share experiences and discuss oncoming topics.

Though systematic surveying of the audience did not take place, all three EOSC champions shared their thoughts and feelings during the monthly meetings about this program. The overall conclusion of the EOSC Champion program was that PhD students are more likely to take part in the open science discussion (even though this series was not officially part of the PhD curriculum), while professors are much harder to involve in such activity. Taking part in any kind of open science activities is barely acknowledged in the research evaluation process of the universities, and this is not easily overcome for individual researchers. While most of the attendees agreed on the global benefit of open science, much feedback was gained about the lack of the financial drivers regarding open research practices.

3.5. Testing RDM Tools

The issue of data repositories has found its way to the hot topics of research policy in recent years in Hungary, and most of the universities and research institutions still provide no such service for their researchers. A cooperation was formed between KIFÜ and the National Laboratory for Digital Heritage (DH-LAB) that serves as a major center for digital humanities in Hungary. The DH-LAB was officially founded in late 2020, while earlier works were also carried out at the same institution [19]. With support from DH-LAB expertise, Invenio was chosen to build repository structure that is freely adaptable to institutional requirements, while technical support is provided by KIFÜ. Invenio is still in beta, which makes it possible to easily shape the required features of the future repository, while being a CERN software makes it safe enough to build on. This work started in H1 2022, and results will be published only at the end of the project. Parallel with and independent of this initiative, other data repository projects commenced in 2022, most importantly the Eötvös Loránd Research Network (ELKH) ARP Data Repository project[5] that aims to provide a data repository covering researchers of the largest publicly funded Hungarian research network of 11 research centers, 7 research institutes and 116 additional supported research groups.

These initiatives are still under development, so we have no possibility to judge any of the outcomes. The many independently started RDM projects, however, make clear the awareness of research data, FAIR principles, and open science criteria both from researchers and research funds sides.

3.6. Hungarian Open Science Forum

The Hungarian Open Science Forum was launched as part of the NI4OS-Europe project by the two Hungarian consortium partners: DEENK and KIFÜ. The forum is organized as an online event. Skiles et al. [20] discusses all the benefits of online events, including costs and wider geographical composition of attendees. These factors were clearly a benefit of the online format, while organizers had to be more conductive to generate discussion.

The main objective is to inform the Hungarian researcher community about recent open science, especially EOSC-related trends. Another important aspect of the forum was to introduce and gain feedback regarding the then-forming National Position Paper on Open Science.

A forum event is usually 90–120 min long, where some presentations are held, and online discussion is formed based on the presentations. For H1 2022, four forum events were organized: The first forum took place on 28 May 2021, followed by the next ones on 24 September 2021; 19 January 2022 and 28 April 2022. Topics of the forum events varied from discussing open science policies of European countries, introducing the National

Position Paper on Open Science and presenting open science practices for life scientists and social scientists (see Table 3).

Table 3. Number of attendees of the Hungarian Open Science Forum events.

Event	Main Topic	Date	Number of Attendees
Hungarian Open Science Forum I	Introducing open science and EOSC	28 May 2021	53
Hungarian Open Science Forum II	Open science policies of European countries	24 September 2021	36
Hungarian Open Science Forum III	National Position Paper on Open Science	19 January 2022	138
Hungarian Open Science Forum IV	Open science practices for life scientists and social scientists	28 April 2022	74

All the Hungarian research institutions and higher education institutions were informed about the forum events directly via e-mail. A total of 150–270 e-mails were sent to promote each event, while social media communication also supported the recruitment process. The number of attendees varied between 36 and 138 (see Table 3). For the analysis, all data of registrants who had not attended the meeting were removed from the dataset.

The third forum had the highest number of attendees. This event was organized together with the National Research, Development and Innovation Office, that is the main research fund body of Hungary. At this event, the vice president for science and international affairs and the open science advisor of the office introduced the newly launched National Position Paper on Open Science. It seems clear that the National Research, Development and Innovation Office attracted more attendees than other forums.

Due to the registration form, we were able to analyze the attendee affiliations and professions (see Table 4). For this, three groups were formed: researchers, librarians (meaning all library staff, including IT specialists and data stewards) and organizers (all KIFÜ and DEENK staff). Where the 'profession' field was left blank during registration, affiliation and e-mail fields helped to determine the most suitable group for the attendee type.

Table 4. Number of attendees of the Hungarian Open Science Forum events broken down by affiliation.

Event	Number of Attendees	Number of Researchers (Among Attendees)	Number of Librarians (Among Attendees)	Number of Organizers (Among Attendees)
Hungarian Open Science Forum I	53	23	24	6
Hungarian Open Science Forum II	36	7	19	10
Hungarian Open Science Forum III	138	82	45	11
Hungarian Open Science Forum IV	74	19	43	12

The forums are generally visited by more librarians than researchers. The only exception is the third forum where the National Research, Development and Innovation Office as co-organizer might have attracted researchers' attention. This effect, however, did not occur at other events. This suggests that open science is indeed considered as a 'library business'. This was reflected by heads of research, who answered the invitation letters, and delegated the librarian of the institution to the forum. The high ratio of librarian attendees of the

events also underlines the importance of librarians in promoting open science. Librarians play a large role in facilitating open science in their research institutions.

4. Discussion

This paper describes all the open science promotion activity carried out in Hungary by KIFÜ as part of the NI4OS-Europe project. Several activities were introduced, including online interviews with researchers, champion programs for early career researchers, an e-learning course, open science newsfeed, online events, etc. By learning the usage data and attendee ratio of 2021 and H1 2022 activities, we might have identified practices that proved to be more successful in Hungarian context compared to others. These data need to be handled cautiously due to small sample size; results are more likely only impressions of the first half of the project activities.

Studying usage and number of visitors, online video interviews and e-learning courses were less attractive in the period, noting that these activities are not meant to be used one-time only. It is clear, however, that additional promotion is needed to reach the target audience, especially via social media and YouTube. The overall impression of an EOSC Champion program was that younger researchers are easier to involve in open science discussions. The greatest barrier seems to be that open science activities are barely acknowledged in the research evaluation process of the researchers.

Relatively high usage and a wide range of posts were recorded for the Hungarian NI4OS open science newsfeed. Analyzing the posts from different aspects we saw that events were overrepresented, and the feed had much more international than local context. The usage of posts confirmed the strategy of overweighting the events as the interest for these posts was greater than expected. The usage data of the posts revealed that although more international posts were available to the users, they still read the ones with local context more. The same results can be seen by analyzing the usage of posts for international and local events. Despite the fact that, due to the pandemic situation, the international events were also held online, the posts of domestic events received much more interest. This result may even indicate language barriers.

By examining which open science areas the posts focused on, we found that the topic of open and FAIR data management was given priority in the communication. The usage of the posts in relation of open science fields showed balanced attention from readers, confirming that the content provided in the newsfeed met the needs of the community. However, it is also important to take into account that the offered content influences the consumption of information, so in open science communication it may be worthwhile to give space to fields that are currently receiving less emphasis.

Analyzing the attendee ratio of four online Hungarian Open Science Forum events, it seems that open science is considered part of a librarian's duty, while researchers can be involved in higher numbers when research funds are also a factor. This suggests that bottom-up open science promotion activity needs to be accompanied by a top-down approach as well. Based on the results, the role of librarians is particularly important in facilitating open science, so emphasis must be placed on their training in this direction to provide the appropriate skills for knowledge transfer.

This study might be of great help in mapping the open science landscape of a Central European country, being the first to assess the promotion activity of an open science project in Hungary. Using multiple data sources for this purpose, we have the possibility to form conclusions of various aspects of the Hungarian open science landscape. This is one of the first data-based research studies that can point out that open science is clearly linked to libraries, and it is generally thought that open science is role for librarians. This finding might help shape the skills of library and information professionals in the future. The paper also shows that the language barrier can be measured in a Hungarian context. This underlines the importance of using national language(s) when promoting open science and science in general. While using English is a must when following international trends,

initiatives of organizing local events and translating statements, white papers cannot be underestimated.

To learn more about the results, it would be worthwhile to collect data from other NI4OS-Europe consortium members regarding their open science promotion activities. This would make it possible to compare usage patterns of different Central and South European countries and learn if the above results can be observed for other research communities.

Author Contributions: Conceptualization, Á.L.; methodology, Á.L. and P.S.; formal analysis, P.S.; investigation, Á.L. and P.S.; resources, Á.L. and P.S.; data curation, P.S.; writing—original draft preparation, Á.L.; writing—review and editing, P.S.; visualization, P.S.; supervision, Á.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Research data regarding this study are openly available via Zenodo. Lencsés, Á.; Sütő, P. Number of Attendees of Hungarian Open Science Forum Events, and Number of Visitors of KIFÜ Open Science Newsfeed, 2022. https://doi.org/10.5281/zenodo.7034816 (accessed on 30 August 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] The interviews are available (in Hungarian only) at https://videotorium.hu/hu/channels/4935 (accessed on 14 October 2022).

[2] The newsfeed originally was published under https://kifu.gov.hu/ni4os/hirek. During Summer 2022, the major redesigning of KIFÜ's web-page concluded in a new site under https://kifu.gov.hu/ni4os-hirek/. All earlier posts have been migrated to the new platform, and earlier URLs have been redirected to the new sites (accessed on 14 October 2022).

[3] Nyílt tudomány és EOSC a gyakorlatban. Available online: https://elearning.kifu.hu/ (accessed on 14 October 2022).

[4] For the newsfeed availability, see note 3.

[5] The acronym ARP refers to 'data repository project' in Hungarian: AdatRepozitórium Projekt.

References

1. Burgelman, J.-C. Politics and Open Science: How the European Open Science Cloud Became Reality (The Untold Story). *Data Intell.* **2021**, *3*, 5–19. [CrossRef]
2. Andreoli-Versbach, P.; Mueller-Langer, F. Open Access to Data: An Ideal Professed but Not Practised. *Res. Policy* **2014**, *43*, 1621–1633. [CrossRef]
3. Armeni, K.; Brinkman, L.; Carlsson, R.; Eerland, A.; Fijten, R.; Fondberg, R.; Heininga, V.E.; Heunis, S.; Koh, W.Q.; Masselink, M.; et al. Towards Wide-Scale Adoption of Open Science Practices: The Role of Open Science Communities. *Science and Public Policy* **2021**, *48*, 605–611. [CrossRef]
4. Dominik, M.; Nzweundji, J.G.; Ahmed, N.; Carnicelli, S.; Mat Jalaluddin, N.S.; Fernandez Rivas, D.; Narita, V.; Enany, S.; Rios Rojas, C. Open Science—For Whom? *Data Sci. J.* **2022**, *21*, 1. [CrossRef]
5. Nosek, B.A.; Alter, G.; Banks, G.C.; Borsboom, D.; Bowman, S.D.; Breckler, S.J.; Buck, S.; Chambers, C.D.; Chin, G.; Christensen, G.; et al. Promoting an Open Research Culture. *Science* **2015**, *348*, 1422–1425. [CrossRef] [PubMed]
6. Robson, S.G.; Baum, M.A.; Beaudry, J.L.; Beitner, J.; Brohmer, H.; Chin, J.M.; Jasko, K.; Kouros, C.D.; Laukkonen, R.E.; Moreau, D.; et al. Promoting Open Science: A Holistic Approach to Changing Behaviour. *Collabra Psychol.* **2021**, *7*, 30137. [CrossRef]
7. Bilicsi, E. Towards Open Access: Supporting Open Access Publishing in Central European Countries. In *Information and Technology Transforming Lives: Connection, Interaction, Innovation Proceedings*; Faculty of Humanities and Social Sciences, University of Osijek: Osijek, Croatia, 2019; pp. 361–368.
8. Holl, A. Az MTA KIK gyakorlata a nyílt tudomány támogatásában. In *Open Science: Nyílt Tudomány Magyar Szemmel*; Gaálné Kalydy, D., Ed.; Magyar Tudományos Akadémia Könyvtár és Információs Központ: Budapest, Hungary, 2021; pp. 161–173. [CrossRef]
9. Karácsony, G.; Görögh, E. New Chapter of Knowledge Management in Hungary: Steps toward Open Science and Tertiary Collaborative Work. *HERJ Hung. Educ. Res. J.* **2017**, *1*, 94–102. [CrossRef]
10. Áts, J.; Deák, N.; Kiszl, P.; Varga, K. Development of Libraries in Hungary. In *Libraries in the Early 21st Century*; Sharma, R., Ed.; De Gruyter Verlag: Berlin, Germany, 2012; pp. 183–202. [CrossRef]
11. Sütő, P. Az EISZ open access szerződéseinek gyakorlati tapasztalatai. In *Networkshop 2020. Országos Online Konferencia. 2020. Szeptember 2–4.*; Tick, J., Kokas, K., Holl, A., Eds.; Hungarnet Egyesület: Budapest, Hungary, 2020; pp. 64–72. [CrossRef]

12. Ignat, T. What Motivates Us to Develop the Focus on Open Science Series? In *Open Science: Nyílt Tudomány Magyar Szemmel*; Gaálné Kalydy, D., Ed.; Magyar Tudományos Akadémia Könyvtár és Információs Központ: Budapest, Hungary, 2021; pp. 127–144. [CrossRef]
13. Lencsés, Á. Kutatási adatok könyvtári kezelése. *Tudományos Műszaki Tájékoztatás* **2021**, *68*, 663–670. [CrossRef]
14. Hoczopán, S.; Molnár, T. Az egyetemi könyvtárak szerepe a nyílt tudomány mozgalom elterjesztésében. In *Open Science: Nyílt Tudomány Magyar Szemmel*; Gaálné Kalydy, D., Ed.; Magyar Tudományos Akadémia Könyvtár és Információs Központ: Budapest, Hungary, 2021; pp. 145–160. [CrossRef]
15. Zeller, R.; Hoczopán, S.; Nagy, G. Kutatási adatkezelést támogató szolgáltatások előkészítése a Szegedi Tudományegyetemen. *Tudományos Műszaki Tájékoztatás* **2021**, *68*, 576–586.
16. YouTube Users Statistics 2022 [Infographics] | GMI. Official GMI Blog. Available online: https://www.globalmediainsight.com/blog/youtube-users-statistics/ (accessed on 30 August 2022).
17. Falus, I.; Ollé, J. *Az empirikus kutatások gyakorlata: Adatfeldolgozás és statisztikai elemzés*; Nemzeti Tankönyvkiadó: Budapest, Hungary, 2008.
18. Diggs-Andrews, K.A.; Mayer, D.C.G.; Riggs, B. Introduction to Effective Mentorship for Early-Career Research Scientists. *BMC Proc.* **2021**, *15*, 7. [CrossRef] [PubMed]
19. Indig, B.; Knap, Á.; Sárközi-Lindner, Z.; Timári, M.; Palkó, G. The ELTE.DH Pilot Corpus—Creating a Handcrafted Gigaword Web Corpus with Metadata. In Proceedings of the 12th Web as Corpus Workshop, Marseille, France, 11–16 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 33–41.
20. Skiles, M.; Yang, E.; Reshef, O.; Muñoz, D.R.; Cintron, D.; Lind, M.L.; Rush, A.; Calleja, P.P.; Nerenberg, R.; Armani, A.; et al. Conference Demographics and Footprint Changed by Virtual Platforms. *Nat. Sustain.* **2022**, *5*, 149–156. [CrossRef]

MDPI

Article

Scholars' Domain of Information Space

Danijela Pongrac [1,*], Mihaela Banek Zorica [2] and Roman Domović [1]

[1] Department of Computer Science, Zagreb University of Applied Science,10000 Zagreb, Croatia
[2] Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb,10000 Zagreb, Croatia

* Correspondence: dpongrac@tvz.hr; Tel.: +385-99-2198-911

Abstract: This article addresses Croatian scholars' information behavior and how they use technology to acquire information in three areas of their work: teaching, research, and administrative activities. Our study aims to find which communication channels scholars utilize to find and share knowledge. Are they using communication channels targeting a broader audience, i.e., formal–explicit communication, or those targeting a narrower one, i.e., informal–implicit communication? The questionnaire used included four questions regarding scholar activities, with nine possible communication channels, scored on a seven-point Likert scale. Considering many channels for each area of activity, a reduction was made through Principal Component Analysis (PCA), to determine latent components in various channels. In finding information for teaching activities, the main communication channel is informal and implicit, while for research and administrative activities, it is formal and explicit. PCA shows a distinction between social and technical domains of science in terms of how scholars collect material for administrative tasks. A further communication channel is reduced to two factors for all questions, where the first factor has formal–explicit and the second has informal–implicit characteristics. This work is part of a larger study aimed at determining the mechanisms of information diffusion within academic institutions, utilizing the Information space model.

Keywords: I-space model; scholars; communication channels

1. Introduction

Given the development of modern technologies and the availability of various tools and modalities of communication, higher education institutions (HEI) can develop and improve ways to exchange information more effectively between their scholars and other stakeholders. Here the emphasis is on scholars and the dominant forms of channel communication from which they explore information for their three basic activities: teaching, research, and administration. Given that scholars have a constant need for information, it is necessary to check whether there are certain differences between different disciplines; in this case between the social and technical fields.

This paper seeks to discover the modalities of taking over and disseminating information through an institution; the way it is disseminated determines the strength of the diffusion of the information itself. In this sense, we are guided by the assumption of the Boisot Information space model (I-space) [1]: the larger the population to which information is directed, the weaker the diffusion, because information is not sufficiently widespread in space. The model also assumes that when the information is well coded and abstract, diffusion is a prerequisite because the explicitness of the content is achieved. However, on the other hand, if there is a large population, the information often does not achieve good enough diffusion.

Accordingly, we explore which communication channels scholars use when collecting information in the three basic activities of academic work. The I-space model divides information and knowledge from a non-codified and non-diffuse, i.e., a tacit and narrow area, to a codified and diffuse, i.e., explicit wider scope. Considering the framework

of I-space and the most common communication channels of information gathering, it is possible to determine the characteristics of individual communication channels, and, following the theory, to determine the prevailing communication channel for a particular activity of academic work. The focus of this research is on the most common forms of communication channel for the transmission of information in the context of HEI, which are divided into explicit and implicit forms of information diffusion given the framework of the I-space conceptual model.

For better understanding and conceptualization, we provide a short overview of the theoretical model of I-space with an emphasis on the dimension of knowledge and communication channels. Furthermore, the general communication features in HEI are presented, regarding formal and informal forms, with special emphasis on the specifics of the Community of Practice—which has the characteristics of both forms—as well as the vertical and horizontal directions of communication. In the conducted research, a descriptive analysis of the obtained results was made, divided into three activity areas for finding information, and one concerning sharing information, regarding the field of science that scholars belong to. Given the many components for each area of activity, a reduction of communication channels was made through Principal Component Analysis (PCA) to identify logical combinations of components and to give a better understanding of the interrelationships between them. Data reduction resulted in two factors within each activity and field of science, which were divided according to their characteristics into explicit–formal and implicit–informal communication channels.

2. Model of I-Space and Communication Forms in HEI

Boisot's I-space model is a three-dimensional entity that explains the forces that direct the flow and distribution of knowledge within a given space [1]. The three dimensions relate to codification, abstraction, and diffusion processes, which drive the flow of data and are considered crucial for information processing. Together, they form the three features of I-space, its conceptual framework, which can explore the behavior of information flow to understand the creation and dissemination of knowledge within selected populations. Codification and abstraction are more subjectively related because abstraction represents a cognitive strategy that reduces and optimizes content, while codification simplifies form. By researching the effects of forces that shape data flow patterns in different parts of I-space, they provide insight into how knowledge is gradually built in the individual's head, in written records and documents and also in organizations, and how long-term migration of knowledge from one part of I-space to another can occur [1]. As the authors of [2] emphasize, the I-Space model is an analytical tool for cultural and institutional analysis, and Boisot approached it uniquely, in terms of institutional analysis based on information. In other words, I-space is a tool for understanding different flows of different types of information, which helps understand the creation and dissemination of information within groups of people. Therefore, I-space at the individual level can also explain the construction of the domain of information from identification, comprehensibility and usability, to structuring and organising data that are part of personal information management, with different forms of communication channels intertwined in all these parts. By considering the systemic relationship between codification and diffusion, which has wide implications on psychological and sociological processes, reference [1] lists four dimensions of knowledge concerning different communication situations, the population in question and the availability of technology.

The first considers personal knowledge, which is often difficult to articulate and is most often communicated implicitly through examples, it is inaccessible because it is related to a particular context. Since there is no common context in personal knowledge, there is no common code, which is needed for transmission. Most implicit or tacit knowledge is uncodified and can be fully shared only with those directly present, which, except for in video conferencing, is usually a limited number, so it is also undiffused (insights, experience, face-to-face conversation). The second, proprietary knowledge, refers to structured

knowledge that is considered codified and un-diffused. According to [3], it is ready for transmission but is intentionally limited to a small population, and only those who know about the existence of this knowledge can access it (institutional cloud, intranet, closed database). However, if such knowledge proves to be useful, it has value and interest in further transmission. The next is public knowledge, which refers to knowledge that is structured, verified and recorded through different types of media, so it is codified. This type of knowledge is widely spread, diffused, and is most often unrelated to its origin In addition, it is mostly impersonal (libraries, open databases, social networks, wikis, institutional internet). The last is common-sense knowledge, which is less codified but most widespread because it is tied to a particular context, thus embedded in social values and beliefs, therefore it is codified but undiffused.

According to the four dimensions, knowledge ranges from completely uncoded and non-diffuse, i.e., personal, to different levels of coding and abstractness, which depend on the efficiency of transmission, i.e., diffusion. Each of these knowledge dimensions can be put into an appropriate form of communication channel, as well as the context within which it is established and built. Therefore, no matter how high the codification and abstraction of information, the domain and the way information is directed can make diffusion more difficult. Therefore, as an assumption of the model, it determines how the population size affects the strength of information diffusion. That is, if the population range is larger, the diffusion is weaker, while if the population range is smaller, the diffusion is stronger [1]. Criticisms of Boisot's model state that codified and uncodified are the only two discrete categories of knowledge, and as such the model is overtly simplified from the perspective of knowledge [4,5]. Boisot himself states that the presentation of the model seems simple, but it is only seemingly so. This is because there are different curves of the flow of information and knowledge in communication situations, from uncodified to codified, where various degrees of abstraction are included [1].

Every function and activity in HEI includes some form of direct or indirect communication where effective communication channels, from the organizational to staff level, are important for disseminating information. Communication channels have a vertical and horizontal line, i.e., from superiors to lower levels and vice versa and between employees at the same hierarchical level. Traditionally HEI relies on bottom-up vertical communication regarding projects and collaboration outside of the institution [6]. Furthermore, [6] explains the establishment of structured relationships as a new type of relationship with external stakeholders, which include specific forms of communication through network events, platforms for cooperation, and partnership agreements between the HEI and various external stakeholders, with the active involvement of academics through teaching and research activities [6]. Associated with new forms of structural relations, [7] explores the Third Mission concept, which integrates a new model of communication as a basis for knowledge transfer through joint activities of academics and external stakeholders. If we look at the organization, there are several types of communication channels, of which the most common are verbal, nonverbal, and written [8]. Verbal communication refers to speech through everyday activities, most often without documentation unless it is about formal meetings and presentations. Nonverbal communication involves the use of body language to send signals such as happiness, contentment, anger, worry, fear, etc. These two types of communication are crucial in understanding and transmitting tacit knowledge among employees. Written communication refers to explicit knowledge and includes codified information, including letters, correspondence, regulations, etc. Written communication is also a formal communication channel that allows longer message processing and possible reuse, such as notices, announcements, manuals, research, etc. In addition to the above channels, an another means of communication can be mentioned in personal communication, or "face-to-face", which includes primarily verbal and nonverbal forms and is one of the "richest" communication channels that can be used within higher education [9]. The greatest advantage of this communication lies in the characteristics of personality and reciprocity. With a wider circle of employees, it improves speaking, writing, and presenta-

tion skills, and the interaction between employees makes it easier to build relationships and greater trust. Group-level communication occurs through departments, project teams, working groups, various committees, and stakeholders. The focus at these levels is on sharing information, discussing different issues and tasks, holding discussions, solving problems, and building consensus. Communication at the organisational level focuses on issues such as vision and mission, statutes, regulations, policies, new initiatives, and organisational knowledge and performance. This communication often has a cascading approach where the administration communicates with the staff through hierarchical channels. Since Web 2.0 has introduced new concepts and tools that are able to operationalize a more society-oriented vision, using these tools it is possible to create, codify, organize and share knowledge, but also spread social activity through personal networks and collaboration in creating new and organizing existing knowledge. This encourages and enables people to achieve greater efficiency through knowledge sharing and virtual interaction through collaboration tools, which has a positive impact on personal knowledge processes [10]. Today, digital communication channels have become effective tools for direct interaction among all actors in HEI. As [11] points out, online communication channels are flexible and allow institutions to present customized information through different devices and for different purposes. Costs associated with online communication channels are independent of the amount of information, distance, or diffusion that is aimed for. In the Croatian example, educational public institutions have a supporting infrastructure, as well as the possibility of integrating cloud technologies by the national academic and research network. The use of open and free tools for communication has intensified because of the pandemic in the last two years, but it has also progressed in the flexibility of the various channels and their effectiveness. We distinguish the most common communication online channels in Croatian HEI: public websites; intranet; cloud infrastructure and software (e.g., Office 365, G-suite); learning management system (LMS); an open database and library; social networks (e.g., Facebook, Twitter, Instagram, YouTube); professional and academic networks (e.g., Linkedin, Academia.edu, ResearchGate, Mendeley); video channels (e.g., YouTube, Teams, Zoom, Meet, Skype); online communities (alumni, informal groups); and instant messaging (e.g., WhatsApp, Viber, Discord).

Furthermore, each organisation consists of some form of a formal and informal network. The term formal structure is used to distinguish public organizational schemes, policies, regulations, and formal hierarchical procedures from non-formal structures such as norms, values, and social groups. Given the characteristics of a formal network, modes of action are easier to show and follow because they are open and public. While hidden or informal networks can be those that build trust between individuals, real sources of influence and power can also be identified through communication channels, which can also be associated with certain negative characteristics: inefficiency, corrupt practices, etc. [12]. Thus, communication networks in higher education institutions can be defined through two groups: formal and informal. Common formal and informal communication channels using new technologies include institution portals and various electronic media, mobile technologies, the cloud, intranet, social channels, video conferencing, blogs, instant messaging podcasts, chats, system wiki, etc.

Formal communication channels, whether written or oral, usually transmit information such as goals, policies, and procedures, which correspond to the set hierarchy. That is, official information through various channels goes to the staff of the next level. This includes meetings of departments, institutions, board meetings, all workers, or working group meetings to enforce organisational rules and regulations. The direction in which formal communication occurs also depends on the structure of the organisation itself, but it most often occurs through two generally different directions: vertical and horizontal [8]. Vertical communication can move down a hierarchy of an organisation or upward, i.e., from a lower organisation to a higher one. Canary and McPhee [8] identify several general purposes of downward communication which are most present within an organisation: the implementation of goals, strategies and tasks; job instructions; procedures and prac-

tices; and performance feedback. Diagonal or horizontal communication occurs among employees at different levels and in different functions. According to [8], horizontal communication falls into some of the following categories: problem-solving within the department; coordination between departments; and advising staff through relevant departments. It is important to emphasize how horizontal communication flows affect the improvement of coordination of activities in a certain level, which allows departments to work with other departments without the need to monitor channels up and down. Many HEI incorporate horizontal communication in the form of working groups, committees, liaison staff, or matrix structures to facilitate such coordination. Ideally, the organisational structure should provide communication flows up and down with horizontal communication, i.e., communication should go in all directions through a formal hierarchy.

Informal communication does unofficially reflect specific channels, as it mostly develops outside the hierarchical structure. It is therefore important because it arises from the social and personal interests of employees and not from the formal requirement of organisational communication. These types of communication channels include social networks, as well as certain informal leisure groups, professional clubs, etc., where the climate is relaxed and pleasant. In addition, through informal communication that occurs within the organisation, not only can the topics of meetings or encounters be discussed spontaneously, but also wider public and social topics. Furthermore, informal or direct types of communication according to [13] are not sufficiently researched in teaching activities, especially through different forms of pedagogical communication between students and professors, considering different multi-channel communication methods.

As knowledge sharing involves the activity of transferring or disseminating knowledge from one person to another, to a group of people, or to an entire organization, information and knowledge from the personal domain are disseminated and linked to the knowledge of a team, department, or organisation. Therefore, the creation or collection of knowledge may come from an individual doing it for an organization, or some groups within that organization, such as a Community of Practice (CoP), yet as [14] point out, it all takes down to on a personal level, where almost everyone performs some activities of creating, collecting and codifying knowledge in the domain of their work. According to [15], values for scholars within the CoP are visible through the following: sharing and accumulating concrete knowledge to solve specific teaching or research problems; building strong links with other academics who possess diverse knowledge, and the ability and skills to build normalized channels for tacit knowledge sharing at a high level; and building an academic reputation in a research field to fulfil one's own and societal values through a contribution to knowledge. Thus, CoP can be characterized more as informal structures with unclear membership and a fluid decision-making process, created by people who share the same interests and a common set of values [16].

In a network, knowledge sharing depends not only on the motivation of individuals to share their knowledge and on the position someone has in the network, but also on the ability to absorb and process knowledge flowing through the network. The effectiveness of knowledge sharing depends on the organisational culture, especially organisational trust. If organisational trust is very low, people will prefer to accumulate knowledge instead of sharing knowledge [15].

Scholars are constantly looking for information because they have a need for a broad knowledge base, with certain differences between different disciplines. The domain context is essential, and it is difficult to make generalizations because scholars from different fields differ in terms of information behavior [17]. The author further states the basic concepts of information behavior that prove to be important for research and relate to the type of information, search context, relevance, prominence, and information overload.

In this sense, the need for information is associated with certain characteristics of the construction of information domains, which relate to the invention, use, and further diffusion of information. Given how information is found and accessed, the influence also exists in the way of communication modalities inside and outside the institution. From the

personal level, from informal and formal groups to the institution as a whole, i.e., public communication, each context has its differences, as presented earlier. In addition, within each context, there is an explicitly tacit form of information diffusion, which is never in the same proportion. Thus, for example, on the personal level, the tacit form prevails, while in the public space of the institution or organisational level, the explicit form prevails.

Given the characteristics of a communication channel, we can determine whether it has a narrow or wide range, and assess the achievement of the diffusion criteria. The intention is for the questionnaire to test an assumption of the I-space model, which states that the larger the target population is, the weaker the diffusion [1]. We examine the strength of diffusion using two established assumptions based on the model assumption and the included communication channels in the survey:

1. If the dominant mode of communication is an implicit–informal form, diffusion is stronger because a smaller circle of people is involved;
2. If the dominant mode of communication is an explicit–formal form directed towards a larger population, the diffusion is weaker.

3. Materials and Methods

This study analyzes the behavior of the scholars through a survey questionnaire, which aims to gain insight into the types of communication channels through which they collect and share information. A link to a survey questionnaire was sent to 383 employees that are listed on websites from seven public polytechnics in Croatia, which are, among other fields, in the technical and social fields of science. By the technical field, we mean the scientific fields of computing and mechanical and electrical engineering, while the social field refers to economics and informatics. A survey was entirely completed by 125 (N) respondents, which was 32% of the sample. Part of the survey questionnaire, regarding communication channels, had 4 questions on the ordinal scale with 9 components per question, with a scale of 7 possible answers (Table 1), [Supplementary Materials]. The components for all questions were the same, except for component 9, where for questions 15 and 16 it refers to libraries and databases, and for questions 17 and 18, where it refers to email. The components, i.e., communication channels, were identified by the authors based on the literature [8,10,11] and personal experience. In Table 2, the 7 possible answers are shown, which include an approximate percentage so that the respondents could determine the answer more precisely. In the following representations, abbreviations are used for each component and answer (Tables 1 and 2).

Table 1. Questions and components in the questionnaire.

Questions			
15.	Finding information for teaching activities	17.	Finding information for administrative activities
16.	Finding information for research activities	18.	Sharing official information

Component			Abbreviation
1	*Conversations with colleagues (directly and indirectly)*		*Conv*
2	*Informal groups inside and outside the Polytechnic*		*Info*
3	*Formal groups of Polytechnics (composition decision)*		*Form*
4	*Internet Polytechnics*		*Inter*
5	*Polytechnic intranet*		*Intra*
6	*Cloud (e.g., Office365)*		*Cloud*
7	*LMS (Learning Management System)*		*LMS*
8	*Social networks*		*SocNet*
9	*Libraries/Databases*	*Email*	*Lib/Data/Email*

Table 2. Scale of response.

Scale	Meaning	Abbreviation
1	Never	Never
2	Rare (<10%)	Rare
3	Sometimes, irregularly (about 30%)	Sometimes
4	Occasionally (about 50%)	Occasionally
5	Often (about 70%)	Often
6	Mostly (about 90%)	Mostly
7	Always	Always

To show the differences between the two fields of science, the responses on the scale were summarized, i.e., the frequencies were summed, to better see the end values and enable a simpler comparison. Answers that refer to 1, 2, and 3 on the scale represent the lowest use and refer to about 30% or less. The answers that refer to 4 on the scale, represent medium values and refers to between 40% and 60%. Answers that refer to 5, 6, and 7 on the scale represent the most frequent use and relate to about 70% or more.

The collected data were processed using the Excel spreadsheet tool and SPSS program for statistical processing. Frequencies, percentages, and the median were used in the descriptive analysis, while in this paper the results are presented in percentages. Considering a large number of channels for each area of activity, a reduction was made through Principal Component Analysis (PCA) to determine new factors to find the latent component in various communication channels and to discover which type of communication is most represented in each activity and with a distinction between science fields.

Table 3 shows the coefficients of internal consistency among the items, i.e., how much the set of items of each question is closely related as a group. Cronbach alpha (α) provides a coefficient of inter-item correlations, that is, the correlation of each item with the sum of all the other items. It is the average correlation among all the items in question [18]. The alpha coefficient (α) is considered acceptable if it is greater than 0.70. Given that this research aims to discover the dominant mode of communication channel for finding information, with the obtained alpha coefficient values (Table 3), we can confirm that the set of components in the four questions has sufficient internal consistency and is reliable for further processing.

Table 3. Cronbach's alpha coefficients.

	Question	α
15.	Finding information for teaching activities	0.756091
16.	Finding information for research activities	0.856856
17.	Finding information for administrative activities	0.80443
18.	Sharing official information	0.805633

4. Results

4.1. Finding Information in the Area of Teaching Activities

Figure 1 shows the percentages of the responses of all respondents (*n* = 125) to the statements for question 15, which queries through which channels scholars most often find information for teaching activities. The components for question 15 are presented in Table 1.

Respondents (26.4%) mostly found information related to teaching in conversations with colleagues, which may indicate informal and implicit (tacit) forms of finding the necessary information. The intranet allows frequent retrieval of information (22.4%), which agrees with the common practice, according to the author's experience, of placing information about subjects, teaching calendars, etc., in that channel of communication. As an occasional possibility for finding teaching activity information, the respondents chose formal groups (28.0%), public internet institutions (24.8%), and informal groups (15.2), (21.6%). The cloud and its services are represented never or infrequently (24.0%), which corresponds with the results of other technologies based on the cloud and are also

poorly represented as a diffusion channel. LMS (39.2%), social networks (36.8%), as well as libraries (23.2%), are the worst represented as a source of information needed for teaching, i.e., these percentages represent the "never" category. The search for information through databases or libraries in this sample shows that there is very little or no use for them in the teaching process, while tacit and informal channels of communication are more present. Is it because polytechnics are declared as higher professional schools, so that information for teaching activities is in the narrower professional groups, both formal and informal, through direct communication?
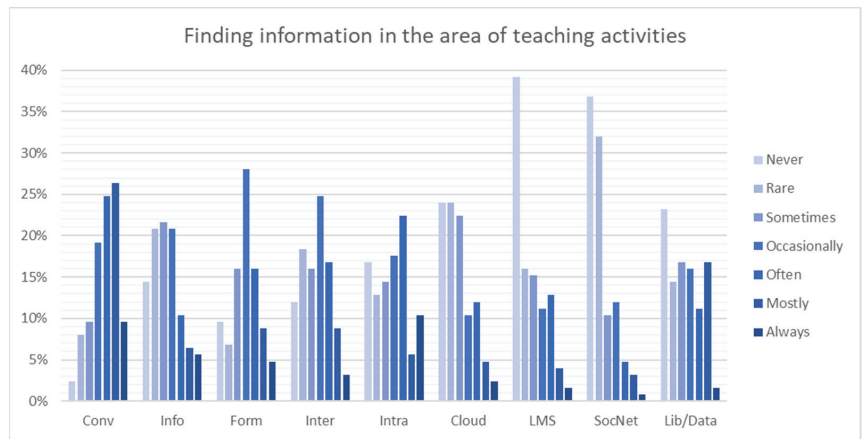

Figure 1. Ways of finding information in the area of teaching activities. (The complete legend is shown in Tables 1 and 2).

Table 4 shows the percentages of answers to question 15, concerning the scholars' affiliation to the technical (T, *n* = 64) or social (S, *n* = 61) field of science. Values with a difference of less than 10% are marked in gray, values with a difference of between 10 to 15% are in black, while bold values show a difference of more than 15% between areas.

Table 4. Ways of finding information in the area of teaching activities regarding the field of science.

	About 30% or Less		About 40% and 60%		About 70% or More	
	T	**S**	**T**	**S**	**T**	**S**
Conv	**13**	**28**	23	15	64	57
Info	53	61	25	16	22	23
Form	48	36	31	25	**20**	**39**
Inter	52	41	30	20	**19**	**39**
Intra	47	41	17	18	36	41
Cloud	73	67	11	10	16	23
LMS	67	74	14	8	19	18
SocNet	80	79	14	10	6	11
Lib/Data	58	51	19	13	23	36

The results indicate that there are certain differences within the conversation channel; the technical field uses it to a greater extent, while the social area uses the channels of formal groups and the internet more to find information for teaching activities. This may indicate that the technical field finds necessary information in more implicit and less formal ways, as communities of practice and internet portals offer information on specific areas of expertise, for example, related to a specific programming language, general programming, etc.

4.2. Finding Information for the Area of Scientific Activities

Figure 2 shows the percentages of the answers to question 16; the channels through which teachers most often find information for research activities. The figure shows the answers of all respondents (*n* = 125). The components for question 16 are presented in Table 1.
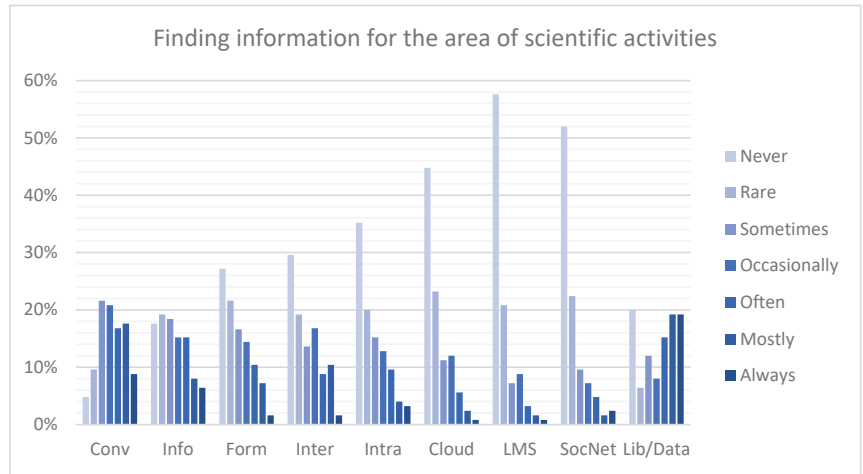

Figure 2. Ways of finding information in the area of scientific activities. (The complete legend is shown in Tables 1 and 2).

Information related to scientific production scholars found that 19.2% mostly or always found information through databases and libraries, i.e., as formally explicit forms, and occasionally in conversations with colleagues (20.8%), i.e., as an informally tacit form. Informal groups are not represented here, or at least very rarely. In addition, within this sample, we can assume that the cloud and related technologies are the least used.

Table 5 shows the percentages of answers to question 16, concerning the scholars' affiliation to the technical (T, *n* = 64) or social (S, *n* = 61) field of science.

Table 5. Ways of finding information in the area of scientific activities regarding the field of science.

	About 30% or Less		About 40% and 60%		About 70% or More	
	T	**S**	**T**	**S**	**T**	**S**
Conv	36	36	20	21	44	43
Info	56	54	20	10	23	36
Form	67	66	16	13	17	21
Inter	**72**	**52**	17	16	**11**	**31**
Intra	75	66	9	16	16	18
Cloud	80	79	13	11	8	10
LMS	86	85	11	7	3	8
SocNet	86	82	9	5	5	13
Lib/Data	44	33	9	7	47	61

The results indicate that there are certain differences in the use of the institution's internet channel and databases or libraries between the fields; the social field uses it to a greater extent than technical field to find information for scientific activities. All other statements indicate no major differences between the social and technical fields.

4.3. Finding Information for the Area of Administrative Activities

Figure 3 shows the percentages of the answers to question 17; the channels through which teachers most often find information for administrative tasks. The figure shows the answers of all respondents (*n* = 125). The components for question 17 are presented in Table 1.
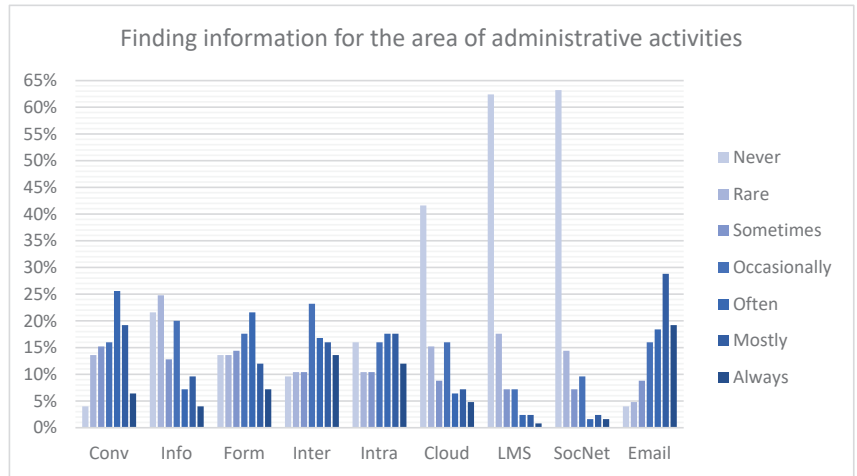

Figure 3. Ways of finding information in the area of administrative activities. (The complete legend is shown in Tables 1 and 2).

For the needs of institutional and administrative work, respondents mostly (28.8%) collect information via email and often (25.6%) through conversation with colleagues and within formal groups (21.6%). Social networks, LMS, and the cloud are the least used. According to the results of this sample, it is obvious that email still has primacy in business communication, although there are various other possibilities for exchanging such information, such as the cloud, which offers significantly higher modalities and platforms for this type of communication, for example, the DMS (Document Management System).

Table 6 shows the answers to question 17, concerning the scholars' affiliation to the technical (T, *n* = 64) or social (S, *n* = 61) field of science.

Table 6. Ways of finding information in the area of administrative activities regarding the field of science.

	About 30% or Less		About 40% and 60%		About 70% or More	
	T	S	T	S	T	S
Conv	31	34	17	15	52	51
Info	58	61	22	18	20	21
Form	47	36	17	18	36	46
Inter	**45**	**15**	25	21	**30**	**64**
Intra	**47**	**26**	19	13	**34**	**61**
Cloud	69	62	16	16	16	21
LMS	89	85	8	7	3	8
SocNet	88	82	11	8	2	10
Email	**25**	**10**	17	15	**58**	**75**

The results indicate that there are noticeable differences within the use of the institution's internet and intranet, and minor differences in the databases or libraries channel; the social field uses it to a greater extent than technical field to find information for administrative activities. In all other components, the use of communication channels shows no major differences.

4.4. Sharing Official Information within the Institution

Figure 4 shows the percentages in the answers to question 18, i.e., the channels through which scholar most often share or forward formal information within their institution. The figure shows answers of all respondents (*n* = 125). The components for question 18 are presented in Table 1.
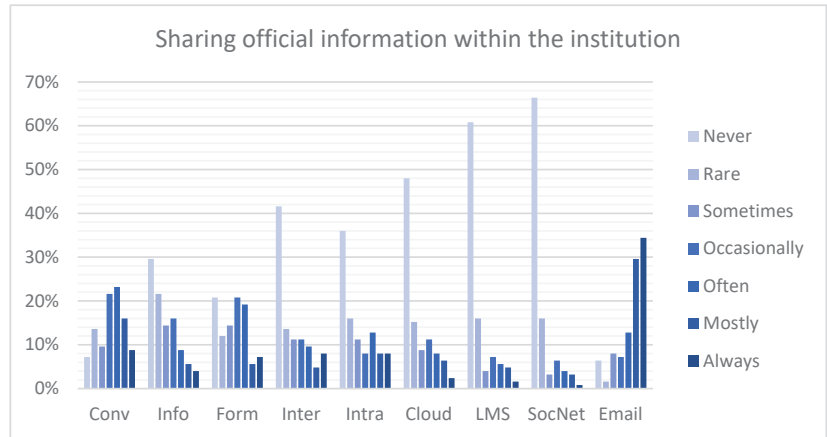

Figure 4. Ways of sharing official information within the institution. (The complete legend is shown in Tables 1 and 2).

The dissemination of information related to formal activities within the institution is always (34.4%) or mostly (29.6%) forwarded by email. If the information is received in some other way, the results of this sample show that, to the greatest extent, the information is forwarded by email. According to [19], the information sent and received takes different forms in accordance with the increasing methods of communication, but also customs, habits, and expectations. Given the long-term use of email, we can say that it is the main and basic form of both business and private communication. Often, transfer of information occurs through conversation (23.2%) or formal groups (20.0%), i.e., through different types of meetings, which most often include formal and informal conversation. It is to be expected that within this context, institutional formal groups are the generators of such information, but they are not the main diffuser. Thus, in addition to explicit form, i.e., formal communication, the implicit form is used to a greater extent. Other components that are never used by most respondents are the cloud and related technologies, such as LMS and social networks. Given the wide possibilities of using the cloud, which combine with real-time communication services, and given the rise in working from home in the last two years, the results in this sample show that this form is not adequately included in the daily work of scholars.

Table 7 shows the answers to question 18 but concerning the scholars' affiliation to the technical (T, *n* = 64) or social (S, *n* = 61) fields of science.

In statements indicating the sharing of formal information within the institution, there are no major differences between the percentages in the responses of social and technical respondents, except for the email channel; social field respondents used email more than technical field respondents.

According to the total years of work in higher education, 68.8% of respondents to this research have been working for more than 10 years. Thus, it is possible to assume that the majority of the respondents have a certain established way of selecting and using communication channels in their work. The differences between respondents who have worked for more than 10 years and those who have worked for less than 10 years did not prove to be significant in any of the information-seeking activities.

Table 7. Percentages of answers for the components in question 18 regarding field of science.

	About 30% or Less		About 40% and 60%		About 70% or More	
	T	**S**	**T**	**S**	**T**	**S**
Conv	28	33	23	20	48	48
Info	66	66	19	13	16	21
Form	55	39	17	25	28	36
Inter	73	59	11	11	16	30
Intra	69	57	6	10	25	33
Cloud	73	70	11	11	16	18
LMS	78	84	9	5	13	11
SocNet	83	89	11	2	6	10
Email	23	8	11	3	**66**	**89**

4.5. Principal Component Analysis (PCA)

PCA is a multivariate method that reduces dimensionality and was chosen for component analysis to make the data clearer and easier to understand [20]. This method forms new latent variables, i.e., components, which are mutually independent, and those that are "sufficiently informative" are retained [21]. Here, we will reduce the number of components for each question.

Before extracting the components, tests to assess the goodness of fit of the data, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity, were performed [22]. Figure 5 shows the values obtained for the technical and social areas where the suitability test indicates moderate and medium index values, ranging from 0.661 to 0.768, with *p*-value < 0.05, which confirms the justification of the factor analysis.

Technical	**KMO**	Bartlett's Test of Sphericity			**Social**	**KMO**	Bartlett's Test of Sphericity		
		Approx. Chi-Square	df	Sig.			Approx. Chi-Square	df	Sig.
15	0,732	153,911	36	,000	**15**	0,768	167,278	36	,000
16	0,766	310,792	36	,000	**16**	0,759	244,599	36	,000
17	0,741	197,253	36	,000	**17**	0,753	160,865	36	,000
18	0,723	174,168	36	,000	**18**	0,661	186,610	36	,000

Figure 5. Kaiser-Meyer-Olkin (KMO) and Bartlett's test values of sampling adequacy.

To reduce the number of components, the eigenvalue, the percentage of variance, and the cumulative percentage of variance were determined for each component. Although there is another way to determine the number of extracted components, for this analysis, a Cattell diagram (Scree plot) was used to evaluate the optimal number of components for extraction through several iterations for both fields of science (Figures 6 and 7). Two factors for both fields are retained, while the other components enter the flatter part of the curve, which means that each subsequent component has a smaller and smaller number of eigenvalues.

Orthogonal Varimax rotation was chosen as the rotation technique, as it is the most common rotation technique in factor analysis and results in factor structures that are not correlated [23]. Given that the main goal is to enable an easier interpretation of the results using this rotation solution, we wanted to show the best fit and suitability, either conceptually or/and intuitively. Furthermore, the criterion for the statistical significance of factor loadings, with 95% certainty, offers a guideline as to whether the size of the examined sample is considered large enough for a certain level of factor loading to be significant [23]. Given that the sample size for the technical area is N = 64, and for the social area is N = 61, the factor loading that can be considered significant, according to [23], with 95% certainty, is >0.70.
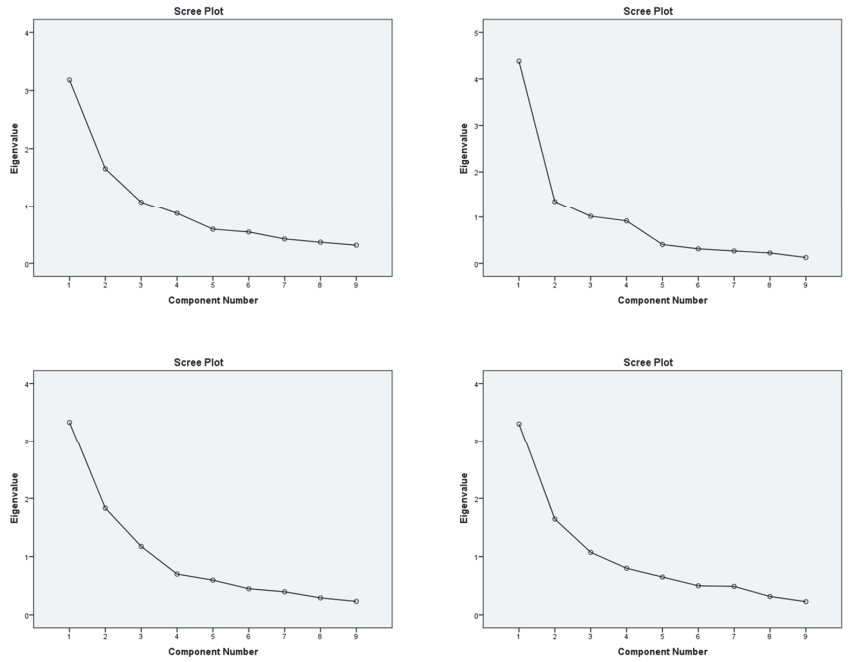
Figure 6. Cattell diagram of components and eigenvalues for the technical field.
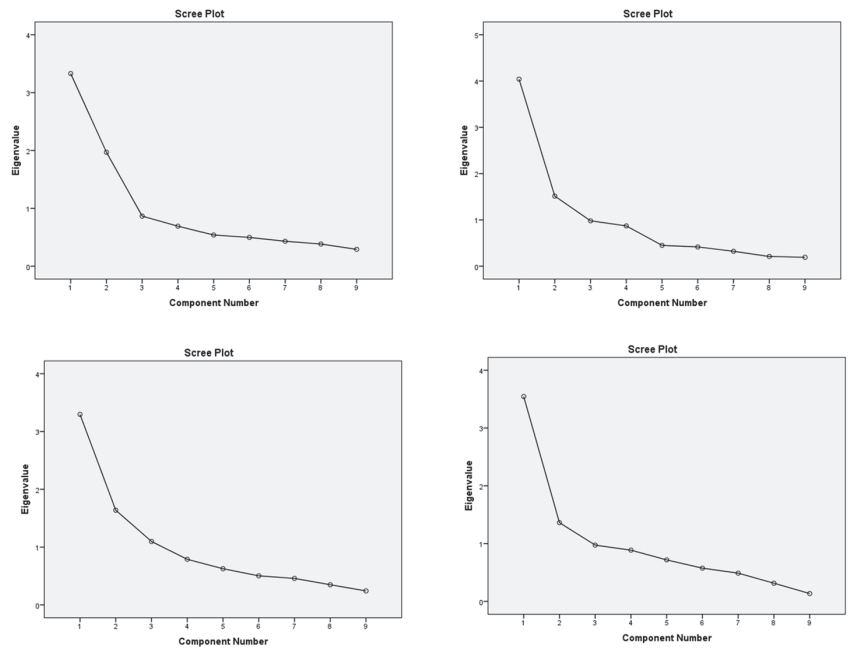

Figure 7. Cattell diagram of components and eigenvalues for the social field.

Figure 8 shows a matrix of rotating components for two areas (T and S) and four questions. Components that have factor loadings above 0.7 are shown, and the others are excluded from further analysis. It is clear that rotation of the factors simplifies the structure by maximising the loading of the components within each factor, which allows us to clearly identify them.

T	Component		S	Component	
	1	2		1	2
15.9	0,778		15.8	0,783	
15.6	0,778		15.6	0,758	
15.8			15.7	0,724	
15.4			15.5		
15.5			15.4		
15.7			15.1		0,798
15.1		0,872	15.3		0,791
15.2		0,822	15.2		
15.3			15.9		

T	Component		S	Component	
	1	2		1	2
16.6	0,867		16.6	0,828	
16.4	0,811		16.7	0,806	
16.5	0,791		16.8	0,737	
16.8	0,783		16.5	0,725	
16.7	0,763		16.4	0,714	
16.3	0,727		16.9		
16.9			16.1		0,860
16.1		0,891	16.2		0,821
16.2		0,862	16.3		

T	Component		S	Component	
	1	2		1	2
17.5	0,854		17.2	0,859	
17.6	0,741		17.1	0,831	
17.4	0,734		17.3	0,821	
17.9			17.9		
17.8			17.8		0,776
17.7			17.6		0,775
17.1		0,879	17.7		0,746
17.2		0,837	17.5		
17.3			17.4		

T	Component		S	Component	
	1	2		1	2
18.5	0,821		18.5	0,822	
18.4	0,794		18.6	0,808	
18.6	0,770		18.4	0,787	
18.3	0,708		18.3		
18.7			18.7		
18.9			18.9		
18.8			18.1		0,854
18.2		0,889	18.2		0,814
18.1		0,846	18.8		

Figure 8. Rotated Component Matrix regarding the science area. Extraction Method: Principal Component Analysis. Rotation converged in three iterations.

The components are often grouped around similar variables, in this case around similar modes of communication. For all four questions there are two factors with several components that are similar. For better visibility, in Figures 9 and 10, we have presented the two obtained factors with their components regarding the area and activities where they are shown. The name of the factor is not assigned, but the essential characteristics that determine the conceptual meaning are indicated. For Factor one, the components that are singled out for both scientific fields from questions 15, 16, and 18 have the communication channel characteristics of explicit–formal, public, and wide-scope. Question 17 indicates the difference between the two fields, where the technical field has the characteristic of explicit–formal, while the social field uses implicit–informal communication channels. For Factor two, the components that are singled out for both scientific fieldsin questions 15, 16, and 18 have the communication channel characteristics of implicit–informal, personal, and narrow-scope. Question 17 indicates the difference between the two areas, where the technical field has the characteristic of implicit–informal, while the social field uses explicit–formal communication channels.

Factor 1

	Technical	factor loading	Component characteristic		factor loading	Social
15	.9 Libraries / Databases	0,778	explicit-formal, public, wide scope of the population		0,783	.8 Social networks
	.6 Cloud	0,778			0,758	.6 Cloud
					0,724	.7 LMS
16	.6 Cloud	0,867	explicit-formal, public, wide scope of the population		0,828	.6 Cloud
	.4 Internet Polytechnics	0,811			0,806	.7 LMS
	.5 Polytechnic intranet	0,791			0,738	.8 Social networks
	.8 Social networks	0,783			0,725	.5 Polytechnic intranet
	.7 LMS	0,763			0,714	.4 Internet Polytechnics
	.3 Formal groups	0,727				
17	.5 Polytechnic intranet	0,854	explicit-formal	implicit-informal	0,859	.2 Informal groups
	.6 Cloud	0,741			0,831	.1 Conversation
	.4 Internet Polytechnics	0,734			0,821	.3 Formal groups
18	.5 Polytechnic intranet	0,821	explicit-formal, public, wide scope of the population		0,822	.5 Polytechnic intranet
	.4 Internet Polytechnics	0,794			0,808	.6 Cloud
	.6 Cloud	0,770			0,787	.4 Internet Polytechnics
	.3 Formal groups	0,708				

Figure 9. Areas and activities referred to by Factor one.

Factor 2

	Technical	factor loading	Component characteristic		factor loading	Social
15	.1 Conversation	0,872	implicit-informal, personally, narrower scope of population		0,798	.1 Conversation
	.2 Informal groups	0,822			0,791	.3 Formal groups
16	.1 Conversation	0,891	implicit-informal, personally, narrower scope of population		0,860	.1 Conversation
	.2 Informal groups	0,862			0,821	.2 Informal groups
17	.1 Conversation	0,879	implicit informal	explicit-formal	0,776	.8 Social networks
	.2 Informal groups	0,837			0,775	.6 Cloud
					0,746	.7 LMS
18	.2 Informal groups	0,889	implicit-informal, personally, narrower scope of population		0,854	.1 Conversation
	.1 Conversation	0,846			0,814	.2 Informal groups

Figure 10. Areas and activities referred to by Factor two.

5. Discussion

Overall, finding information for teaching activities dominates conversation communication channels, which points to informal and implicit forms of finding information, with frequent use of the intranet, and occasional use of the internet and formal groups of the institution. Since verbal and nonverbal communication form part of the informal methods of seeking information, according to [8], they form the basis for the understanding and

transfer of tacit knowledge between employees. When we look at the difference between the science fields, social scientists use formal groups and the internet more than technical scientists, but use less conversation channels. Within this sample, respondents from the technical field are characterized by using informal ways to request information for teaching activities, which corresponds to the characteristics of the CoP. This may include, inter alia, finding information for teaching purposes within different professional groups sharing the same interests and values [16].

Croatian scholars in this sample find information related to scientific activities through databases or libraries, and often in conversations with colleagues. The characteristics of the two forms of explicit and implicit ways can be intertwined in their appearance within this activity. Very often we start research based on an idea formed in a conversation with colleagues, then continue research through explicit forms, to exchange certain knowledge again within a narrower scope of the population. There are also certain differences between the science fields; social scientists use more internet than technical scientists, who uses conversation channels more; however, databases and library channels are used equally.

Information seeking for the purposes of administrative activities include the email channel, whether initiated by conversation or formal group activities. To a lesser extent, the intranet, internet, and formal groups can be singled out, which are used as channels occasionally, although they very often represent the basis for any search with regard to administrative tasks and related documentation. We can also look at emails and formal groups in the context of vertical communication, and conversation in the context of horizontal communication, bringing together the different categories of activities mentioned in [8]. When we look at the differences between science fields, there are noticeable differences in the use of the internet and intranet institution channels, which are favored by social science. They both use conversation and email to a great extent.

When sharing official information, and given that it also includes administration to a greater extent, the email channel comes to the fore, showing the highest usage values of all activities. As another sharing channel, conversation stands out, in addition to formal groups. There is only one difference between the science fields, regarding the email channel, which is used to a much greater extent within the social science group of scholars. Thus, administrative activities, whether searching for or sharing information, correspond to a formal network structure that includes a procedural hierarchy, policy, and organisational schemes, and is generally public [9].

Considering the obtained results for the two assumptions given in the I-space model [1], and three basic groups of scholars' activities, the following conclusions can be drawn for the obtained data:

- In finding information for teaching activities, the most common form of communication is implicit–informal, and it is to be assumed that there is a stronger diffusion of information;
- In finding information for the needs of research activities, the most common form of communication is explicit–formal, thus the diffusion is smaller within the population;
- In finding information for the needs of administrative activities, the most common form of communication is explicit–formal, thus the diffusion is smaller within the population.

According to PCA results, the number of components was reduced to two factors for each scholar's activities. The first factor revealed that the components in the technical science field, and all questions, have explicit–formal characteristics. For the social science field, they are mostly explicit–formal, except for question 18 where informal and implicit dominates. In the second factor, although it has a higher factor loading, only two components are present that have the characteristics of an implicit–formal form of communication. There is an exception in the field of social science, where for information on administrative activities, the channel characteristics correspond to an explicit–formal mode.

However, it is necessary to state the most common possible shortcomings in this type of analysis, such as the inadequate selection of the number of components and insufficient clarity of data, which is a subjective aspect with many differences in opinion [21,22]. It should be noted that the key communication channels for searching and sharing information were

determined by factor analysis, but there is no possibility to go into deeper elaboration using this method. In addition, through descriptive analysis, it was shown that the responses were scattered due to a scale of seven responses and an insufficiently large sample. Generalizing on the basis of one sample, regardless of its size, is always problematic, therefore all conclusions are presented in the form of possible applications in the context of the given sample. In this research, a purposive sample was used from selected public Croatian polytechnics that had a social and technical field in their curriculum; therefore, in further research, the sample can include other polytechnics, as well as universities. Given that similar research, which includes all three activities of academics, has not been found outside of Croatia, the disadvantage is that a sufficiently good comparison is not possible with regards to the context of the activity.

6. Conclusions

From the descriptive analysis, it can be concluded that for the needs of teaching activities, the surveyed Croatian scholars find information through direct communication through conversation (tacitly), while for the needs of research activities they find information in databases or libraries (explicitly). In administrative activities, if the information is obtained or shared, the most common channel of communication is email. To a certain extent, there is a difference in frequencies between the social and technical science fields when finding information for administrative activities. There are several contributions from this research:

- According to the results of this research, certain newer technologies, such as the cloud, are not used enough. With their greater involvement in communication channels, access to modalities would be significantly increased, and various flexible solutions would be offered;
- The results within this sample indicate that libraries and databases are to a greater extent included only for the needs of information in scientific work, while they are used the least for teaching activities. In this context, it is necessary to ensure and offer, in a transparent manner, various modalities of access to libraries and databases, given that there are a certain number of higher education institutions that do not have a library within the institution for various reasons;
- Although the explicit–formal type of communication prevails through the four basic activities of academics, implicit–informal channels have great value for each activity, and this is most reflected in teaching activities. Given that the surveyed sample are all Croatian polytechnics that are by nature oriented towards the profession, it can be assumed that personal and informal forms of information flow play a major role. In doing so, one should consider whether they are formal professional groups (communities of practice) or more isolated groups and consider their possible support and development.

Future research can be focused on specific forms of communication, such as formal groups, that are proved to be an explicit and implicit link between different forms of communication. It is important to further investigate the form of formal groups, their appearance, modalities, influence, and functionality.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Boisot, M.H. *Information Space: A Framework for Learning in Organizations, Institutions and Culture*; Routledge: London, UK, 1995.
2. Wang, L.; Buckland, M. From Fief to Clan: Boisot's Information Space Model as a Documentary Theory for Cultural and Institutional Analysis. *Proc. Doc. Acad.* **2016**, *3*, 10. [CrossRef]
3. Fascia, M. Knowledge transfer discussion: Paper 1.002. *Theor. Model. EMRI J. Multicult. Res.* **2014**, *3*, 1–29. Available online: https://www.academia.edu/7138485/Knowledge_transfer_discussions_Paper_1_002_Theoretical_Models (accessed on 7 April 2022).
4. Mohajan, H.K. The Impact of Knowledge Management Models for the Development of Organizations. *J. Environ. Treat. Tech.* **2017**, *5*, 12–33. Available online: https://mpra.ub.uni-muenchen.de/83089/1/MPRA_paper_83089.pdf (accessed on 7 April 2022).
5. Fascia, M.; Fascia, S. The value of knowledge transfer mechanisms. *J. Strategy Oper. Econ. JSOE* **2017**, *3*, 1.
6. Broström, A.; Feldmann, A.; Kaulio, M. Structured relations between higher education institutions and external organisations: Opportunity or bureaucratisation? *High. Educ.* **2019**, *78*, 575–591. [CrossRef]
7. Compagnucci, L.; Spigarelli, F. The Third Mission of the university: A systematic literature review on potentials and constraints. *Technol. Forecast. Soc. Change* **2020**, *161*, 120284. [CrossRef]
8. Canary, H.E.; McPhee, R.D. *Communication and Organizational Knowledge. Contemporary Issues for Theory and Practice*; Taylor & Francis: New York, NY, USA, 2011.
9. Badau, K.M. Administration of universities and channels of communication in Nigeria. *Arts Humanit. Open Access J.* **2018**, *2*, 302–306. [CrossRef]
10. Razmerita, L.; Kirchner, K.; Sudzina, F. Personal Knowledge Management: The role of Web 2.0 tools for managing knowledge at individual and organisational levels. *Online Inf. Rev.* **2009**, *33*, 1021–1039. [CrossRef]
11. Constantinescu-Dobra, A.; Coțiu, M.A. Communication Channels. In *Communication Management*; IntechOpen: London, UK, 2021. [CrossRef]
12. Goes, S. Understanding informal networks in higher education institutions: Theoretical concepts from a Russian and Norwegian perspective. *Barents Stud. Peoples Econ. Politics* **2015**, *2*, 34–49. Available online: https://nordopen.nord.no/nord-xmlui/handle/11250/2382156?locale-attribute=en (accessed on 7 April 2022).
13. Chernenko, O.; Roienko, S.; Balanutsa, O.; Ivashchenko, B.; Romankova, K. The role of communication and dialogue during studies in higher education institutions. *Laplage Rev.* **2021**, *7*, 548–556. [CrossRef]
14. Dalkir, K. *Knowledge Management in Theory and Practice*; MIT Press: Cambridge, MA, USA, 2013. [CrossRef]
15. Yu, D.; Zhou, R. Tacit Knowledge Sharing Modes of University Teachers From the Perspectives of Psychological Risk and Value. *Int. J. High. Educ.* **2015**, *4*, 214. [CrossRef]
16. Jordanne, C. Faculty Learning Communities to Support Technology Integration: A Literature Review Integration. *Transform. Dialogues Teach. Learn. J.* **2016**, *9*, 1–19. Available online: https://td.journals.psu.edu/td/article/view/1063 (accessed on 7 April 2022).
17. Vilar, P. Information behaviour of scholars. *Libellarium* **2014**, *7*, 17–39. [CrossRef]
18. Cohen, L.; Manion, L.; Morrison, K. *Research Methods in Education*; Routledge, Taylor & Francis Group: London, UK, 2007.
19. Jones, W.; Dinneen, J.D.; Capra, R.; Diekema, A.R.; Pérez-Quiñones, M.A. Personal Information Management. In *Encyclopedia of Library and Information Science*, 4th ed.; Levine-Clark, M., McDonald, J., Eds.; Taylor & Francis: New York, NY, USA, 2017; pp. 3584–3605. [CrossRef]
20. Introduction to SAS. UCLA: Statistical Consulting Group. Available online: https://stats.oarc.ucla.edu/sas/modules/introduction-to-the-features-of-sas/ (accessed on 2 June 2022).
21. Fazlić, S.; Đonlagić, S. *Primjena Faktorske Analize u Identificiranju Dimenzija Kvalitete Visokoobrazovne Usluge*; Poslovna Izvrsnost Hrcak: Zagreb, Croatia, 2016; Available online: https://hrcak.srce.hr/170734 (accessed on 16 August 2022).
22. Williams, B.; Onsman, A.; Brown, T. Exploratory factor analysis: A five-step guide for novices. *Australas. J. Paramed.* **2010**, *8*, 1–13. [CrossRef]
23. Hair, J.; Black, W.; Babin, B.; Anderson, R. *Multivariate Data Analysis*, 7th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2010.

Measuring and Promoting the Success of an Open Science Discovery Platform through "Compass Indicators": The GoTriple Case

Stefano De Paoli [1,*], Emilie Blotière [2], Paula Forbes [1] and Sona Arasteh-Roodsary [3]

[1] Division of Sociology, Abertay University, Bell Street, Dundee DD11HG, UK
[2] CNRS National Center for Scientific Research, 8Cr de Humanités, 93300 Aubervilliers, France
[3] Max Weber Stiftung, Rheinallee 6, 53173 Bonn, Germany
* Correspondence: s.depaoli@abertay.ac.uk

Abstract: Previous research on indicators for measuring the success of Open Science tends to operate at a macro/global level and very rarely addresses the need to measure success at the level of a single project. However, this previous research has the merit of arguing for the definition of indicators that offer an alternative to more traditional bibliometric measures or indicators that focus on mere performance. This paper is the outcome of work conducted for a specific project that aims to build a discovery platform for social sciences and humanities, the platform GoTriple. GoTriple is designed taking inspiration from Open Science principles and has been built through a user-centered approach. The paper details the practice-led work conducted by the GoTriple team for assessing the meaning of the term success for the project and to identify indicators. To this end, this paper proposes the concept of compass indicators and presents how the project team arrived at the definition of this concept. The paper also highlights a distinction between compass indicators, which are modest measures, and key performance indicators, which tend to be tied up with measurable objectives. Compass indicators are defined as indicators that do not aim to achieve a specified numerical target of success but rather explain the journey of a project toward achieving certain desirable outcomes and offer insights to take action. Compass indicators defined for the project embrace areas such as diversity, inclusivity, collaboration, and the general use of the platform. In the final discussion, the paper offers reflections on the potential relevance of the notion of compass indicators and closes with a discussion of the next steps for this work.

Keywords: indicators; open science; discovery; diversity; user; social sciences; humanities

1. Introduction

Social Sciences and Humanities (SSH) research is divided into several disciplines, sub-disciplines, and languages. While this specialization makes it possible to investigate a variety of topics, it also leads to fragmentation. Use and reuse of SSH publications is suboptimal, interdisciplinary collaborations are often missed, and as a result the societal, economic, and academic impacts of SSH are limited. Having SSH disciplines embrace Open Science more concretely can help in addressing some of these issues by breaking down barriers and supporting wider collaborations. In line with this, the Horizon 2020 TRIPLE project (https://project.gotriple.eu/about/, accessed 7 August 2022) seeks to offer a contribution to address some of these issues and break down the siloed approach often common in SSH. It is trying to do so with the design, development, and dissemination of the GoTriple discovery platform (https://www.gotriple.eu/). GoTriple is a digital solution based on cutting edge web technology and design which should facilitate better discovery for SSH and foster wider collaboration. With the term discovery we mean the capacity to find, expose, and display material such as literature, data, projects, people etc., that researchers would need for their work [1].

The aim of GoTriple is to make it much easier for scientists, citizens, and business organizations to access scientific publications, data, data processing platforms, and data processing services, with a focus on SSH. This effort is equally linked with the development of Open Science as it is with the fostering of Open Access. The Organisation for Economic Co-operation and Development (OECD) provides the following definition of Open Science: "To make the primary outputs of publicly funded research results—publications and the research data—publicly accessible in digital format with no or minimal restriction" [2] (p. 7). They add yet another important aspect: "Open Science is about extending the principles of openness to the whole research cycle, fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done". The TRIPLE project, via the development of GoTriple, shares the ambition to contribute to Open Science. The technical and philosophical work underpinning the development of GoTriple have been described in detail in the paper by Dumouchel et al. [3], whilst the paper by Achenbach et al. [1] discusses the differences between GoTriple and other mainstream discovery services and, in particular, Google Scholar. GoTriple had been initially conceived based on the results of the OPERAS design study which were presented in detail in a white paper [4] and other studies, for instance, the "European survey on scholarly practices and digital needs in the arts and humanities", conducted by the Digital Methods and Practices Observatory Working Group (DiMPO) [5]. These publications identified two important gaps: the difficulty to find resources and researchers in SSH because of their huge fragmentation and the general difficulties that exist in their discovery. They highlighted the existence of a variety of diverse SSH digital repositories as well as the numerous small SSH units in Europe. It is this fragmentation of publications, projects, and people that requires efforts to be made to bring resources together and offer the SSH community digital tools and services that can act as single access points. The GoTriple platform therefore has the ambition to try to address some of the known barriers of interconnection and interoperability across SSH disciplines and communities. There is also a visible gap in the existing European Open Science Cloud service marketplace (https://search.eosc-portal.eu/search/service?q=*, accessed 7 August 2022), which largely offers services for natural sciences, where on the contrary, SSH services are fairly limited in both numbers and scope. The GoTriple platform therefore is a next necessary step forward for value generation from the SSH community. By becoming an established European discovery platform for SSH resources, GoTriple aims to experiment with and develop new ways to conduct research and especially interdisciplinary research supporting SSH. It is then important to also reflect on what success for a project such as GoTriple could mean. Consequently, in this paper, we discuss the strategy which has been developed by the project for measuring the success of the platform. In doing so, the paper seeks to present an answer to the following practical research problem: when considering the success of an Open Science discovery platform, what should we measure and how? In particular, we aim to explore important aspects associated with the creation and nurturing of a healthy and varied user community of SSH scholars and other actors.

The TRIPLE project's methodology is based on the combination of two interlinked pillars: the first is the adoption of Open Science practices for the design and development; and the second is a user-centered approach to design with a focus on understanding the researchers' needs and co-designing core aspects of the platform with them. These two pillars have supported the creation of GoTriple as an innovative multilingual and multicultural discovery solution for SSH. GoTriple presently supports discovery in 10 languages (Italian, Spanish, Portuguese, German, French, Polish, Croatian, Slovene, Ukrainian, and English) and work is ongoing to include additional languages. The platform's ambition is to provide a central access point that allows discovery at a European scale in the above languages. This is achieved through the harvesting of open repositories, archives, or other open publishing platforms, offering researchers access to publications and other resources. The platform technology is also based on the Isidore (https://isidore.science/, accessed 8 August 2022) search engine [6]. Compared to Isidore, GoTriple additionally

boosts several innovative services and specific tailored features (see Figure 1 for the main home page). The innovative services include a trust building system, an annotation tool, a crowdfunding solution, and a recommender system, as well as a visual search interface. Tailored features include various dashboards which can be used by the user to explore publications, such as a dashboard related to publications in specific SSH disciplines (see Figure 2), user profiles, and others.
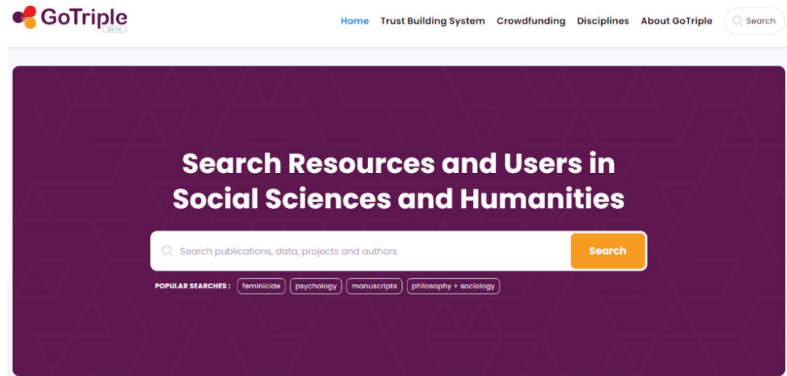

Figure 1. The home page of GoTriple with the search engine. A beta prototype of GoTriple was released in October 2021 and a near final version was released in September 2022. At the time of writing, concrete actions for building a lively user community have started via a user engagement strategy. A further important aspect is that the platform will be one of the dedicated services of OPERAS (https://operas-eu.org/, accessed 10 August 2022) [7], the research infrastructure supporting open scholarly communication in SSH in the European Research Area. Moreover, GoTriple will be a dedicated service of the European Open Science Cloud (EOSC) (https://eosc-portal.eu/, accessed 10 August 2022), thus further contributing to achieving Open Science in Europe.
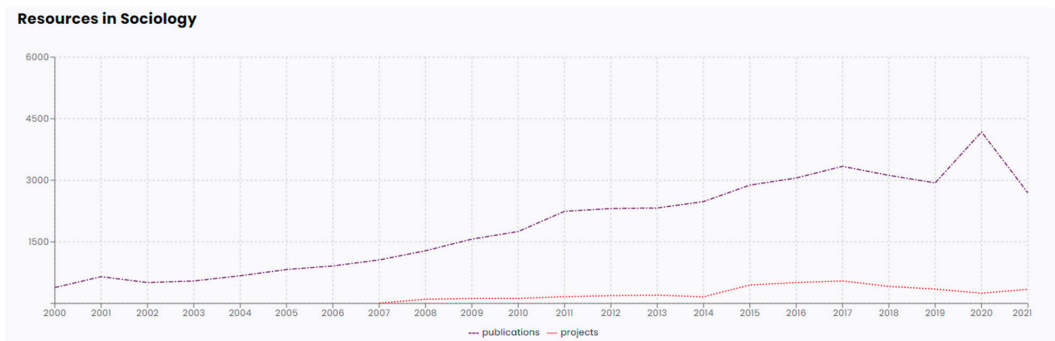
Figure 2. Example of a component of the GoTriple discipline dashboard (see https://www.gotriple. eu/disciplines, accessed 12 August 2022) showing the number of projects and publications discoverable for sociology over time.

One problem related not so much to the design but to the achievements of the GoTriple platform is being able to define a set of indicators that can be used to measure success. Obviously for TRIPLE, with it being a European Union funded project, the partners had to propose a set of key performance indicators (KPIs), to measure certain achievements related to the receipt of public funding and tied to the project objectives. These include the number of documents harvested or the number of users at the end of funding. A performance indicator is usually defined as "an item of information collected at regular

intervals to track the performance of a system" [8] (p. 1). Whilst KPIs have a role in assessing performance in achieving certain critical objectives for, e.g., an organization, and they should help to substantially increase a certain performance [9], they are essentially tied with targets [10] to be achieved over time. Not meeting a KPI is often synonymous of failure as the performance was not what was expected and the objectives were not met.

In our specific case, whilst meeting the funding KPIs is fundamental for measuring the achievement of high-level objectives, we also need to be mindful that the KPIs proposed at the time of submitting a project proposal are not necessarily a good representation of how the project has evolved over time and they are potentially obsolete after the end of the funding period. Moreover, it is the position of the authors that adopting strict KPIs may not be the only approach to measure the success of an Open Science platform such as GoTriple, as success should be seen as a journey which can go in very many directions more than as a performance target. Indeed, the creation of a new platform is a practical problem which requires reflection, discussion, user involvement, refinement, and further action. For this reason, the TRIPLE project partners have conducted work to define an alternative set of indicators to measure success for an Open Science community platform, based on a modest practice-led approach [11,12] rather than one of target achievements. There also is discussion of the importance of adopting both qualitative and quantitative indicators for measuring Open Science, e.g., [13,14]. For practical reasons which will be discussed later, the indicators we propose in this work are quantitative in nature, however they are not targets and are used consequently in an inductive manner as potential guides for further action based on our expert interpretation. The goal of this paper is then to present part of the practical work that was conducted to address the problem of "what and how" to measure success for GoTriple. As we will see later in the paper, we have defined the indicators for measuring success in GoTriple as **compass indicators**. This specific name was adopted to move away from the idea of measuring success as a target to achieve and instead define success more as a journey where trajectories could be evaluated and actions taken.

2. State of the Art

Within the field of Open Science and Open Access publishing there is debate about the need to measure success in ways that differ from the closed model of publishing. See [13–16] for some contributions to this debate. Traditionally, bibliographic metrics have been used to measure the success of publications and their impact, and these include aspects such as the number of citations, the impact factor of a journal, or the H-index of a scholar. However, in time, the scientific community started to look at alternative metrics to assess the value of academic publishing and a debate on how to measure success ensued. This debate goes back more than ten years and is tied to the advent of the digitalization of publishing and social media [17], which has shown the limits of measuring the impact of academic publishing based on only traditional bibliographic metrics (arguably originally developed before digital publishing). On one side, there has been a call to include metrics based on social media, for example the number of visualizations, downloads, or bookmarks for a publication. On the other hand, it is now recognized that there are a variety of outputs of science (beyond publications) which include, for example, datasets, but also self-publishing outputs (e.g., scientific blogs) and others. This has led to the definition of alternative metrics, or altmetrics, and the publication of a manifesto [18]. Authors have noted how altmetrics can provide an important contribution to the development of Open Science [18], for example, by widening the potential audiences or allowing the impact of a wider variety of outputs to be assessed. Authors have also pointed to the need to go beyond the idea of altmetrics, proposing concepts such as "open metrics" [19] and "next-generation metrics" [13]. This last proposition emphasizes that metrics cannot offer a "one size fits all solution" (p. 15) to measure the success of Open Science and offer a set of criteria for the development of responsible metrics for Open Science which include: (1) robustness (the use the best data available); (2) humility (quantitative measures should go alongside

qualitative ones and expert judgements); (3) transparency (ensuring that transparency of the data collection process for stakeholders); (4) diversity (ensuring metrics support plurality), and (5) reflexivity (assume that metrics can impact a project and reflect on them and update them).

Several organizations have waded in on this debate, for example, UNESCO which has promoted a recommendation on open science [20], with "a common definition, shared values, principles, and standards for open science at the international level and proposes a set of actions conducive to a fair and equitable operationalization of open science for all at the individual, institutional, national, regional, and international levels." (p. 5). Shared values identified by UNESCO include quality and integrity, collective benefits, equity and fairness, and diversity and inclusiveness. The Open Science Leadership Forum has proposed a set of success factors of Open Science, which include, amongst others, increased equity in research, better recognition of early career researchers or increased accountability of the research enterprise [21]. Similarly, the European Commission [22] has established specific initiatives on Open Science, including work on indicators and alternative metrics [23]. The document "Indicator frameworks for fostering open knowledge practices in science and scholarship" [24] defines a set of frameworks for establishing relevant indicators for success in Open Science. This document offers recommendations at four levels and includes novel infrastructures, capabilities, best practices, and incentives systems. The document also provides a comprehensive table of potential indicators for each of these actions, based on analysis from the previous literature.

We can see in the first two examples above (UNESCO and the Open Science Leadership Forum) the importance which is given to diversity, inclusion, and the collective benefit that Open Science can bring. However, we need to remark again that these propositions are tied with measuring success at a rather macro or global level (e.g., in a country) and not at the level of a specific project. In the third example (from the EU Commission), there is an emphasis on the need to also operate at levels which are not necessarily global to capture the diverse contexts in which Open Science is achieved; the four actions offer different levels of analysis and vary from the EU-wide level to the level of a single research institute. However, they still are operationalized at a level which is higher than that of a single project (such as GoTriple).

Overall, it has been recently argued [14] that "diversity, equity, and inclusion are key components of Open Science" and developing metrics for success must reflect this commitment. This surely also applies at the level of a single Open Science project. Indeed, all the above reflections and criteria are a useful frame for the definition of success metrics, and we will reflect on some of them later in the paper. Moreover, we need to remark that some of the propositions above (in particular altmetrics) are tied to Open Science related to publications and we should remember that GoTriple is not concerned with publishing. Nonetheless, what is relevant for us from the above discussion is: (1) the path to finding alternative approaches to measure success which are not necessarily hard-wired to KPIs as targets to meet and (2) that these measures should reflect, amongst others, issues such as equality, diversity, and inclusion. Moreover, it is clear that the importance of using available data and to reflect on the transparency of what is being measured at all levels (see in particular [13]).

In an albeit tangential area to Open Science, Open Source Software (OSS) projects have also grappled with the issue of defining metrics for success for quite some time. Whilst in this paper we are not concerned directly with the problem of evaluating the success of the GoTriple as an OSS, there are interesting insights that can be derived from the literature. Willinsky [25] noticed the potential convergence between Open Source and Open Science in several areas but especially "a common commitment to a larger public sphere". Over time, OSS projects had to define metrics of success going beyond those of traditional information system development. Moreover, the OSS efforts are also generally operating at a project level rather than at a macro-global level. There is somehow a parallel here with Open Science in terms of proposing some alternative metrics, but this is often a

localized endeavor. Whilst OSS has become more and more tied with corporate interests and managers still have to demonstrate things such as the return of investments [26], OSS is also based on relatively large communities of contributors (often volunteers) attached to specific projects.

That of the measure of success of OSS projects is indeed a debate with an even longer standing than that on Open Science and it dates to at least the early 2000s. For example, Crowston et al. offered one of the earlier accounts [27] recognizing that OSS is a form of software development which differs from traditional information system development. Therefore, "additional success measures that might be appropriate for OSS" (p. 1) need to be considered, alongside traditional metrics such as those proposed by, e.g., Seddon [28]. In OSS, while some metrics clearly have to do with the quality of the code or with the management of risks of a project, some other metrics relate to the quality of the management of heterogeneous communities, whilst other metrics deal directly with the success of a community and with aspects such as diversity and inclusion [29]. Diversity, inclusion, and collective action are therefore keywords that recur in both OSS and in Open Science. For example, the Linux Foundation has established a project called CHAOSS—Community Health Analytics Open Source Software [30], which places emphasis on, e.g., more traditional metrics related indeed with development and risk management but it also has metrics with a strong focus on equality, inclusivity, and diversity. As has been argued by Goggins et al. [31], this proposition supports a shift from looking at communities merely in terms of measures to considering instead what is defined as the "health" of a community and the related OSS project: "the potential of projects to continue developing and maintaining quality software" [32] (p. 31).

Another set of contributions needs to be mentioned in the context of this review in relation to research assessment. There is a lively discussion on reforming the way research is assessed, and this also ties with the potential development of metrics. Our goal is not to assess research or evaluate researchers but rather to assess the success of a platform for Open Science in SSH. Nonetheless, we should also learn from t literature. Efforts for reforming research assessment are being promoted by the Coalition for Advancing Research Assessment, also in collaboration with the EU Commission, which culminated with the publication in 2022 of the Agreement on Reforming Research Assessment [33]. The agreement includes high-level principles as well as practical commitments for reform. Amongst other things, the document emphasizes again a focus on diversity and inclusion, stresses the importance of also having also qualitative evaluation based on peer-review alongside quantitative indicators, and also to the need to limit the use of institutional rankings or measures such as the H-Index or impact factor alone. The document effectively traces a plan for the reform of research assessment and offers clear steps for this. Similar remarks were made by Gadd [34] with particular emphasis on empowering researchers, at the expenses of rankings, strict evaluations, or the measurement of what is not necessary. In this context, it is also important to mention the work of the HuMetricHSS initiative which in a white paper [35] has offered a set of recommendations to support a better alignment of institutional values (in, e.g., universities) with evaluation systems for researchers and impact, with a focus on reappointment, promotion, and tenure. This contribution is interesting because it focuses specifically on humanities and social sciences. The authors argue that "the values that institutions of higher education profess to care most deeply about—articulated through university mission statements, promotional materials, and talking points— are often not the values enacted in the policies and practices that shape academic life" [35] (p. 10). This requires a better re-alignment of values with the evaluation of researchers. For example, many institutes claim to support equality, diversity, and inclusion however achieving this for, e.g., tenure is quite complex. For instance, the focus on research excellence fosters elitism which translates into the perpetuation of inequality. Hence the formulation of recommendations to realign institutional values with the evaluation of researchers. These contributions are an example of the debate around research evaluation, however this is an area which does not concern GoTriple directly, since what we are interested in here

is identifying ways to measure success for an online platform and an online community. Research evaluation clearly is a task for institutes or specific national programs (such as the evaluation of universities carried out by ANVUR in Italy). Nonetheless this literature again points toward the importance of considering equality, diversity, and inclusion, whilst promoting the need for alternative forms of evaluation.

Lastly, it is also not surprising to see some overlap of what was mentioned earlier with the idea of responsible research and innovation (RRI). This is a policy and a framework of the EU Commission for the Horizon 2020 programme [36] aimed at considering the potential effect of science and innovation and can be seen as "the ongoing process of aligning research and innovation to societal values, needs, and expectations" [37] (p. 1). A common definition found on various European H2020 project websites states that RRI is "an approach that anticipates and assesses potential implications and societal expectations with regard to research and innovation, with the aim to foster the design of inclusive and sustainable research and innovation" [36]. RRI is anticipated to have an immediate effect on key areas that promote and monitor both the sustainability and the social justice/inclusion of any research project undertaken as part of European research. RRI is based on six principles which include: (1) governance; (2) ethics; (3) gender equality; (4) public engagement; (5) open access; (6) science education. Open Access, as we can see, is one component of this policy. However, more relevant to our discussion is the focus on equality, diversity, and inclusion. The Commission has also offered a set of indicators to measure the above six principles in practice [38]. These indicators operate at a high level, national, or supra-national (e.g., the European Union) and are not suitable to be adopted or adapted into metrics for any specific project. Some are also not relevant for our discussion. Nonetheless, these are examples of indicators that place emphasis on the gender balance or on the need to foster better social inclusion of a variety of stakeholders in science and innovation. Like in the previous discussion in Open Science, OSS, and research evaluation, diversity, inclusion, and equality recur as potential areas where success needs to be found—and somehow measured.

As is clear from the review of the literature, we are interested in the idea of embracing alternative ways of measuring success for an Open Science platform, looking at areas such as diversity, inclusion, or collaboration. Moreover, we are interested in moving away from hard-wired KPIs traditionally used to measure whether a project has met its objectives over time. We also need to identify measures that are modest and that can be located at the level of a single project rather than at a micro or even global level.

3. Materials and Methods

The approach and methodology we adopted to define the indicators of success for GoTriple can be seen as: (1) practice-led and (2) emergent. For Candy [10], practice-led research is "concerned with the nature of practice and leads to new knowledge that has operational significance for that practice". In practice-led approaches, artefacts play an important role in the creation of knowledge as they "function as a means of realizing a thing which has to be perceived, recognized and conceived, or understood" [11] (p. 159). Practice-led research is normally associated with creative and artistic practice, but in fact it parallels action research in the field of design, where it is far more established as an approach [39]. In practice-led research, new knowledge is generated through first the process of making things (e.g., a new design, a new creative artefact) and then second on reflecting on how these things should be, e.g., perceived or understood. It is therefore a two-pronged approach based on making first and reflecting second. Moreover, practice-led research can be, to an extent, distinguished from both the qualitative and quantitative research paradigms [40], as it locates the main outputs not in written texts (i.e., the scientific papers) using either numerical of qualitative data, but in practical outcomes. This mirrors our own work for GoTriple; the project team engaged in the practice of designing and developing the platform (the making), and then with the artefact at hand there is a process of documenting and reflecting on how the artefact (which includes its ideal user

community) could allow the generation of indicators to measure certain aspects of the artefact itself and what these indicators represent. The real output of this work therefore is not this manuscript per se, but it is the production of a set of indicators that can be used for the practical advancement of the platform. Therefore, the definition of the indicators follows the process of making the artefact (the platform in our case) and is the key output of the research presented here. In practice-led research, the role of reflection and documentation is fundamental [41], where documents (of any sort, whether they be diaries, notes, or anything else) about the practice support further reflection on the construction of the artefact and the knowledge that was generated. Through the making of GoTriple, several documents have been produced including deliverables, for example, as well as specific documents related to the process of defining the indicators which played an important role in the decision-making.

Our approach to the definition of success indicators can also be seen as an emergent process. We did not start with a specific theory or view of the world from which we derived assumptions and deductions. Rather, as we will see later, the definition of success indicators emerged through the practice-led process of making GoTriple as well as from internal discussion. This aligns with the idea that phenomena are not fixed by their specific properties but rather emerge from various social and material practices as discussed by DeLanda [42] (p. 4), according to whom "separate causes simply add or mix themselves in their joint effect, so that we can see their agency in action in that effect". Indeed, the indicators for success defined for GoTriple emerged as an outcome of several elements mixing together: the creation of GoTriple itself, the availability of certain data via the platform, the internal discussion amongst the project partners, the reflection of what is being carried out in the area of Open Science metrics (i.e., the literature), the existing project documents, and so on. Thus, the practice-led process of creating a new discovery platform and the blending of various emergent factors coalesced together in the creation of the GoTriple indicators for success.

The material we propose here is tied to the use of various sources of knowledge related to the project as well as internal discussion and decision-making operated by the TRIPLE project team. If one wanted to formalize and summarize the material we used for defining the indicators, one could see this as composed of a mix of: (1) knowledge generated from the participatory research conducted for the GoTriple design; (2) existing assumptions on the basis of the TRIPLE project; and (3) actions and decisions taken directly by the project team during various sessions of discussion and debate on the indicators. Moreover, as a fourth relevant point, we were able to rely on insights driven by the actual development of GoTriple. This last point is especially important in relation to the knowledge of data that would become available from the platform.

As stated earlier, the design of GoTriple is based on a user-centered approach. During the earlier stages of the project, we conducted qualitative interviews ($n = 25$) with SSH researchers across Europe (see [43] for the data, which is currently restricted). These researchers were representative of different disciplines, career levels, ways of working, and so on. While these interviews were conducted to gather the user needs for the design and not directly for the purpose of building success indicators, the analysis of these data (see Forbes et al. [44]) also offered some knowledge for fostering an internal discussion around the indicators. This includes, for example, observations around the struggle of early career scholars (for example in making their research emerge) or issues around the dominance of English as a publishing language at the expense of other languages. These interviews were conducted following ethical guidelines approved by the Ethical Committee of Abertay University in 2019. In any case, no excerpts of the interviews are shown in this paper, as the knowledge generated from the interviews served as basis for some decisions around the definition of the indicators.

The GoTriple platform has also been developed based on original assumptions. Some of these can be currently read on the project website (see https://project.gotriple.eu/about/, accessed 12 August 2022). That is, when we proposed our project to the funding agency,

we also offered a set of initial assumptions warranting funding support. For example, as stated earlier, GoTriple is a discovery platform for SSH. Much emphasis of the project has always been based on the multilingual aspects of the discovery. Moreover, the focus was on favoring a more collaborative approach amongst the SSH disciplines. Again, these assumptions have connections with inclusivity, diversity, and collaboration and have influenced the definition of our indicators.

The two above are elements of the materials that were given in a certain sense. In practical terms much of the work for the definition of the indicators has been the outcome of internal meetings of an ad hoc group of project partners. This group included amongst others, the project coordinator, the leaders of user research, the leader of communication and dissemination, the leader of the sustainability work. These discussions, specifically five meetings, took place online between March 2021 and June 2022 and an additional meeting took place in person, plus some one-to-one meetings between the coordinator and partners with specific knowledge, which ended with a final presentation of the indicators to the whole consortium in March 2022. A further three meetings took place afterwards to refine the indicators and evaluate their feasibility. Ad hoc discussions also took place with other partners, in particular with the technical team in order to assess the availability of the data and the definition of a long-term technical strategy (which could last beyond the project's funding). Documentation played an important role during and for the meetings. In each meeting, we took comprehensive notes and minutes about the various options that were evaluated, the decisions that were taken, the reasons for these decisions, and how the work for the preparation of the indicators should have been organized.

4. Results

In this section, we report the decision-making process for the definition of the compass indicators of GoTriple, alongside our understanding of what it means to achieve success in an Open Science discovery platform. To understand the process and the results, we have to put things into context and present what was achieved in the different meetings conducted, as well as comments on the various routes that were evaluated, and the choices made during our practice-led activities.

4.1. The Vague Notion of "Compass Indicators"

The first meeting for the definition of the success indicators took place on 23 March 2021. When the partners participating in the work met for the first time, there were originally different ideas on what the challenge at hand was. We were tied with a specific task of the project stating that we would "evaluate the success of TRIPLE solution in an iterative process every 7 months", but beyond this, the task description was not restricting us to a specific approach nor was it indicating specific measures or other metrics of sort to adopt. Whilst this loose definition of the task did not offer much guidance, it was also an opportunity to explore different options and ideas. Essentially the practical decision to be taken revolved around this simple question/problem: what to measure and how?

The first meeting took place as a form of brainstorming where different ideas and their merits were discussed among participants (seven in total from four different organizations, including the project coordinator, the lead of user research, the lead of platform sustainability, and the lead of platform interface design). In the beginning, two ideas were floating around on "what to measure", as potential starting points; on one side evaluating whether people/users were happy with the actual GoTriple platform and on the other side defining a set of KPIs such as the number of users, the number of documents available, and so on.

Discussion ensued on different aspects of these two starting points. Evaluating whether users are happy with the platform is something more akin to evaluating the usability [45] of the platform and not a measure of success as such. Moreover, in the project there was already work planned to conduct a full usability evaluation (also encompassing a final satisfaction questionnaire). Thus, we agreed that these were not the kind of success measures we were looking for and that this would lead to a duplication of work. On

KPIs, discussion ensued on two problems: (1) how the definition of indicators for success would duplicate other KPIs which were required anyway for the funding agency and (2) whether KPIs (even if monitored over time) would indeed be the best way to capture the success of GoTriple. For example, while the overall number of users registered on the platform could be considered a good measure for success as well as a KPI, it is also a very generalist measure which tells little, for example, about the composition of the user base, and if the project is achieving success just beyond increasing numbers, for example, around diversity or collaboration. We started therefore reflecting on the need to have dynamic measures/indicators, the sort of compasses that would allow us to understand how the project was achieving success and also those that would serve as an indication of where we could take actions now and in the future.

This last consideration is connected with the practice-led approach of generating knowledge which has operational capacity for decision-making. For example, what if in building the user base we found that certain European countries are underrepresented or not represented at all? This would then arguably require the team to take some actions for promoting the platform more in the under-represented country. Then, we agreed to abandon the idea of KPIs as targets to meet and began to adopt this, a notion of "compass indicators". That is, indicators whose main objective is to give us a sense of the direction of where the platform and its community are going, but also indicators that would help us take some actions to change the direction. Essentially the main achievement of the first meeting had been the proposition to adopt this, yet still vague, notion of compass indicators.

In parallel, we also discussed the problem of "how to measure". At this stage we still were not in the position to have identified what to measure, and therefore the "how" problem was still very loosely approachable. However, something important was clarified. Since the work with the users for the design of GoTriple had been tied with conducting largely qualitative data collection (e.g., interviews, codesign), an initial proposition was to gather evidence for success again with a similar process. However, discussion ensued on whether this was the best approach. On one side, we had already been demanding a lot from our potential user base, and involving them in further interviews might just increase their burden. On the other side, gathering primary data in this way was seen as lacking sustainability. In other words, running questionnaires or interviews is time-consuming and requires significant resources, and in the long term (considering the potential end of funding), this was considered not sustainable. Thus, a proposition was discussed to try to obtain measures based on data generated directly by the GoTriple platform. For example, if we wanted to understand the representation in the user base of the different European countries, we could simply try to obtain this data directly from the GoTriple database if the information was available or use some proxy measures that could provide an indication of this. This proposition of reusing the platform data has some significant advantages, as it would: (1) reduce the burden on the collection of data; (2) allow the direct reuse of data already potentially held in our database; (3) support the process of automation of the measurements; and (4) facilitate the creation of visualization offering an immediate view, rather than waiting for the process of analysis of primary data. It is clear then that our decision to adopt quantitative measures of success over qualitative ones has been dictated by practical reasons tied with the availability of resources in the future, after the end of funding. Qualitative measures do require a lot of resources for data collection and analysis. Quantitative measures based on existing data from the platform do require an initial investment in automating the process, but then in the long term, they can be run with limited resources.

4.2. Preliminary Brainstorming on the Compass Indicators

A second meeting took place on the 11th of May 2021, with a restricted number of participants (three in total, with members from the coordinator team and the lead of user

research present). At this stage, having agreed on the notion of "compass indicators", the next task was to start defining what these compasses were in more detail.

The first part of the discussion concentrated on brainstorming about potential indicators and broad areas/categories of measure. A good discussion took place in relation to the **diversity** of the GoTriple user community, in particular with the proposal to consider the following indicators:

- The balance of female/male users
- Good coverage of all European countries in the user base
- Good balance at career levels
- Good balance of age levels
- Good representation at discipline levels

Some of these indicators were inspired by the user research conducted in the early stages of the design. Through interviewing potential users, we often heard, for example, stories of the struggles that early career researchers face in order to emerge, the difficulties that older scholars face when dealing with new technologies and of the importance that the SSH scholars place on the definitions of disciplines. Here, we see a further clear connection between the practice-led approach of making a new artefact (i.e., the outcomes of the user-centered design of GoTriple) and the process of making sense of it. The discussion on diversity also clearly connects with what is seen in the literature review about alternative metrics for, e.g., Open Science, which have a strong focus on diversity and inclusivity.

The team agreed that the goal was not necessarily to achieve a certain balance or representation from the outset, for example a perfect balance between female and male users but to use the indicator to assess the current balance and take action in case of a clear unhealthy unbalance. The same goes for the other indicators. For example, GoTriple is aimed at all SSH scholars at all career levels. If for instance we had observed that we had a very low number, of e.g., PhD students enrolled versus people at the top of their career (e.g., professors), then this could give us a stimulus to look more into this, understand why, and then foster better and tailored communication for the under-represented category of users and achieve better diversity. It thus is clear that the "compass" is not a hard-wired KPI, where achieving a certain balance is a measurable objective of a performance in itself. Instead, it is a measure of the current situation, providing knowledge to the platform managers on where and potentially how to take actions.

Beyond diversity, during the second meeting, we also discussed other indicators, initially under the label of "**technical indicators**"; later (see next section) this label had been changed to better reflect our goals. The initial proposed indicators under this label included:

- The number of documents available in each of the 10 languages
- The findability of documents in "rarer" languages

These two potential indicators are reflective of the multilingual nature of GoTriple. In addition, we did not speak of balance here, as clearly it is not possible to have a balance of documents in different languages. However, the idea was to have an indication of representation for each of the languages supported by the platform. Moreover, there is an emphasis on making sure that languages that perhaps have less strength in terms of, e.g., publications (e.g., Croatian), are represented sufficiently in the material discoverable through GoTriple.

A third area discussed was related to the label **collaboration**. Indeed, one of the goals of GoTriple is to foster collaboration across SSH and make sure that the users can do things together, across disciplines. The following potential indicators were initially discussed:

- The number and types of interactions between users
- Whether there is a dense relationship or not, based on a network analysis
- The number of times the user looks at other profiles
- The use of innovative services

Some of these indicators still appeared rather vague during the meeting's discussion. For example, the number and types of interactions was just a signpost we could use to better figure out what kind of interactions are possible via GoTriple. For example, sending a direct message to another user could be considered as a meaningful type of collaboration, as a user is trying to contact another user for a purpose. The indicator of "dense relationships" instead represents a potential measure of whether interactions on the platform are sustained and happen frequently, or if they are instead just a one-off event. However, this indicator was later abandoned as it would require further analysis, and it is thus not currently included in the current indicators. The number of times users look at profiles is an indicator that there is interest in looking at other people and what they are doing. Finally, the fourth indicator is related to the use of the GoTriple innovative services. Some of these services have their own user registration (separated from the platform registration), whilst others are embedded directly in the platform (such as the recommender system) and thus measuring the use of the innovative services would tell us which services are providing added value to our user community.

While much of this second meeting was devoted to fleshing out some initial potential indicators, a brief discussion also ensued on "how" to measure some of them. For example, some indicators could be measured through data collected at the moment of the user registration process or from information that could become available from the GoTriple user profiles. For example, through the completion of their GoTriple profiles (see an example in Figure 3), users could insert information about themselves, including, e.g., their title, position, gender, and so on. Other indicators (in particular those related with publications) could be measured from the metadata available from the Isidore search engine, which include, e.g., the discipline or the language of a publication. Moreover, other information could be retrieved from the web analytics tools of GoTriple. These analytics tools have been set up for the traditional purposes of monitoring the use of the platform, but clearly, they can also provide data relevant for measuring the compass indicators.

4.3. Refining the Compassess and How to Take Action

Having provided an initial definition of potential compass indicators, the team started to reflect on how to improve them, if there were more indicators to consider and if some of the already proposed indicators could be abandoned (either because they were not very meaningful or too difficult to measure). Discussion in the subsequent three meetings allowed the team to refine the broad categories of the compass indicators. In particular, one of these meetings entailed an intense discussion with the platform developers in order to have a clear idea of what could and could not be measured.

The first essential outcome of these meeting has been the definition of the final four broad categories of the indicators, as follows:

- Diversity and inclusivity
- Platform usage
- Collaboration
- User experience quality

The broad categories of "**diversity and inclusivity**" and "**collaboration**" were kept from the previous discussion. Moreover, the group decided that what were earlier defined as "technical indicators" were not reflective of the strong user dimension of GoTriple. A decision was taken to rename this as "**platform usage**". A decision was further taken to include a fourth broad category, mostly to measure the user retention (that is a measure of the continued use of the platform), and this was called "**User experience quality**".
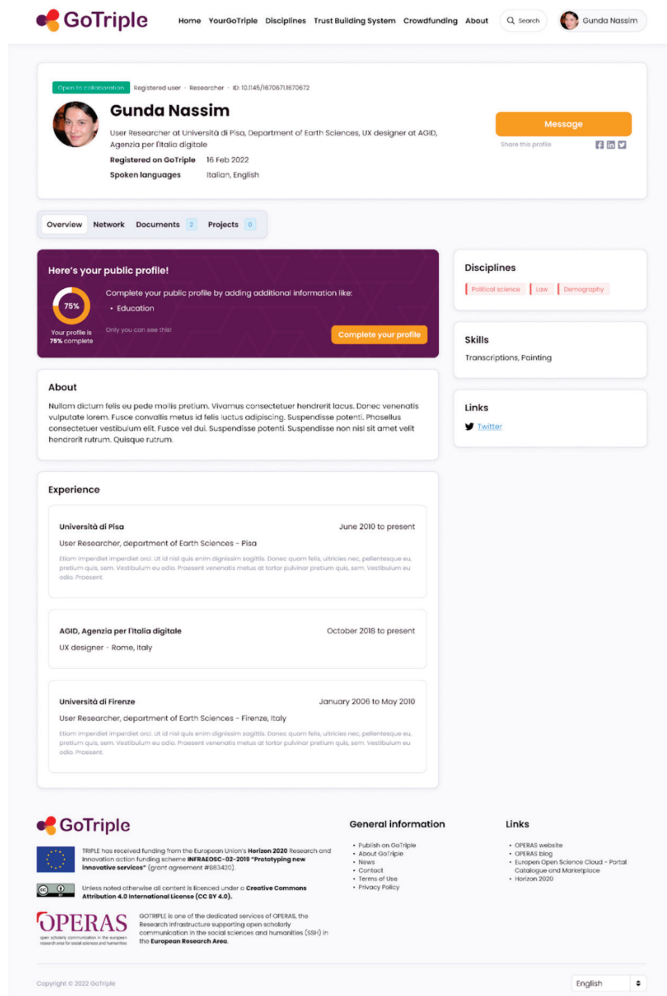
Figure 3. Example of a completed user profile on GoTriple.

With the broad categories at hand, detailed compass indicators have then been identified for each of them. Table A1 in the Appendix A shows all the compass indicators defined for GoTriple. Some indicators, as we have discussed previously, had already been defined in the initial meetings (and were at this stage refined); others were further proposed and accepted. We will comment now on some of these indicators, to give an idea of the underlying thinking. Readers can consult Table A1 for the complete list.

For example, the "user type" compass indicator under the "diversity and inclusivity" category captures whether the user is a researcher/academic or another category of user such as a journalist, an individual from the private sector, or an individual from an NGO, for example. This indicator could indeed further enhance the understanding of the diversity in the user base and tell us if GoTriple has managed to entice users beyond research and academia. This information (the type of user) is provided by the user at the time of registration (see Figure 4). Additionally, for another example, we decided to adopt the indicator "number of followed/following" within the collaboration category. This can tell us something about how the network develops on the platform and how users are engaging with others. It indicates the activity of interaction among users.
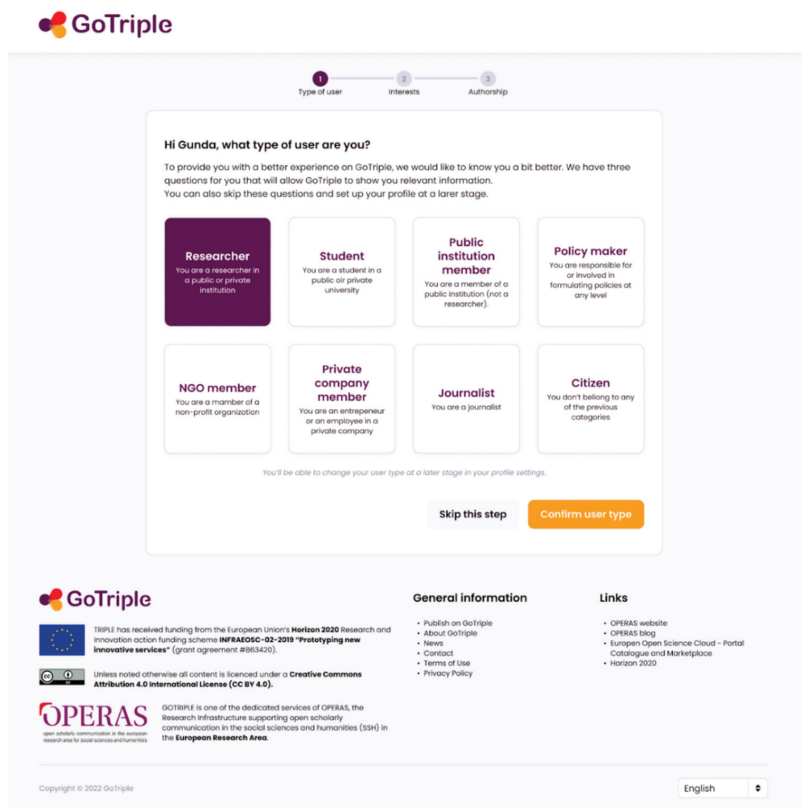
Figure 4. Step of the GoTriple registration process, where users are asked what "type" of user they are (researchers or another type).

Alongside the definition of the specific indicators, a discussion took place in relation to the "type of action" that could be taken based on the knowledge gathered from the compass indicators and also "how" actions could be taken to modify/improve the situation. We should remember again that these indicators were developed with the declared goal to tell us the "direction of travel" of GoTriple and to offer knowledge to take actions to steer this direction differently where necessary. It is worth commenting on some of these elements to provide clarity on the undying thinking.

For example, in the category of diversity and inclusivity, for the indicator "gender" what we are seeking to measure is the balance in the user base of female and male user. This ideally will be carried out using some baseline data to compare the data, i.e., by using baseline data of the gender balance of SSH in Europe. Should the indicator offer a picture of an imbalance, then direct action will be taken to promote the platform more to, e.g., male scholars, with targeted advertising on social media for example. Indeed, if the indicator is very different from the EU SSH gender balance, it means the platform is not inclusive enough on this specific aspect, and we should carry out some research to understand why, before acting. Similar to this is the plan for the indicator "location" (that is the location of the user, i.e., the city where the user is based). This is a proxy for us to understand in which country a user is based (e.g., a user with the location Rome could be assumed to be working in Italy) and this is information that the user could include in their profile. This information can tell us about the spread of the user base across Europe, if GoTriple has coverage of all the countries, and if there is reasonable representation. Again, this indicator could then be compared with a potential baseline (if available), e.g., the number of SSH

scholars in each country, to better understand the representation. Should the indicator tell us that there are countries which are under-represented, then actions could be taken with promotional activities such as mailing lists or events in the local/national language. Variation in the time of the actions could also help the team understand the effect of the promotional activities.

In the category of "platform usage", particular attention was applied to assess the multilingual aspect of GoTriple. For example, the "document language" indicator captures whether the search function facilitates discovery of documents in all the languages supported. If this is not the case (compared to what should be the expected numbers), then action could be taken, for example in the form of increasing the number of Open Access repositories harvested by GoTriple in the under-represented languages. The expected number is given by a measure of the languages of documents given by regular searches and the number of documents that the user clicks on (on this second aspect, it is possible with an analytics tool to track the language of the clicked document). This would then allow for better coverage of the under-represented languages and ideally increase the number of publications available in that said languages.

Similarly, for the category "collaboration", the compass indicator "profiles clicks" measures the access of users to other users' profiles. This is seen as a proxy to measure the fact that users are trying to establish potential connections, discover other people to collaborate with, and ideally decide to make contact with them. Through the platform analytics, it is possible to measure the number of profiles clicks and also see what kind of profiles are the most clicked. Should we assess that the profiles as not being used enough or that only some categories of users are receiving relevant clicks whilst others lag behind, then again, some forms of action could be taken. Possible actions include: (1) making the profiles more visible in the search results; (2) making the profiles stand out more on the home page; (3) improve the quality of the profile page and data; (4) and increase the number of profiles.

Overall, from these examples, we see the approach that was adopted: each compass indicator is accompanied by a definition of what kind of actions could/should be taken in order to improve the indicator, as well as how this will be achieved with a definition of activities.

4.4. Automation and Dashboard

During the last meetings, discussions also ensued on the actual process of gathering the measures (i.e., the data) for the compass indicators. The initial idea was to have an ad hoc process with the measures essentially collected manually from the GoTriple database (roughly every 7 months, as this was the period stipulated in the project task to report the results). Some participants at the meetings voiced, however, considerations on the lack of sustainability for this approach and they also drew attention to the labor intensity of a manual ad hoc process. The discussion then pivoted on the merits of instead having a fully automated process. Meanwhile, as we briefly discussed earlier, some indicators could be obtained directly through the GoTriple analytics tools (and could hence already be automated), while others would necessarily require access to information held in the GoTriple database. To make the measurement of compass indicators sustainable over time, the team therefore agreed to implement a process of automating the gathering of the indicator data, in particular, those outside the analytics and related to database information. Automating the process was seen as having some advantages. Firstly, whilst it would require an initial investment of time, this investment would however allow for the better sustainability of the compass indicators in the long term, since the process would then be in place and functioning with minimal maintenance and/or resources. Secondly, automating the process could allow the team to use these indicators to create visualizations. This would allow one or more dashboards to be built, all the indicators to be gathered in one place and for the use visualizations to derive potential insights for supporting decision-making. Moreover, the internal discussion also highlighted the potential opportunity to bring back

some of these visualizations directly within the GoTriple platform and to the users. Indeed, a public dashboard with some relevant indicators could be realized showcasing how the platform is doing in certain areas, such as the country balance, or the balance of languages and so on. This will also have the advantage of sharing some of the compass indicators with the users and, as a consequence, potentially empower the user community to take direct action. Currently an initial version of the dashboard is available only for internal use (see Figure 5), covering a limited set of indicators, in particular those gathering data that are available from the analytics tools. For building other indicators with the data from the GoTriple database, additional work will be necessary, with the need to create ad hoc Python scripts as well as tailored visualizations. At the time of writing, work has started to implement these scripts and we expect to have initial working versions of all these indicators by early 2023. Additionally of note is that the user engagement and onboarding of GoTriple are also planned to take place from December 2022 and therefore most of the defined indicators cannot be measured as of yet and require the building and presence of a community of registered users.
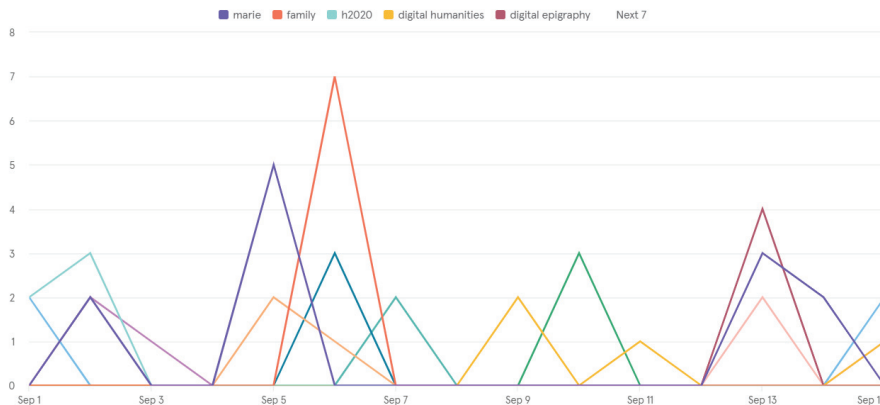

Figure 5. Example of an initial dashboard concept based on existing data retrieved from the GoTriple platform, displaying the key search terms for the period 1–15 September 2022.

5. Discussion

There is consensus that measuring the success of Open Science initiatives and platforms requires moving away from traditional bibliographic metrics to embrace alternative metrics, encompassing aspects such as inclusivity or sustainability. There is indeed debate on how to develop metrics that are appropriate to capture the specific nature of Open Science, aimed at building better scientific cooperation and wider access to knowledge through digital technologies. In this paper, we proposed the concept of **compass indicators**, as a potential avenue to understand success in a localized Open Science platform and community. This concept is the outcome of the practice-led work of designing and developing a discovery platform (GoTriple) for SSH in Europe. Compass indicators could be seen as modest indicators, as they do not seek to target performance necessarily, but rather offer insights into the direction of travel of a project. We also need to remember that in practice-led research, new knowledge is generated through the process of making things. We have been involved in the production of GoTriple and the need to measure its success connects to the task of reflecting on how GoTriple is used and understood. Therefore, the concept of compass indicators is not something derived from theory assumptions but is an emergent component of the practice of making GoTriple. Things come together (such as the interface design, the software, the data collection, the building of a user base, the meetings of the project team, etc.) whereby separate elements of the practice-led work for

making a platform "mix themselves in their joint effect, so that we can see their agency in action in that effect". [42] (p. 4).

We can now formally define compass indicators as indicators that can tell the "direction of travel" of a project (in this case an Open Science discovery platform) and that allow the promoters of such a project to take actions in order to steer or change this direction. They are not therefore a measure of performance but a measure of direction. For instance, GoTriple is a European platform for SSH research and as such it strives to achieve diversity and inclusion which may require, for example, having a good balance of representation for scholars at different stages of their career, having a representation of all European countries in the user base or having a good representation of documents in the different languages supported by the platform. However, success is not in the achievement of, e.g., a target balance or representation, but rather on the understanding of how the project is doing and whether some actions need to be taken. A compass indicator effectively: (1) tells us what the current situation is (e.g., the current representation of countries in the user base); (2) signals the project team to whether some action may need to be taken (e.g., an assessment that there is under-representation of certain countries, based on existing knowledge about the number of SSH scholars in each country); (3) proposes some potential actions that can be taken to improve the situation and steer "the direction of travel" of the project (e.g., run communication campaigns aimed at the under-represented countries, for example in the national language, to enroll more users); and (4) allows to re-evaluate the effects of the action on the indicators after some time and assess if further actions are needed. Achieving, e.g., a balance or representation (of e.g., countries), is not therefore the ultimate goal; the goal is to travel toward that balance.

It has been a clear outcome of our practice to detach success from the idea of hard-wired targets that need to be achieved and met (such as KPIs). Of course, the aim is still to, e.g., achieve a good balance of inclusivity and diversity in several areas, or, e.g., have a balanced platform usage, but the direction of travel is more important than the achievement of a measure, even if there is an imbalance in the first place this is not seen as something unsuccessful. Success in itself is the realization that, in some areas, more work is required to improve things such as the platform inclusivity/diversity, the usage, the collaboration, or the user experience quality. Compass indicators therefore put a different light on the meaning of success and see this as something dynamic, subject to changes and ideally improvements. Therefore, the "compass" label in our concept offers a strong analogy which serves as a reminder that success is most likely a journey, and indicators are the signposts of this journey.

It is useful to remember that in their proposition for next-generation metrics for Open Science, Wildson et al. [13] proposed a set of criteria for generating robust metrics: (1) robustness; (2) humility; (3) transparency; (4) diversity, and (5) reflexivity. Our proposition for compass indicators embraces most of these criteria. We have seen how we seek to obtain the data that are available directly from the platform in order to support a robust process of data gathering (via automation and visualization). Whilst the indicators we propose are largely quantitative, expert judgement will be exercised on them by the project team in order to make decisions for action. Moreover, initial knowledge for some indicators was derived by qualitative research conducted for the platform design. We plan to exercise transparency by bringing back the indicators to the GoTriple users via a public dashboard with the intent to empower the users in understanding how the project is doing in areas such as, e.g., inclusivity. Much emphasis on the definition of our indicators was placed on diversity and inclusion, especially around the composition of the user base and the language representation. Compass indicators more than targets are providing material to the project team as well as the platform users to reflect on the trajectory of the platform and to offer insights to take actions.

We have also seen that much of the existing propositions around metrics for Open Science take a rather global or macro approach for the definition of metrics (e.g., [20,24]). However, these propositions offer little insight on how to measure the success of Open

Science at the level of a single, perhaps small, project, where the health of a user community is a primary concern and where the achievement of Open Science is a rather situated endeavor. Compass indicators have a localized nature and are again more modest than macro indicators. Nonetheless, they serve an important function, which is supporting project teams in their decision-making. In this sense, they are more similar to the processes of measuring the health of Open Source communities [32], as they are targeted at a specific project/community and their merit exists within that context.

It is also important to recognize the debate about the adoption of qualitative indicators for Open Science, e.g., [13,14], alongside quantitative ones, which was briefly mentioned in the introduction. Small projects such as GoTriple have to consider the limits of their resources. Therefore, our position in this debate is that practical considerations should guide what kind of indicators should be adopted, rather than, e.g., epistemological reasons. In our case, we agreed for numerical/quantitative indicators as a way of offsetting the cost in terms of resources for maintaining our measures in the long term. The approach we propose has the advantage of allowing to us to conduct measures with very limited resources. On the other hand, maintaining qualitative indicators would require substantial research work for gathering the data on a continuous basis. This is something which was not seen as sustainable in the long term for us.

Overall, there are some lessons that can be learned from this experience for other Open Science digital platforms which focus on offering services to end-users. The first is to adopt a modest approach to measure success, looking at this as a journey and not as a hard measure to be achieved. It is therefore not a matter of performance but of reflection on the evolution of a project and its contribution to Open Science. The second lesson is to always keep the user at the center, seeking to develop measures that can support both an understanding of the user community and that can be brought back to the community to foster a dialogue with them. The third lesson is to reuse existing data as much as possible. This has advantages in terms of the sustainability of the process as well as supporting better reflexivity on the current evolution of the user community. The reuse of existing data allows further reflection but also restitution to the user for example with visualizations.

6. Conclusions

What was described in this paper is the process and the negotiations that led the GoTriple project team to the emergent definition of the compass indicators. GoTriple was fully released in September 2022, and from December 2022 this will be followed by a user engagement strategy which aims to build a thriving community of users around the platform. It is at this stage that the compass indicators will start to play their role, as they will be used to measure the status of the nascent GoTriple community and its evolution (i.e., in terms of diversity, collaboration, etc.) over time. The next steps of this work will be to complete the automation of the data collection from the GoTriple database with ad hoc scripts and the creation of the associated dashboards as detailed earlier. Moreover, for the evaluation of some of the indicators, we assumed we could compare our measures with some European benchmarks or baseline data. At the time of writing, we have not yet drawn up a possible list of these benchmarks and this is one of the next significant tasks for our work. This might require additional considerations, as, for example, baseline data may not be available at all, and if available they may not exist just for SSH. We would also need to consider whether to adopt an EU-wide approach or not. For example the "She Figures" report from the EU Commission [46] gives aggregate figures at national and the EU-28 level of the proportion of women in research, including a separate breakdown for humanities and social sciences. This provides a useful baseline for our gender indicator. However, there are significant differences across countries, and in this case, working with the EU-28 aggregate figure may seem appropriate. For other indicators, however, finding a baseline figure may be more complicated and will require careful scrutiny of the existing literature and reports. For example, this may apply for the indicator of the community composition at the career level and what a healthy breakdown of our community for this

indicator should consequently look like. A similar consideration could be made for the baseline data about the distribution of users across countries. In this case, one could consider gathering data about the total number of SSH scholars in each country, drawing some ratios, and using these ratios as a baseline. However, it is not obvious that these specific data exist and if they exist, they may be in national reports (rather than in EU-wide reports), which would increase the difficulties in building the baseline.

In conclusion, this paper has proposed the concept of compass indicators as a potential avenue to measure success in a small Open Science research project and community. We assume success not as a performance or targets to be achieved, but as a journey to be undertaken. Compass indicators are thus the signposts of this journey and indicate the direction of travel of an Open Science project, such as GoTriple. We presented the practice-led work undertaken for the definition of the concept. Further work will be carried out at a later stage to assess the value of these indicators and how they can support a thriving Open Science project. With an initial user base and the technical instruments in place, then the monitoring of the compass indicators will start. We will then strive to assess the value of these indicators, but also to report back the results in a follow-on publication.

Author Contributions: S.D.P., writing—original draft preparation; conceptualization; and methodology. E.B., writing—review and editing; conceptualization; and methodology; P.F., writing—review and editing; and conceptualization; S.A.-R., writing—review and editing; and conceptualization. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The interview data mentioned in this paper were deposited in Zenodo on 3 December 2021 with restricted access https://zenodo.org/record/5752095#.YwX1vXbMI2w.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. List of GoTriple Compass Indicators with full description.

The GoTriple Compass Indicators		
Title	**Indicator**	**Type of Action and How**
1st Category: Indicators about Diversity/Inclusivity		
Gender	Balance of female/male users	If the indicator is significantly different from the EU SSH gender balance (our potential baseline), it means the platform might not be sufficiently inclusive. We should then conduct some initial research to understand the underlying issues. Afterward, we should put in place some activities to improve the balance, such as better promotion of the platform or modifications to the design where necessary.
Location	Coverage of all European countries in the user base	Should the indicator signal that there is under-representation of certain countries (based on potential baseline data of the number of SSH scholars in Europe), then action should be undertaken. The focus will be put on the less covered countries, with promotional activities covering these potential channels: mailing lists, events, and social activities (e.g., workshops/webinars in the local language). Variation in time of the indicator can help the team understand the effect of the promotional activities.

Table A1. *Cont.*

The GoTriple Compass Indicators		
Title	**Indicator**	**Type of Action and How**
User type	The type of user (academic researcher, individual from the private sector, NGO, someone interested in citizen science, policy maker, etc.)	The approach here could be to run experimental campaigns to involve specific groups of users (i.e., journalists) regardless of the initial metrics, and see if we can detect any improvement in the on-boarding of users from the specific user group. Percentage of variation in time can tell us about the direction of the indicator. Possible actions to improve this indicator include: - Changing the communication on various channels - Selecting different channels of communication to non-researchers - Making the platform more usable for non-researchers
User disciplines	Representation at disciplinary level	Should we detect from the indicator a significant under-representation of certain disciplines in the user base, we could focus some marketing activities on the less present disciplines. (We must compare the percentage of users for each discipline with the percentage of documents. We expect to have a relationship between the two: if for a discipline there are few documents, we cannot expect to have many users)
Search terms	The interest in words, i.e., search terms used for the discovery on the platform. Interest in what language is used during the discovery activities	Even if not an actionable metric, this can be of some interest for the SSH community to know the most popular arguments and search terms. We could also breakdown these data with location, type of user, and disciplines of interests. These data will also be displayed in the public GoTriple dashboard.
Docs claiming	What percentage of registered users have claimed at least one document?	Low percentage values might indicate that: - The users do not really see the benefit of having a completed profile. We should then work to make these benefits more evident or add more benefits. - The UX of claiming a document is not clear enough, which will require looking into the platform interface and potentially make some changes - We should stress the importance and relevance of the claiming documents process more to the users (like Academia and ResearchGate do so, with regular emails)
Docs claimed vs. proposed	Number of claimed documents vs. number of documents whose authorship has been suggested by the platform	Since it is difficult to set a referral percentage here, we should work on experiments to try to raise the value. Actions could be similar to the "docs claiming" indicator and require stressing more the relevance of the claiming documents process for the completion of a user profile.

Table A1. *Cont.*

The GoTriple Compass Indicators		
Title	**Indicator**	**Type of Action and How**
2nd Category: Indicators about Platform Usage		
Search language efficacy	Efficacy of the language search of publications	This is like a funnel with three steps: Step 1. In the Results page list, the user clicks on a language in the language facet Step 2. The search results are updated Step 3. The user clicks on a result (aspect to track: language of the document clicked) The factor to investigate is the percentage of found documents (step 3) which have the same language selected by the user (step 1). We could observe the conversion rate of the funnel for a given time and then complete some modifications to the interface to see if we have improved it or not. Running A/B testing is also a further possibility with different designs, looking at which one delivers more search language efficacy.
Documents' language	Findable documents in all the languages	This information can be retrieved directly from the GoTriple database, which can indicate the number of documents for each language. If some languages are less present (compared to the expected values), then we should consider adding to GoTriple repositories that provide documents in the under-represented languages.
Recommender	Recommender system: Users clicking on the recommendations offered to them by the platform	Since we do not have a referral value for this indicator (i.e., something that tells us the value of the recommendations for a user, especially one which is not registered to the platform), we can run campaigns to improve it. Possible actions to improve the indicator: - The Recommender system integration and visibility could be improved - The Recommender algorithm itself could be improved in order to offer more targeted recommendations
TBS	Trust building system use and uptake of the system by the GoTriple users	Possible actions to improve the indicator should the use and uptake not be satisfactory: - TBS integration and visibility could be improved (maybe users do not see value in using the TBS) - Overall, the TBS experience and functionalities could be improved
Pundit	PUNDIT Annotation tool use and uptake	Possible actions to improve the indicator should the use and uptake not be satisfactory: - Pundit integration and visibility could be improved on the interface - Overall, the Pundit experience and functionalities could be improved

Table A1. *Cont.*

The GoTriple Compass Indicators		
Title	**Indicator**	**Type of Action and How**
Crowdfunding	Crowdfunding (based on the Wemakeit platform)	Possible actions to improve the indicator should the use and uptake not be satisfactory: - Wemakeit integration and visibility could be improved on the interface - The CrowdFunding projects presentation is not effective enough; we should provide guidance on how to present a project - Projects promotion/marketing is not effective enough: we should provide guidance on how to promote a project
Visualizations	Discovery/visualization service: Button from the platform to connect	Possible actions to improve the indicator should the use and uptake not be satisfactory: - Increasing the promotion of visualizations on the user interface - Improving the visibility of the visualizations in the interface
Searches	Search engine: Number of searches, where we should expect the real growth of the platform and probably more advanced features	Possible actions to improve the indicator should the use and uptake not be satisfactory: - Making the search engine more visible in the header of the platform (in the home page is probably already visible) - Considering adding an autocomplete - If the results are not satisfying, actions should defined. Some research is probably needed here, and probably the quality of documents should be improved
3rd Category: Indicators about Collaboration		
Profiles clicks	Number of visualizations of profiles	Should the visualization of profiles not be satisfactory, we could run campaigns to improve the reach of profiles on GoTriple. Possible actions to improve the indicator include: - Making the profiles more visible in the search results - Making the profiles standout on the home page - Improving the quality of the profile page and data - Increasing the number of profiles (more actions to engage users to add more profiles)
Follows	Number of Follows	Possible actions to improve metrics should the follow button not be used enough: - Making the "Follow" action more visible - Adding the "Follow" action in other contexts in the UI - Making the users better understand why they should follow their peers - Adding more benefits from following another user - More explicit calls to action related to the recommender system "Profiles you should follow"

Table A1. *Cont.*

The GoTriple Compass Indicators		
Title	**Indicator**	**Type of Action and How**
Messages	Number of messages between users and breakdown of other dimensions such as type, location, and disciplines.	Possible actions to improve this indicator should we see not enough messaging taking place between users: - Making the "Message" action more visible - Adding the "Message" action in other contexts in the UI - Explaining to users why they should get in touch with other users - Showing the users that there is a "community" using a messaging system and that they should join in
Status	Current use of the status in the profile (e.g., open for collaboration)	Possible actions to improve metrics: - Users do not know how to set the status - Users do not see the benefit of setting the status - Make the status statistics more visible in the interface (such as a chart saying X% of GoTriple users are willing to collaborate)
Registered users	Registered users	Should we see users coming to the platform but not registering, possible actions to improve this indicator include: - If the users do not see the benefit of being registered, we should show them the benefits by highlighting actions that require registration and adding a registration wall (with lazy registration) - If the registration journey is too difficult, research would be required to understand why - Create targeted campaigns to make users register
4th category: Indicators on Quality (Retention) "User experience quality"		
Quick survey	Offer a short survey asking: "Did you find what you were looking for?"(There would need to be a pop-up question asked to the user)Other questions may be asked in following campaigns	Possible actions to improve this indicator if the short survey results are not positive: - Improving the search function in the UX/UI - If the results are not satisfactory: actions to be defined, some research is probably needed here to investigate the issue more closely, and quality of documents should be improved
Retention	Considering the type of usage of the platform we expect users to keep coming back to GoTriple on a monthly basis	Possible actions to improve this indicator if retention is not satisfactory: - Improving the overall UX/UI - Conducting user research to understand the main issues with a pool of users

References

1. Achenbach, K.; Błaszczyńska, M.; De Paoli, S.; Di Donato, F.; Dumouchel, S.; Forbes, P.J.; Kracker, P.; Vignoli, M. Defining discovery: Is Google Scholar a discovery platform? An essay on the need for a new approach to scholarly discovery. *Open Res. Eur.* **2022**, *2*, 28. [CrossRef]
2. OECD. *Making Open Science a Reality. Science, Technology and Industry Policy Papers*; OECD Publishing: Paris, France, 2015; Volume 25.
3. Dumouchel, S.; Blotière, E.; Breitfuss, G.; Chen, Y.; Donato, F.D.; Eskevich, M.; Forbes, P.; Georgiadis, H.; Gingold, A.; Gorgaini, E.; et al. GOTRIPLE: A User-Centric Process to Develop a Discovery Platform. *Information* **2020**, *11*, 563. [CrossRef]
4. Mounier, P.; Ferwerda, E.; Dumouchel, S.; Gatti, R.; Gingold, A.; Radovic, R.; Smaniotto, A.; Stojanovski, J.; de Vries, S.; Waaijers, L. OPERAS Platforms and Services White Paper. *Zenodo* **2018**. [CrossRef]

5. Dallas, C.; Chatzidiakou, N.; Benardou, A.; Bender, M.; Berra, A.; Clivaz, C.; Cunningham, J.; Dabek, M.; Garrido, P.; Gonzalez-Blanco, E.; et al. European Survey on Scholarly Practices and Digital Needs in the Arts and Humanities. Project Report. DARIAH; DIMPO. *OpenAIRE* **2017**. [CrossRef]
6. Pouyllau, S.; Mine, J.l.; Capelli, L.; Bunel, M.; Sauret, N.; Busonera, P.; Desseigne, A.; Baude, O. ISIDORE celebrates its 10th anniversary. *Huma-Num.* **2021**. [CrossRef]
7. Mounier, P. Publication favela or bibliodiversity? Open access publishing viewed from a European perspective. *Learn. Publ.* **2018**, *31*, 299–305. [CrossRef]
8. Fitz-Gibbon, C.T. (Ed.) *Performance Indicators*; Multilingual Matters: Clevedon, UK, 1990.
9. Parmenter, D. *Key Performance Indicators: Developing, Implementing, and Using Winning KPIs*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
10. Bishop, D.A. Key performance indicators: Ideation to creation. *IEEE Eng. Manag. Rev.* **2018**, *46*, 13–15. [CrossRef]
11. Candy, L. Practice based research: A guide. *CCS Rep.* **2006**, *1*, 1–19.
12. Mäkelä, M. Knowing through making: The role of the artefact in practice-led research. Knowledge. *Technol. Policy* **2007**, *20*, 157–163. [CrossRef]
13. Wilsdon, J.; Bar-Ilan, J.; Frodeman, R.; Lex, E.; Peters, I.; Wouters, P.F. Next-Generation Metrics: Responsible Metrics and Evaluation for Open Science. 2017. Available online: https://op.europa.eu/en/publication-detail/-/publication/b858d952-0a19-11e7-8a35-01aa75ed71a1 (accessed on 12 August 2022).
14. Pourret, O.; Irawan, D.E.; Shaghaei, N.; Rijsingen, E.M.V.; Besançon, L. Toward More Inclusive Metrics and Open Science to Measure Research Assessment in Earth and Natural Sciences. *Front. Res. Metr. Anal.* **2022**, *7*, 13. [CrossRef]
15. Bornmann, L. Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *J. Informetr.* **2014**, *8*, 895–903. [CrossRef]
16. Sugimoto, C.R.; Work, S.; Larivière, V.; Haustein, S. Scholarly use of social media and altmetrics: A review of the literature. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 2037–2062. [CrossRef]
17. Priem, J.; Hemminger, B.H. Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday* **2010**, *15*. [CrossRef]
18. Priem, J.; Taraborelli, D.; Groth, P.; Neylon, C. Altmetrics: A Manifesto, 26 October 2010. Available online: http://altmetrics.org/manifesto (accessed on 12 August 2022).
19. Herb, U.; Impactmessung, Transparenz & Open Science. Young Information Scientist 2016, 1. Available online: http://eprints.rclis.org/29991/ (accessed on 9 September 2022).
20. UNESCO. Recommendation on Open Science. 2021. Available online: https://www.unesco.org/en/natural-sciences/open-science (accessed on 21 August 2022).
21. Ali-Khan, S.E.; Jean, A.; MacDonald, E.; Gold, E.R. Defining Success in Open Science [version 2; peer review:2 approved]. *MNI Open Res.* **2018**, *2*. [CrossRef]
22. European Commission. Open Science. 2020. Available online: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en (accessed on 23 August 2022).
23. Gerber, A.; Forsberg, E.M.; Shelley-Egan, C.; Arias, R.; Daimer, S.; Dalton, G.; Cristóbal, A.B.; Dreyer, M.; Griessler, E.; Lindner, R.; et al. Joint declaration on mainstreaming RRI across Horizon Europe. *J. Responsible Innov.* **2020**, *7*, 708–711. [CrossRef]
24. Wouters, P.; Ràfols, I.; Oancea, A.; Kamerlin, L.; Holbrook, B.; Jacob, M. Indicator Frameworks for Fostering Open Knowledge Practices in Science and Scholarship. 2019. Available online: https://op.europa.eu/en/publication-detail/-/publication/b69944d4-01f3-11ea-8c1f-01aa75ed71a1 (accessed on 12 August 2022).
25. Willinsky, J. The unacknowledged convergence of open source, open access, and open science. *First Monday* **2005**, *10*. Available online: https://firstmonday.org/ojs/index.php/fm/article/download/1265/1185?inline=1 (accessed on 13 August 2022). [CrossRef]
26. Linux Foundation. Measuring Your Open Source Program's Success. 2017. Available online: https://www.linuxfoundation.org/tools/measuring-your-open-source-program-success/ (accessed on 12 August 2022).
27. Crowston, K.; Annabi, H.; Howison, J. Defining Open Source Software Project Success, ICIS 2003 Proceedings 2003, 28. Available online: https://aisel.aisnet.org/icis2003/28 (accessed on 16 August 2022).
28. Seddon, P.B. A respecification and extension of the DeLone and McLean model of IS success. *Inf. Syst. Res.* **1997**, *8*, 240–253. [CrossRef]
29. Aue, J.; Haisma, M.; Tómasdóttir, K.F.; Bacchelli, A. Social Diversity and Growth Levels of Open Source Software Projects on Github. *Proc. 10th ACM/IEEE Int. Symp. Empir. Softw. Eng. Meas. Ser. ESEM* **2016**, *16*, 1–6. [CrossRef]
30. Linux Foundation. CHAOSS Metrics. 2022. Available online: https://chaoss.community/wp-content/uploads/2022/04/English-Release-2022-04-18.pdf (accessed on 12 August 2022).
31. Goggins, S.; Lumbard, K.; Germonprez, M. Open source community health: Analytical metrics and their corresponding narratives. In Proceedings of the 2021 IEEE/ACM 4th International Workshop on Software Health in Projects, Ecosystems and Communities (SoHeal), Madrid, Spain, 29 May 2021; pp. 25–33.
32. Link, G.J. Open Source Project Health. USENIX PATRONS Login, Fall 2020. 2020, pp. 31–35. Available online: https://www.usenix.org/system/files/login/issues/login_fall20_issue.pdf#page=33 (accessed on 12 August 2022).

33. Coalition for Advancing Research Assessment. Agreement on Reforming Research Assessment 2022. Available online: https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf (accessed on 21 October 2022).
34. Gadd, E. Better, Fairer, more Meaningful Research Evaluation-In Seven Hashtags. Impact of Social Sciences Blog. 2018. Available online: https://blogs.lse.ac.uk/impactofsocialsciences/2018/09/27/better-research-evaluation-in-seven-hashtags/ (accessed on 21 October 2022).
35. Agate, N.; Long, C.P.; Russell, B.; Kennison, R.; Weber, P.; Sacchi, S.; Rodhy, J.; Thornton-Dill, B. Walking the Talk: Toward a Values-Aligned Academy. 2022. Available online: https://hcommons.org/deposits/item/hc:44631/ (accessed on 21 October 2022).
36. European Commission. Responsible Research and Innovation. 2019. Available online: https://web.archive.org/web/20190702224847/https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation (accessed on 18 August 2022).
37. Rome Declaration. Rome Declaration on Responsible Research and Innovation in Europe. 2014. Available online: https://ec.europa.eu/research/swafs/pdf/rome_declaration_RRI_final_21_November.pdf (accessed on 4 September 2022).
38. Spaapen, J.; Strand, R.; Bauer, M.W.; Hogan, E.; Revuelta, G.; Stagl, S.; Paula, L.; Guimaraes Pereira, A. *Indicators for Promoting and Monitoring Responsible Research and InnovationReport from the Expert Group on Policy Indicators for Responsible Research and Innovation*; Directorate-General for Research & Innovation Science with and for Society: Brussels, Belgium, 2015; Available online: https://op.europa.eu/en/publication-detail/-/publication/306a7ab4-f3cb-46cb-b675-9697caf5df19 (accessed on 22 August 2022).
39. Skains, R.L. Creative practice as research: Discourse on methodology. *Media Pract. Educ.* **2018**, *19*, 82–97. [CrossRef]
40. Haseman, B. A manifesto for performative research. *Media Int. Aust.* **2006**, *118*, 98–106. [CrossRef]
41. Mäkelä, A.M.; Nimkulrat, N. Reflection and Documentation in Practice-Led Design Research. Nordes 2011 (4). Available online: https://archive.nordes.org/index.php/n13/article/view/98/82 (accessed on 3 September 2022).
42. DeLanda, M. Emergence, causality and realism. *Archit. Theory Rev.* **2012**, *17*, 3–16. [CrossRef]
43. De Paoli, S.; Forbes, P. TRIPLE Research Interviews with Academics and Researchers (0.1) [Data set]. *Zenodo* **2021**. [CrossRef]
44. Forbes, P.; De Paoli, S.; Błaszczyńska, M.; Maryl, M. TRIPLE Deliverable: D3.1 Report on User Needs (Draft). *Zenodo* **2020**. [CrossRef]
45. Gould, J.D.; Lewis, C. Designing for Usability: Key Principles and What Designers Think. *Commun. ACM* **1985**, *28*, 300–311. [CrossRef]
46. European Commission. She Figures 2021: Gender in Research and Innovation | Statistics and Indicators. 2021. European Union.. 2021. Available online: https://ec.europa.eu/info/sites/default/files/research_and_innovation/strategy_on_research_and_innovation/documents/ec_rtd_shefigures-2021-report.pdf (accessed on 21 October 2022).

Adoption of Transparency and Openness Promotion (TOP) Guidelines across Journals

Inga Patarčić [1],* and Jadranka Stojanovski [2,3]

1   Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
2   Department of Information Sciences, University of Zadar, 23000 Zadar, Croatia
3   Ruđer Bošković Institute, 10000 Zagreb, Croatia
*   Correspondence: inga.patarcic@mdc-berlin.de

Abstract: Journal policies continuously evolve to enable knowledge sharing and support reproducible science. However, that change happens within a certain framework. Eight modular standards with three levels of increasing stringency make Transparency and Openness Promotion (TOP) guidelines which can be used to evaluate to what extent and with which stringency journals promote open science. Guidelines define standards for data citation, transparency of data, material, code and design and analysis, replication, plan and study pre-registration, and two effective interventions: "Registered reports" and "Open science badges", and levels of adoption summed up across standards define journal's TOP Factor. In this paper, we analysed the status of adoption of TOP guidelines across two thousand journals reported in the TOP Factor metrics. We show that the majority of the journals' policies align with at least one of the TOP's standards, most likely "Data citation" (70%) followed by "Data transparency" (19%). Two-thirds of adoptions of TOP standard are of the stringency Level 1 (less stringent), whereas only 9% is of the stringency Level 3. Adoption of TOP standards differs across science disciplines and multidisciplinary journals (N = 1505) and journals from social sciences (N = 1077) show the greatest number of adoptions. Improvement of the measures that journals take to implement open science practices could be done: (1) discipline-specific, (2) journals that have not yet adopted TOP guidelines could do so, (3) the stringency of adoptions could be increased.

Keywords: transparency and openness promotion; TOP guidelines; TOP Factor; open science; publishing policies

1. Introduction

Science advances knowledge through research and disseminates results via different kinds of research outputs, among which are the most visible scholarly publications. Although scholarly publishing has been witnessing a transition from 'publishing as fast as possible' towards open science practices of 'sharing knowledge as early as possible' [1], many research outputs are stored behind the publisher's paywall, and thus, are inaccessible to a broader audience [2,3]. Recent open science (OS) initiatives call for a fundamental change in how data, materials or results are produced and published, and establish novel practices on how researchers engage and communicate with the public [4]. In general, novel practices in scholarly communication, enabling open access to publications, data, code, methods, educational materials and transparent and open peer review process, have been established to create a more effective and inclusive system of science [5].

Coordinated efforts of publishers, funders, policymakers, institutions, libraries and researchers as a targeted group are required to achieve a greater level of openness in science [6]. Although most scientists agree to embrace disciplinary norms and values of transparency, openness, and reproducibility [7], that was not necessarily the case in practice [8–10]. Furthermore, Baker (2016) showed that "more than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half

have failed to reproduce their own experiments" [11]. The benefits of sharing the research outputs are increased transparency and trust, reproducibility and reuse, increased visibility, readability, citation and impact, long-term archiving and preservation, recognition and reputation, and better collaboration opportunities. Still, many scientists are not eager to implement open science norms in their everyday practices if it is left exclusively to their decision and efforts. However, different '(deposit) mandates' have proven to be effective incentives [12]. Thus, along with national/institutional open science policies and funders' requirements [13], the journal's policies and requirements represent one of the key elements of the incentive system that promotes the adoption of open science practices.

In 2015, the Transparency and Openness Promotion (TOP) guidelines were defined in order to become a shared standard for open practices that journals can adopt to promote open science [14]. TOP guidelines consist of eight modular standards that can be used to evaluate journal's policies on data citation, data transparency, material transparency, code transparency, design and analysis, study pre-registration, analysis pre-registration and replication. Standards were defined to have three tiers of increasing stringency-Level 1, Level 2, and Level 3-that move scientific communication toward greater openness; from mentioning specific open science practices towards encouraging, requiring or enforcing them. For example, if journals require researchers to state whether and where code is available, this qualifies them for code transparency standard Level 1. On the other hand, Level 2 code transparency standard demands code to be posted to a trusted repository, whereas Level 3 additionally requires "reported analyses reproduced independently prior to publication" [14,15]. Although similar requirements were made for data and materials transparency standards, other standards have custom-made requirements; such as Level 1 data citation standard is met if a journal clearly describes the citation of data in guidelines.

Since journals can adopt one or more guidelines with different stringencies, in 2020, the Center for Open Science launched the TOP Factor-a quantitative score that reflects a degree of adherence to the TOP guidelines and "Registered reports/Open science badges" interventions [15]. Overall, TOP Factor includes a total of ten subscales (TOP standards plus two effective interventions: "Registered reports" [16] and "Open science badges" [17]), and can go up to 29 which indicates the highest adherence.

Seven years after TOP guidelines were announced, over 5000 journal and funder signatories expressed their support for the guidelines [18] and policies of at least 2000 journals were examined to define a TOP Factor score. Although a few studies reviewed TOP Factors for discipline-specific journals [6,19–22], according to our knowledge, such analysis has not yet been performed in a larger sample of journals and across scientific disciplines. Thus, published studies did not report on how much the adoption of TOP standards differs across scientific disciplines and how well each standard is adopted in general? Likewise, prior to our study, it was unknown which standards are counted in each TOP Factor score. In order to answer those questions, we analyzed TOP Factor scores and individual levels of adoption of TOP guidelines for two thousand journals reviewed by the Center for Open Science.

2. Materials and Methods

The Center for Open Science has been evaluating journal policies based on the degree to which they comply with the TOP Guidelines and reporting them as a TOP Factor metric. The latest version of the TOP Factor metric scopes 2000 journals and explains the steps journals are taking to comply with each TOP standard. In other words, the metric provides per journal information on: (1) per standard stringency level (Level 0–3), (2) description of journal's policy that corresponds to each TOP standard, and 3) the TOP Factor. Importantly, nine categories that build a TOP Factor can score from 0 to 3, whereas one category-"Open Science badges"-scores 0 to 2 [23].

We downloaded the TOP Factor (v33, 29 August 2022 3:12 PM) metric [24] and analyzed its content with an in-house R script. First, we extracted levels of stringency required

for each TOP standard across journals and wrote R-scripts that produced figures in the Results section and calculated the mean and median values.

Second, in order to get statistics about the implementation of the TOP guidelines across discipline-specific journals, we extracted information about journal's disciplines from the Scopus content database. We downloaded SCOPUS content coverage [25]. Scopus was selected as a multidisciplinary bibliographic database indexing 27.253 active journals. We selected only the first sheet of the .xlsx file (Scopus Sources May 2022) [25] and imported information about 43,016 active and non-active journals into R. We removed all inactive journals from our analysis and worked with 27.253 active ones. In the Scopus content coverage, journals are categorized into four top-level disciplines of science: life, social, health and physics. However, some journals belong to different combinations of the aforementioned top-level disciplines. We defined such journals as multidisciplinary.

We paired information reported in the Scopus database and TOP Factor metrics by matching the journal's names, or when a name match was not identified, we identified a match based on E-ISSN or P-ISSN identifiers. With such an approach, we managed to match 1824 (91%) journals.

Lastly, to test whether percentages of discipline-specific journals are equal between Scopus content and TOP Factor metric content, we performed Pearson's Chi-squared test using the chisq.test() function in R (number of journals per scientific discipline was used as an input for the test).

Analysis was made in R version 4.1.0. with an in-house R-script deposited in Zenodo [26] and Github [27].

3. Results

3.1. Most Journals Adopt a Single TOP Standard and Most Standards Are Adopted with Stringency Level 1

We identified a total of 4661 examples of adoption of TOP standards and two additional interventions ("Registered reports" and "Open science badges") in 2000 journals from the TOP Factor metric. In general, identified adoptions were of the stringency Level 1 (less stringent, 67%), followed by Level 2 (N = 1105, 24%) and Level 3 (N = 412, 9%) with median value of 1 and mean = 1.4 (Table 1).

That held true even when we compared individual standards: depending on the standard stringency Level 1 was identified in 62% to 86% of the journals. For seven out of eight TOP standards, Level 3 policies were adopted in less than 6% of journals, however, the "Replication" standard had an adoption across 37% of journals. On the other hand, two interventions-"Registered reports" and "Open science badges"-were generally adopted with stringency Level 3 and Level 2, respectively. Across TOP standards, mean and median levels of adoption were below 1 (with an exception of data citation standard which median of adoption was equal to 1). However, if we filtered out journals that did not implement given standards, we showed that the mean leave of adoption was generally slightly higher than 1, with an exception of replication standard (mean = 1.8, median = 1), "Registered reports" (mean = 2.8, median = 3), and "Open science badges" (mean = 1.9, median = 2)

The majority of journal policies adopted "Data citation" standards (N = 1192, 60%), followed by 45% of journals that adopted "Data transparency", 36% adopted "Design/Analysis reporting guidelines", and 29% adopted "Analysis and Code" transparency (Figure 1). The other four standards were not so frequently adopted: 15% of the journals adopted "Replication", 14% "Materials transparency", 10% "Study pre-registration" and 9% "Analysis plan pre-registration" standards. Likewise, only 10% of the journals required "Registered reports", and 6% issued "Open Science badges".

Table 1. Categories of TOP standards and their implementation across journals. Stringency levels were analyzed and reported as the number, mean value and median of journals that adopt a given standard. To obtain statistics for columns 3 and 4, all examples of not implemented TOP standards were filtered out of the analysis. For example, 808 journals that do not implement "Data citation" standards were excluded when mean and median values of the subsample were calculated. To obtain statistics for columns 5 and 6, we calculated the mean and median value across all 2000 journals.

| Category of TOP Standards/Interventions	Number of Journals that Adopt Specific TOP Standard/Intervention (Level 1/Level 2/Level 3)	Mean Level of TOP Standard Adoption (Subsample)	Median Level of TOP Standard Adoption (Subsample)	Mean Level of Adopted TOP Standard (All Journals)	Median Level of Adopted TOP Standard (All Journals)
Data citation	1192 (734/446/12)	1.4	1	0.8	1
Data transparency	906 (724/158/24)	1.2	1	0.6	0
Analysis/code transparency	583 (427/120/36)	1.3	1	0.4	0
Materials transparency	274 (181/84/9)	1.4	1	0.2	0
Design/analysis reporting guidelines	715 (545/142/28)	1.3	1	0.5	0
Study pre-registration	193 (165/22/6)	1.2	1	0.1	0
Analysis plan pre-registration	178 (150/22/6)	1.2	1	0.1	0
Replication	295 (183/2/110)	1.8	1	0.3	0
Registered reports	202 (20/1/181)	2.8	3	0.3	0
Open Science badges	123 (15/108/0)	1.9	2	0.1	0
Number of policies	**4661 (3144/1105/412)**	**1.4**	**1**	**0.3**	**0**
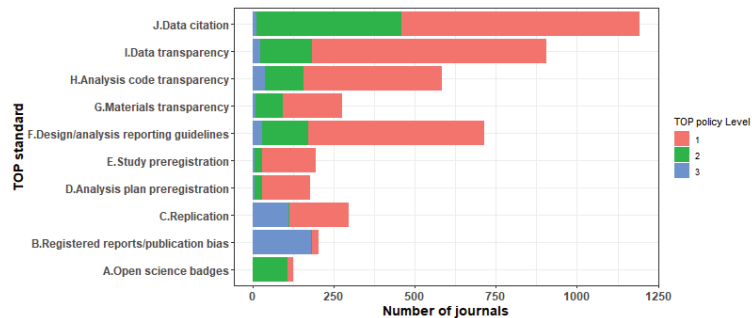
Figure 1. Histogram of the number of journals that adopt each individual TOP standard and corresponding stringency level. Stringency levels 1, 2 and 3 of each TOP standard are indicated in different colors, whereas each row is a single TOP standard.

Interestingly, almost one-fourth of the analyzed journal policies did not adopt any of the TOP standards; in other words their TOP Factor score equals to 0 (N = 455, 23%, Figure 2a). A total of 561 journals adopted a single TOP standard (N = 561, Figure 2b, Supplementary Table S1), 70% of which were an adoption of the "Data citation" standard (Supplementary Table S2). In the case of journals that adopt two different standards, the most frequent adoption was of "Data citation" and "Data transparency" standards (47%, N = 123), followed by a combination of "Data citation" and "Design analysis reporting guidelines" (19%, Supplementary Table S3). When three categories of standards were adopted together by a journal, "Data citation", "Data transparency" and "Design analysis reporting guidelines" was adopted in 25% of journals, followed by 21% of journals adopting "Data transparency", "Analysis code transparency" and "Materials transparency" (Supplementary Table S4).
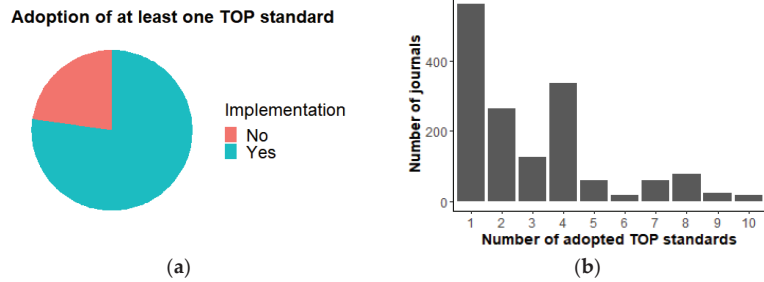
Figure 2. Overview of the adoption of TOP standards across 2000 journal policies from the TOP Factor metric. (**a**) Pie chart that shows the number of journals that implement at least one TOP standard: No-indicates journals that do not implement TOP standards (N = 455, TOP Factor = 0), Yes-indicates a number of journals that implement at least one TOP standard (N = 1545, TOP Factor > 0). (**b**) Histogram of the number of journals that adopt one or more TOP standards. An individual journal can adopt up to eight TOP standards, however, two additional interventions-"Open Science badges" and "Registered Reports"-were added to the standards for this plot. Thus, category 10 corresponds to journals that adopt all 8 standards and 2 interventions.

The high proportion of journals adopted four different standards (N = 337); 74% of which correspond to the adoption of "Data citation", "Data transparency", "Analysis code transparency" and "Design analysis reporting guidelines" standards (N = 245, Supplementary Table S5). A total of 78 journals adopt all eight TOP standards, whereas 17 journals issue "Open Science badges" and require "Registered reports" on top of the adoption of all eight TOP standards (Supplementary Table S1).

3.2. Adoption of TOP Standards Differ across Disciplines of Science

Journals reviewed in the TOP Factor metric articles are multidisciplinary (35%), and publish articles from social sciences (33%) or other disciplines: "Health", "Life" and "Physical" (16%, 9% and 6% of journals, respectively). However, "active" journals from the Scopus content coverage publish mostly articles from social sciences (32%), followed by 22% of multidisciplinary journals, 21% of "Physical", 17% of "Health" and 7% of journals in the category "Life" (Table 2). These two databases have significantly different content when journal disciplines are considered (Pearson's Chi-squared test X-squared = 328.21 with p-value $< 2.2 \times 10^{-16}$).

Table 2. Number and percentage of discipline-specific journals in the TOP Factor metric and Scopus database.

Discipline of Science	Number of Journals in the TOP Factor Metric	Share of a Total Number of Journals in the TOP Factor Metric	Number of Journals in Scopus	Share of a Total Number of Journals in the Scopus
Social	595	33%	8799	32%
Health	297	16%	4603	17%
Physical	118	6%	5848	21%
Life	162	9%	1937	7%
Multidisciplinary	642	35%	5973	22%
NA *	10	1%	93	0.34%

* NA = not available information.

We stratified TOP Factor metrics based on the TOP standards and disciplines of science and showed that multidisciplinary journals have the highest number of adoptions across all eight TOP standards (N = 1418, Supplementary Table S6). "Materials transparency"

standard is an exception since it was adopted by an equal number of multidisciplinary and social sciences journals (N = 81). Journals from the field of social sciences had the second highest number of adoptions of "Data citation" (N = 259), "Data transparency" (N = 220), "Materials transparency" (N = 81), "Study pre-registration" (N = 52), "Analysis plan pre-registration" (N = 42) and "Replication" standards (N = 87). However, a higher number of journals from health sciences (N = 122 & N = 184) than from social sciences (N = 107 & N = 101) adopted "Analysis code transparency" and "Design analysis reporting guidelines", respectively. Only 25 "Physical" journals, as compared to 252 multidisciplinary journals, adopt "Design analysis reporting guidelines".

In terms of shares of a total number of journals, we observed that across all fields of science the "Data citation" standard was most frequently adopted (24–30% of journals), followed by "Data transparency" (18–21%). Interestingly, "Design analysis reporting guidelines" standard was frequently reported in "Health" (22%), multidisciplinary (17%) and "Life" sciences (16%), but was not so frequent in "Social" and "Physical" disciplines (9%). Likewise, four standards: "Materials transparency", "Study pre-registration", "Analysis plan pre-registration" and "Replication" were less frequently, but generally evenly, adopted (3–9%) across disciplines. Issuing of "Open science badges" (N= 31 & N = 55) or requesting "Registered reports" (N = 56 & N = 73) was shown to be done mainly by multidisciplinary and social sciences related journals, whereas, for example, only two "Physical" journals required "Registered reports" (Figure 3).
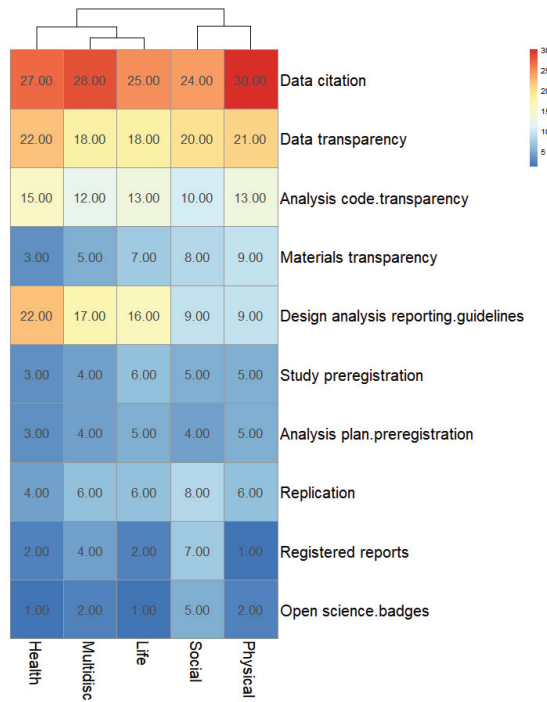

Figure 3. Heatmap of the percentage of discipline-specific journals that adopt each individual TOP standard. Statistics are reported per column (e.g., column's values sum up to 100).

4. Discussion

An increasing number of journals started adopting a widely appreciated set of TOP guidelines to promote research transparency, openness and reproducibility [6,14,15,19–22]. Journals adopt policies on data citation, data transparency, material transparency, code transparency, design and analysis, study pre-registration, analysis pre-registration and

replication, "in a progressive manner, with policies, such as on the availability of data and code, increasing in strength and rigour over time" [28].

We identified 4661 adoptions of TOP guidelines in 2000 journals, and as expected, the great majority of the journals implement a single TOP standard of the stringency Level 1, where standards are just articulated, stated or described. Although, due to differences in methodology, our results cannot be directly compared to the previous discipline-specific studies [20–22], our finding could provide an explanation for the observed low median values of the TOP Factor. Similarly to the same studies, we identified differences in the number and level of adoption across the TOP standards: "Data citation" and "Data transparency" were the most frequently adopted set of guidelines, thereby rewarding concerned researchers for not receiving more credit for sharing data [29], for the effort they have spent engaging in open practices [14]. In addition, proper data citation supports collaboration and reuse of data, proper attribution and credit and enables reproducibility of findings [30]. However, it remains to be seen what proportion of articles actually report well-formed links to data, data itself and if there is an added value in providing such links [28]. Interestingly, journals frequently adopted the combination of four standards: "Data citation", "Data transparency", "Analysis code transparency" and "Design analysis reporting guidelines", thereby incentivizing openness across all scientific processes: data, code, analysis protocols and design and reducing vague and incomplete reports that decrease confidence in scientific results.

Although standards for pre-registration of studies facilitate the discovery of research, two standards that address pre-registration were more widely adopted by social sciences related journals and less by journals from health and physics disciplines. This is certainly surprising for the field of health, given that application and registration of research are mandatory in many countries (e.g., for trials) and the standard practice of publication of many health journals [31]. Similar was with the category of "Replication" which recognizes the value of replication for independent verification of research results and scientific progress [32]. Our findings indicate that disciplines out of the social sciences can develop their policies in the direction of pre-registration and requesting replication. However, this should be reviewed in the light that journals reported in the TOP Factor metric represent a biased subset of existing science journals, when it comes to the disciplines they cover: for example, journals in the field of physics were significantly underrepresented in our sample. Consequently, these results should be considered cautiously because the distribution of journal disciplines between the "sample of journals" from the TOP Factor metric database and a "population of journals" from the Scopus database differs significantly.

Our study has a set of limitations. Firstly, we did not evaluate policies of the journals ourselves, and thus, we rely on results provided by the COS. Unfortunately, the method of selecting journals included in the TOP Factor by the COS staff or volunteers is not fully transparent, and since we observed a high percentage of journals without any standard in place, or the absence of journals that implement TOP guidelines, we acknowledge but do not fully interpret discipline bias. Therefore, our findings cannot be considered as an objective presentation of journals' transparency and openness policies but only as an analysis of the present coverage of TOP Factor metric.

This study could certainly be expanded by analyzing the level of implementation of specific standards in practice, especially by using the TRUST process [15]. Namely, research has shown that the mere presence of a standard's statement does not mean that it will be respected in practice [33]. Additionally, the appropriateness of the specific TOP Factor standards for different disciplines could be examined. There is already a discussion and developed methodology on this topic published in [15]. Although it was not our study's topic, we also recorded a journal selection bias towards large publishers using common platforms for policies' recommendations and dissemination, which we plan to investigate further in our next study. Additionally, we are planning to add a component of time to get an insight in the evolution of requirements from publishers for adoption of open science practices.

5. Conclusions

The majority of the journal policies reported in the TOP Factor metric align to at least one of the TOP's standards, most likely "Data citation" (70%) and with the stringency Level 1. We identified standard-specific and discipline specific differences in implementation of the TOP guidelines that indicated that the improvement of the measures that journals take in order to implement open science practices could be made in three directions: (1) journals that have not yet adopted policies that promote open science could do so, (2) the stringency of the requirements for open science practices for journals who adopted such policies could be increased, and (3) discipline-specific actions could be made. For example, journals from social and physical sciences could more often implement "Design analysis reporting guidelines", whereas "Materials transparency" standards should be more often requested in all disciplines. "Design/analysis reporting guidelines" standards could be more frequently implemented by the journals of physical science, whereas "Registered reports" and "Open Science badges" could be largely deployed by "Life", "Physical", "Health" disciplines, etc. However, since the distribution of journal disciplines between the TOP Factor metric and global distribution of journals according to bibliographic databases differs significantly, these results should be considered with caution.

Supplementary Materials: The following supporting information can be downloaded at: https://zenodo.org/record/7361822 (accessed on 25 November 2022), Supplementary Tables S1–S6: Supplementary Table S1. Table of the number of journals that adopt zero to eight TOP standards and two interventions.; Supplementary Table S2. Table of the number and percentage of journals that adopt a single TOP standard.; Supplementary Table S3. Table of the number and percentage of journals that adopt two TOP standards and corresponding combinations.; Supplementary Table S4. Table of the number and percentage of journals that adopt three TOP standards and corresponding combinations.; Supplementary Table S5. Table of the number and percentage of journals that adopt four TOP standards and corresponding combinations.; Supplementary Table S6. Table of the number of journals that implement each TOP standard across different disciplines of science.

Author Contributions: Conceptualization, I.P. and J.S.; methodology, I.P. and J.S.; software, I.P.; validation, I.P. and J.S.; formal analysis, I.P.; data curation, I.P.; writing—original draft preparation, I.P.; writing—review and editing, J.S.; visualization, I.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. TOP Factor data (v33, 29 August 2022 3:12 PM) can be found at https://osf.io/kgnva/files/osfstorage/5e13502257341901c3805317 (accessed on 3 September 2022) whereas Scopus content version (existJuly2022.xlsx) can be downloaded from https://www.elsevier.com/solutions/scopus/how-scopus-works/content?dgcid=RN_AGCM_Sourced_300005030 (accessed on 11 September 2022). The code and data presented in this study are available here: https://zenodo.org/record/7361822 (accessed on 25 November 2022) and can be cited as: [dataset] Inga Patarcic, & Jadranka Stojanovski. 2022. Code & Data for "Adoption of Transparency and Openness Promotion (TOP) guidelines across journals"; Zenodo. Version 4; https://zenodo.org/record/7361822 (accessed on 3 September 2022).

Conflicts of Interest: Since J.S. was one of the editors of the Publications Special Issue, the peer review process and editorial decision were performed independently.

References

1. European Commission; Directorate-General for Research and Innovation; Peters, I.; Frodeman, R.; Wilsdon, J.; Bar-Ilan, J.; Lex, E.; Wouters, P. *Next-Generation Metrics: Responsible Metrics and Evaluation for Open Science*; Publications Office of the European Union: Luxembourg, 2017; Available online: https://data.europa.eu/doi/10.2777/337729 (accessed on 9 September 2022).
2. Rabesandratana, T. The world debates open-access mandates. *Science* **2019**, *363*, 11–12. [CrossRef]
3. Rentier, B. Open science: A revolution in sight? *Interlend. Doc. Supply* **2016**, *44*, 155–160. [CrossRef]

4. Watson, M. When will "open science" become simply "science"? *Genome Biol.* **2015**, *16*, 101. [CrossRef] [PubMed]
5. Bezjak, S.; Clyburne-Sherin, A.; Conzett, P.; Fernandes, P.L.; Görögh, E.; Helbig, K.; Kramer, B.; Labastida, I.; Niemeyer, K.; Psomopoulos, F.; et al. The Open Science Training Handbook Written by 14 international experts during the FOSTER book sprint. In Proceedings of the Open Education Global Conference (oeglobal2018), Delft, The Netherlands, 24–26 April 2018. [CrossRef]
6. Grant, S.; Wendt, K.E.; Leadbeater, B.J.; Supplee, L.H.; Mayo-Wilson, E.; Gardner, F.; Bradshaw, C.P. Transparent, open, and reproducible prevention science. *Prev. Sci.* **2022**, *23*, 701–722. [CrossRef]
7. Anderson, M.S.; Martinson, B.C.; De Vries, R. Normative dissonance in science: Results from a national survey of US scientists. *J. Empir. Res. Hum. Res. Ethics* **2007**, *2*, 3–14. [CrossRef] [PubMed]
8. Ioannidis, J.P.; Munafò, M.R.; Fusar-Poli, P.; Nosek, B.A.; David, S.P. Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends Cogn. Sci.* **2014**, *18*, 235–241. [CrossRef]
9. John, L.K.; Loewenstein, G.; Prelec, D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* **2012**, *23*, 524–532. [CrossRef]
10. O'Boyle, E.H., Jr.; Banks, G.C.; Gonzalez-Mulé, E. The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *J. Manag.* **2017**, *43*, 376–399. [CrossRef]
11. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **2016**, *533*, 452–454. [CrossRef]
12. Harnad, S. Open Access to Research. Changing Researcher Behavior Through University and Funder Mandates. *JeDEM EJournal EDemocracy Open Gov.* **2011**, *3*, 33–41. [CrossRef]
13. Holbrook, J.B. Open science, open access, and the democratization of knowledge. *Issues Sci. Technol.* **2019**, *35*, 26–28. Available online: https://issues.org/wp-content/uploads/2019/06/Holbrook-Philosophers-Corner-SP19-ISSUES-26-28.pdf (accessed on 13 September 2022).
14. Nosek, B.A.; Alter, G.; Banks, G.C.; Borsboom, D.; Bowman, S.D.; Breckler, S.J.; Buck, S.; Chambers, C.D.; Chin, G.; Christensen, G.; et al. Promoting an open research culture. *Science* **2015**, *348*, 1422–1425. [CrossRef] [PubMed]
15. Mayo-Wilson, E.; Grant, S.; Supplee, L.; Kianersi, S.; Amin, A.; DeHaven, A.; Mellor, D. Evaluating implementation of the Transparency and Openness Promotion (TOP) guidelines: The TRUST process for rating journal policies, procedures, and practices. *Res. Integr. Peer Rev.* **2021**, *6*, 9. [CrossRef]
16. *Registered Reports: Peer Review before Results Are Known to Align Scientific Values and Practices*; COS: Charlottesville, VA, USA, 2022; Available online: https://www.cos.io/initiatives/registered-reports (accessed on 14 September 2022).
17. Kidwell, M.C.; Lazarević, L.B.; Baranski, E.; Hardwicke, T.E.; Piechowski, S.; Falkenberg, L.S.; Kennett, C.; Slowik, A.; Sonnleitner, C.; Hess-Holden, C.; et al. Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biol.* **2016**, *14*, e1002456. [CrossRef] [PubMed]
18. *TOP Guidelines*; Center for Open Science: Charlottesville, VA, USA, 2022; Available online: https://www.cos.io/initiatives/top-guidelines (accessed on 28 September 2022).
19. Nosek, B.A.; Hardwicke, T.E.; Moshontz, H.; Allard, A.; Corker, K.S.; Dreber, A.; Fidler, F.; Hilgard, J.; Struhl, M.K.; Nuijten, M.B.; et al. Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **2022**, *73*, 719–748. [CrossRef] [PubMed]
20. Cashin, A.G.; Bagg, M.K.; Richards, G.C.; Toomey, E.; McAuley, J.H.; Lee, H. Limited engagement with transparent and open science standards in the policies of pain journals: A cross-sectional evaluation. *BMJ Evid. Based Med.* **2021**, *26*, 313–319. [CrossRef]
21. Spitschan, M.; Schmidt, M.H.; Blume, C. Transparency and open science reporting guidelines in sleep research and chronobiology journals. *bioRxiv* **2020**. [CrossRef]
22. Hansford, H.J.; Cashin, A.G.; Wewege, M.A.; Ferraro, M.C.; McAuley, J.H.; Jones, M.D. Open and transparent sports science research: The role of journals to move the field forward. *Knee Surg. Sport. Traumatol. Arthrosc.* **2022**, *30*, 3599–3601. [CrossRef]
23. *TOP Factor Rubic*; Center for Open Science: Charlottesville, VA, USA, 2022; Available online: https://osf.io/t2yu5/ (accessed on 28 September 2022).
24. *TOP Factor*; Center for Open Science: Charlottesville, VA, USA, 2022; Available online: https://osf.io/kgnva/files/osfstorage/5e13502257341901c3805317 (accessed on 3 September 2022).
25. *Scopus Content*; Elsevier: Amsterdam, The Netherlands, 2022; Available online: https://www.elsevier.com/solutions/scopus/how-scopus-works/content?dgcid=RN_AGCM_Sourced_300005030 (accessed on 11 September 2022).
26. Patarčić, I.; Stojanovski, J. Code & Data for "Adoption of Transparency and Openness Promotion (TOP) Guidelines across journals" (Version 4, 2022) [Data Set]. Zenodo. Available online: https://zenodo.org/record/7361822 (accessed on 25 November 2022).
27. Patarčić, I. TOP Paper Publications. GitHub Repository, 2022. Available online: https://github.com/IngaPa/TOP_Paper_Publications2022 (accessed on 30 September 2022).
28. Colavizza, G.; Hrynaszkiewicz, I.; Staden, I.; Whitaker, K.; McGillivray, B. The citation advantage of linking publications to research data. *PLoS ONE* **2020**, *15*, e0230416. [CrossRef]
29. Fane, B.; Ayris, P.; Hahnel, M.; Hrynaszkiewicz, I.; Baynes, G.; Farrell, E. The State of Open Data Report. *Digit. Sci.* **2019**, *16*, 29. [CrossRef]
30. Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone, M. (Ed.) FORCE11: San Diego, CA, USA, 2014. [CrossRef]
31. Wood, A.J.J. Progress and deficiencies in the registration of clinical trials. *N. Engl. J. Med.* **2009**, *360*, 824–830. [CrossRef]

32. Ebersole, C.R.; Mathur, M.B.; Baranski, E.; Bart-Plange, D.-J.; Buttrick, N.R.; Chartier, C.R.; Corker, K.S.; Corley, M.; Hartshorne, J.K.; Ijzerman, H.; et al. Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability. *Adv. Methods Pract. Psychol. Sci.* **2020**, *3*, 309–331. [CrossRef]

33. Gabelica, M.; Bojčić, R.; Puljak, L. Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *J. Clin. Epidemiol.* **2022**, *150*, 33–41. [CrossRef]

MDPI

Article

Leveraging Open Tools to Realize the Potential of Self-Archiving: A Cohort Study in Clinical Trials

Delwen L. Franzen

Berlin Institute of Health at Charité—Universitätsmedizin Berlin, QUEST Center for Responsible Research, Charitéplatz 1, 10117 Berlin, Germany; delwen.franzen@bih-charite.de

Abstract: While open access (OA) is growing, many publications remain behind a paywall. This limits the impact of research and entrenches global inequalities by restricting access to knowledge to those that can afford it. Many journal policies allow researchers to make a version of their publication openly accessible through self-archiving in a repository, sometimes after an embargo period (green OA). Unpaywall and Shareyourpaper are open tools that help users find OA articles and support authors to legally self-archive their papers, respectively. This study leveraged these tools to assess the potential of green OA to increase discoverability in a cohort of clinical trial results publications from German university medical centers. Of the 1897 publications in this cohort, 46% ($n = 871/1897$, 95% confidence interval (CI) 44% to 48%) were neither openly accessible via a journal or a repository. Of these, 85% ($n = 736/871$, 95% CI 82% to 87%) had a permission to self-archive the accepted or published version in an institutional repository. Thus, most of the closed-access clinical trial results in this cohort could be made openly accessible in a repository, in line with World Health Organization (WHO) recommendations. In addition to providing further evidence of the unrealized potential of green OA, this study demonstrates the use of open tools to obtain actionable information on self-archiving at scale and empowers efforts to increase science discoverability.

Keywords: green open access; self-archiving; clinical trial; scholarly communication; shareyourpaper; unpaywall

1. Introduction

Open access (OA) refers to the free online access and largely unrestricted sharing and re-use of scholarly research [1]. While there is evidence OA is growing [2–6], many publications remain hidden behind a paywall in subscription journals. This limits the reach and impact of research and entrenches global inequalities by restricting access to knowledge to institutions and individuals that can afford it [7]. The UNESCO Recommendation on Open Science adopted in 2021 outlined several priority areas of action toward achieving open science globally. This included the recommendation to support non-commercial publishing models and promote existing flexibilities in intellectual property systems to broaden access to knowledge for the benefit of scientists and society [8].

One way of increasing access to published research is self-archiving in an OA location (green OA) [9]. In many cases, publisher or journal self-archiving policies allow researchers to make a version of their publication openly accessible, sometimes after an embargo period. These policies typically outline several permissions that differ as to what version can be archived, where it can be archived, and when it can be archived. Self-archiving is also enabled through national or consortia-based licensing of electronic journals [10] as well as through author rights retention, with some universities having adopted rights-retention OA policies that make it unnecessary to obtain permission from publishers [11]. Moreover, several countries have introduced clauses in copyright law that allow researchers to make a version of their publication openly available under certain conditions, regardless of publisher policies [12,13].

The potential of self-archiving to broaden access to research has been clearly demonstrated. An analysis of self-archiving policies of the largest 100 publishers by output volume found that 80.4% of 1.1 million articles published in subscription-based journals could be shared as a postprint (author accepted manuscript or publisher version) in a repository one year after publication [14]. At the same time, a synthesis of previous studies estimated realized green OA to be around 12% [9], suggesting a largely unrealized potential of green OA. Factors contributing to the limited uptake of self-archiving are thought to include a lack of awareness [15] and concerns over copyright infringement [16]. Embargoes on self-archiving specific versions may also contribute to limited uptake of this practice. Suggestions to increase self-archiving have included introducing funder and institutional mandates [17] and providing tools and services that make it quick and easy to accomplish [18].

Efforts to assess and bridge the gap between opportunity and practice have been supported by the development of resources, such as SHERPA/RoMEO, which aggregates self-archiving policies at the level of journals in a machine-readable format. These resources are increasingly being integrated with libraries' deposit systems, albeit in different ways [10,19–21]. Building on these resources, Shareyourpaper by OA.Works (https://shareyourpaper.org, accessed on 19 December 2022) is a tool that supports authors to legally self-archive their papers by distilling applicable policies into machine-readable, self-archiving permissions at the level of individual articles. If a permission is found, the paper is automatically deposited in the generalist repository, Zenodo. All authors need to do is provide the requested version of the paper. While Shareyourpaper can be integrated with institutional repositories and is therefore scalable for libraries, the deposit in Zenodo also empowers efforts to increase self-archiving beyond a given institution (see more information at [22]). Taken together, Shareyourpaper's approach of bringing together automated permissions checking and an automated deposit workflow makes it a promising tool to increase self-archiving at scale.

In this study, Shareyourpaper was used in combination with Unpaywall (OurResearch), an established tool to determine the OA status of publications, to assess the potential of self-archiving to increase the discoverability of clinical trial results. Clinical trials are the backbone of evidence-based medicine. They inform regulators, public health agencies, and doctors of which interventions are safe and effective to use. Public health crises have repeatedly highlighted the practical and ethical importance of providing equitable access to the outputs of health and clinical research [23,24]. The World Health Organization (WHO) Joint Statement on Public Disclosure of Results from Clinical Trials states "[. . .] publications describing clinical trial results should be open access from the date of publication, wherever possible" [25]. This raises the following research questions (RQ):

– **RQ1:** How many clinical trial results publications are openly accessible?
– **RQ2:** For clinical trial results publications that are closed access, what is the potential of self-archiving to achieve compliance with WHO guidelines?

These research questions were addressed building on a validated cohort of publications describing the results of clinical trials conducted at German university medical centers [26]. This study focused on the potential of self-archiving based on journal and publisher policies. While the right of secondary publication introduced in German copyright law is in principle a powerful tool to drive self-archiving at scale, its use in practice has been limited [27], and implementation attempts have generated as yet unresolved legal disputes [28,29]. Furthermore, self-archiving mandates at the level of German research institutions and funders remain uncommon. In this context, taking advantage of journal and/or publisher self-archiving policies remains an important avenue to broaden access to past and future research.

2. Materials and Methods

2.1. Trial and Publication Screening

This study used the previously developed IntoValue cohort of clinical trials and associated publications [26]. The IntoValue cohort comprises interventional clinical trials

conducted at a German university medical center that were registered in ClinicalTrials.gov or the German Clinical Trials Register (DRKS) and completed between 2009 and 2017. In line with WHO definitions [30], trials in this cohort include all interventional studies and are not limited to highly regulated drug trials. The earliest results publications associated with these trials were found through manual searches [31,32]. As updated registry data was downloaded on 1 November 2022, the IntoValue inclusion criteria were re-applied: interventional, study completion date between 2009 and 2017, complete based on study status, and conducted by a German university medical center. The sample was further limited to journal publications with a unique Digital Object Identifier (DOI) that resolved in Unpaywall and were published between 2010 and 2020.

2.2. Determination of OA Status

The OA status of publications in this cohort was obtained with Unpaywall. Unpaywall harvests content from legal sources, such as publishers, repositories, and preprint servers, and has limited coverage of personal websites. It does not harvest content from academic social networks for which concerns have been raised about the persistence of content [2,33]. Thus, the following definition of OA was used in this study: articles that are free to read online in a journal or OA repository. Unpaywall was queried via its API using the R package, UnpaywallR (https://github.com/quest-bih/unpaywallR, accessed on 17 December 2022), and all available OA locations were extracted for each publication. These include gold (openly available in an OA journal), hybrid (openly available under an open license in a subscription-based journal), green (openly available in a repository), bronze (openly available on the journal page but without a clear open license), and closed access. As publications can have several OA types, a hierarchy was applied such that only one OA type was assigned to each publication in descending order: gold, hybrid, bronze, green, and closed access. Thus, green OA in this study refers to publications that were only openly accessible via a repository. Table A1 outlines how OA types were derived in this study. Unpaywall was queried via its API on 17 December 2022.

2.3. Determination of the Potential to Self-Archive

To obtain self-archiving permissions for the publications in the cohort, Shareyourpaper was queried via its API using a custom-made Python script (https://github.com/delwen/oa-archiving-permissions, accessed on 17 December 2022). The 'best permission' in the API response focuses on how a paper can be self-archived in an institutional repository. Here, publications were defined to have the potential for green OA if a 'best permission' was found for archiving either the accepted or published version in an institutional repository and if the embargo (if applicable) had elapsed by the query date. Permissions for the submitted version were not considered. Publications without a 'best permission' and publications with a 'best permission' but no information on the embargo period, archiving location, or version were considered as unclear. Table 1 outlines how article-level permissions were derived in this study. Shareyourpaper was queried via its API on 17 December 2022.

The realized potential of green OA was estimated based on (a) the number of publications that were only openly accessible in a repository (green OA) and (b) the number of closed-access publications for which a self-archiving permission was found (based on the criteria defined in Table 1). By virtue of being archived, green OA publications were assumed to have had a permission for self-archiving in a repository. Neither the version nor the OA location of green OA publications was systematically checked.

Table 1. Criteria used to determine self-archiving permissions based on the Shareyourpaper API query response.

Permission	Combination of Fields
Permission to archive the accepted version in an institutional repository (based on "best permission")	can_archive = TRUE AND "institutional repository" IS IN locations AND (embargo_months = 0 OR embargo has elapsed) AND "acceptedVersion" IS IN versions
Permission to archive the published version in an institutional repository (based on "best permission")	can_archive = TRUE AND "institutional repository" IS IN locations AND (embargo_months = 0 OR embargo has elapsed) AND "publishedVersion" IS IN versions

Combination of fields used to determine the self-archiving permission of publications in this study.

2.4. Software, Code, and Data

Data processing was performed in R (version 4.0.5) [34] and Python 3.9 (Python Software Foundation, Wilmington, DE, USA). All the code generated in this study is available in GitHub under an open license: https://github.com/delwen/oa-archiving-permissions (accessed on 17 December 2022). The data presented in this study are openly available in Zenodo [35].

3. Results

3.1. Trial and Publication Screening

The IntoValue dataset includes interventional trials registered in ClinicalTrials.gov or DRKS conducted at a German university medical center and completed between 2009 and 2017 (*n* = 3788). After applying the exclusion criteria, the sample included 1897 unique clinical trial results publications that resolved in Unpaywall and were published between 2010 and 2020. Figures A1 and A2 provide flow diagrams of the trial and publication screenings, respectively.

3.2. OA Status of Publications

Of the 1897 clinical trial results publications examined in this study, 54% (n = 1026/1897, 95% CI 52% to 56%) were openly accessible in a journal (gold, hybrid, bronze) or in a repository (green). Across all years, the cohort included 432 gold OA (23%, 95% CI 21% to 25%), 141 hybrid OA (7%, 95% CI 6% to 9%), 310 bronze OA (16%, 95% CI 15% to 18%), and 143 green OA (8%, 95% CI 6% to 9%) publications (Figure 1). The smaller contribution of green OA likely reflects the hierarchy of OA locations used, whereby articles that were openly accessible both in a journal and in a repository were not counted as green OA. The proportion of openly accessible publications increased from 50% (n = 19/38, 95% CI 35% to 65%) in 2010 to 73% (n = 67/92, 95% CI 62% to 81%) in 2020. This largely appeared to be due to an increase in gold and hybrid OA. In turn, across all publication years, 46% (n = 871/1897, 95% CI 44% to 48%) of publications were not openly accessible via a journal or repository (Figure 1) (RQ1).
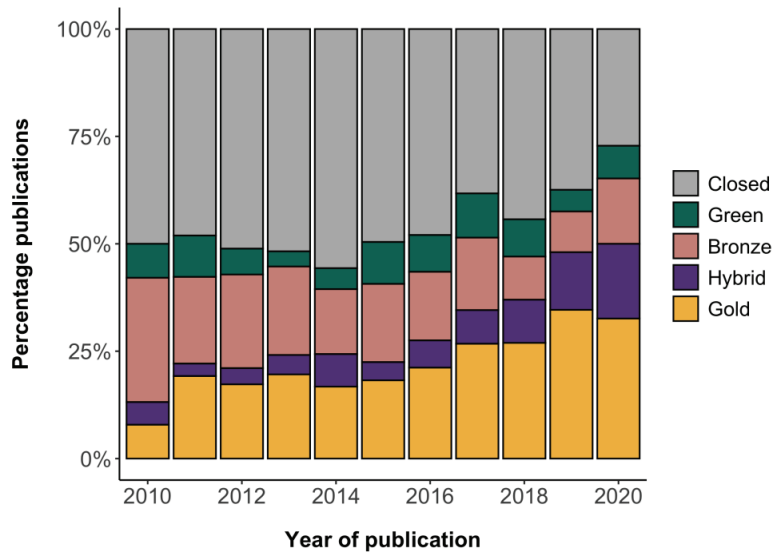
Figure 1. OA status of the clinical trial results publications in the cohort. Given the hierarchy used, green OA represents publications that were only openly accessible in a repository.

3.3. Self-Archiving Permissions of Publications

Focusing on closed-access publications, 85% (n = 736/871, 95% CI 82% to 87%) had sufficient information in Shareyourpaper to derive a self-archiving permission (see Methods). The remaining closed-access publications for which a permission could not be determined include thirty-five (4%, 95% CI 3% to 6%) without a 'best permission' in the API response, ninety-seven (11%, 95% CI 9% to 13%) with no information on the embargo period, and three (0.3%, 95% CI 0% to 1%) with insufficient information on the version and location of self-archiving (dark blue in Figure 2).
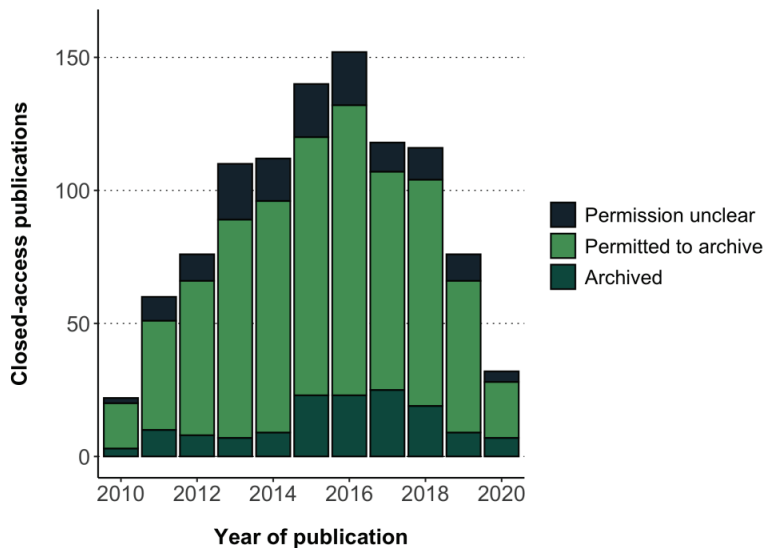
Figure 2. Realized potential of green OA for otherwise closed-access clinical trial results publications in the cohort.

Based on the criteria used in this study (Table 1), all 736 closed-access publications for which a self-archiving permission could be determined in Shareyourpaper had a permission to archive either the accepted or published version in an institutional repository (light green in Figure 2). More specifically, this included permissions to archive the accepted version only (n = 672), the published version only (n = 40), and both the accepted and published version (n = 24). Embargoes associated with these self-archiving permissions ranged between 0–24 months, with most publications having an embargo of 12 months (n = 636) (Figure A3). Taken together, 85% (n = 736/871, 95% CI 82% to 87%) of the closed-access publications in this cohort could be made accessible in an institutional repository, in large part based on permissions issued by journals or publishers to self-archive the accepted version after an embargo of 12 months (RQ2).

The realized potential of green OA for otherwise closed-access articles was estimated based on the number of green OA publications (n = 143) and the number of closed-access publications for which a self-archiving permission was found (n = 736). Thus, 143 of 879 (736 + 143) of otherwise closed-access publications with a self-archiving permission were made openly accessible in a repository (dark green in Figure 2). This corresponds to 16% (95% CI 14% to 19%) realized green OA in this cohort across all publication years.

4. Discussion

4.1. Overall Discussion

This study leveraged open tools to assess the potential of green OA to increase discoverability and generate actionable information on self-archiving at scale. This was demonstrated in a cohort of results publications from clinical trials conducted at German university medical centers. Of the 1897 clinical trial publications published between 2010 and 2020, 46% (n = 871/1897) were not openly accessible via a journal or repository. Of these, 85% (n = 736/871) had a permission to self-archive the postprint in an institutional repository. Thus, many of the closed-access publications in this cohort could be made openly accessible in a location that supports long-term preservation, in line with WHO guidelines [25].

These findings corroborate the largely unrealized potential of green OA found in previous studies [14,18,36]. One study found that 39.2% of a cohort of global health research articles published in journals that allowed self-archiving had been made available via this route [36]. This is higher than the 16% realized green OA found in the present study and may partly reflect the inclusion of a broader range of locations considered as green OA that are not harvested by Unpaywall (e.g., academic social networks). In any case, the gap between opportunity and practice is surprising, given both the practical and ethical relevance of clinical trials and the demonstrated higher impact of self-archived papers compared to paywalled papers based on citations [36,37]. Besides some of the known barriers to self-archiving, such as lack of awareness of self-archiving [15], the unrealized potential of green OA in this cohort likely also reflects the policy context and particularities in the research system in Germany. While many research and funding institutions in Germany have committed to OA, few institutional OA policies mandate self-archiving. This contrasts with other countries, such as the U.K. where high green OA levels have been linked to OA mandates within the Research Excellence Framework [38,39] and the U.S. where the White House Office of Science and Technology Policy recently updated policy guidance to make federally funded research freely available without delay [40]. Moreover, in Germany, researchers based at universities have a high degree of autonomy, which makes it difficult to enforce compliance [3]. The outcome of yet unresolved legal disputes relating to the right of secondary publication introduced in German copyright law will likely shape future efforts to promote self-archiving. A pilot to promote the use of such a clause in the Netherlands ("Taverne Amendment") led to almost 3000 publications being deposited through institutional repositories [13].

Comparisons with other studies are challenging given the use of different approaches and operationalizations of green OA, and applications across different disciplines. However, the overall OA share of 54% in this cohort of clinical trial publications is higher than the 43%

OA share for universities in Germany between 2010 and 2018 obtained in a recent study [3]. In terms of the share of exclusive green OA, the finding that 8% of publications were only openly accessible in a repository (and not the journal) is in line with a previous analysis of the overall literature using Unpaywall, which used the same definition of green OA [2]. Higher estimates of exclusive green OA in other studies (e.g., of 12% [9] and 27.2% [36]) likely reflect the inclusion of non-repository deposit locations (e.g., personal websites).

This study is novel in its use of Shareyourpaper to obtain self-archiving permissions at the level of individual articles. Shareyourpaper aims to derive the most advantageous permission to legally self-archive a paper based on all the relevant policies that may apply for that paper. This includes funder mandates and institutional OA policies. There is increasing interest in monitoring OA at the institutional level and evaluating the impact of interventions on OA uptake [38]. Shareyourpaper seems like a meaningful tool to support such efforts, also given that the underlying data is open and fully machine-readable. Moreover, Shareyourpaper narrows the gap between opportunity and practice by automating the deposit workflow (including metadata entry, permissions, and version checking). This makes it possible to act on self-archiving permissions obtained at levels beyond that of an individual institution (e.g., clinical trials across multiple institutions) and thus empowers efforts to increase self-archiving at scale.

The dissemination and open availability of clinical trial results is essential to support evidence-based decision-making by providers and patients alike and to fulfill ethical obligations to study participants. A previous study focusing on clinical trials in this cohort completed between 2014 and 2017 found that 30% had not reported results five years after trial completion [32]. Non-reporting of clinical trial results distorts our understanding of the medical evidence base, hampers evidence synthesis, and undermines medical decision making [41]. Restricting access to clinical trial results behind paywalls risks further exacerbating these issues. Based on the OA definition used, the findings in this study suggest that compliance with the WHO guideline to make results publications openly accessible *where possible* is low. Most of the closed-access publications in this cohort had a permission to self-archive the accepted version in an institutional repository 12 months after publication. Yet, in most cases this possibility was not exploited to increase discoverability. The approach described in this study is being used as part of an ongoing pilot intervention at the Charité to improve clinical trial transparency, including leveraging publisher self-archiving permissions to make trial results publications openly accessible [42]. In brief, we developed and disseminated trial-specific report cards with feedback on a trial's transparency and recommendations for improvement. If a clinical trial is found to have a results publication that is not accessible in the journal but could be self-archived, the report card recommends researchers to self-archive the publication using Shareyourpaper or by contacting their institutional library.

4.2. Strengths and Limitations

One of the strengths of this study is the development of a fully automated approach based on open tools (Unpaywall and Shareyourpaper) to generate actionable information on self-archiving. This approach can empower efforts to increase science discoverability at scale. The underlying code used to query the APIs is openly available and can be adapted for further use, including the application of different criteria to derive self-archiving permissions. Furthermore, this approach was demonstrated in a validated cohort of clinical trial results publications at the level of German university medical centers. Assessing the realized potential of green OA at this level is meaningful for several reasons: (1) sharing the results of clinical trials is of ethical and practical relevance, (2) WHO guidelines state that clinical trial results publications should be openly accessible where possible, and (3) an analysis at this level can inform interventions to increase self-archiving at a broad scale while also allowing the impact of institutional policies to be evaluated.

This approach also faces limitations. The findings depend on the information in Unpaywall and Shareyourpaper being accurate and up to date. Unpaywall has previously

been shown to provide a conservative estimate of the actual percentage of OA in the literature [2]. In turn, a recent study found changes in Unpaywall OA classifications over time, which may reflect previous errors or a true change in OA status [43]. However, it is unclear whether this affected this study, also because OA classifications were defined in UnpaywallR (Table A1). Taken together, some publications reported as closed may in fact have been openly accessible. Moreover, based on the OA definition used in this study, publications that were only free to read in a location other than a journal or repository (e.g., ResearchGate) were not considered as OA. Publications with no 'best permission' or no embargo information in Shareyourpaper may have been archivable, and other self-archiving routes beyond those examined in this study could be considered. Moreover, while Shareyourpaper takes funder mandates and institutional OA policies into account, they were not considered in this study. Thus, the potential of self-archiving to increase discoverability may be higher than shown here. Finally, the proportion of realized green OA reported in this study is an estimate since: (a) green OA publications were assumed to have had a permission for self-archiving, and (b) neither the version nor the deposit location of green OA publications was checked and may have included other forms of deposit beyond self-archiving of the postprint in an institutional repository.

4.3. Future Research Directions

Libraries may adapt the approach described in this study to support their efforts to populate institutional repositories. This could be achieved by promoting the use of Shareyourpaper among authors at their institution or by embedding the described tools into existing self-archiving workflows. One strength of this approach lies in the ability to relatively quickly flag publications that can be made openly accessible via self-archiving. Providing an institutional record of publications exists, this approach could serve to identify the target group for institutional initiatives that aim to raise awareness of and promote self-archiving, thereby increasing their impact. This includes training workshops and consultation services but also targeted interventions to promote deposits to an institutional repository (e.g., during Open Access week). Whether implementing such an intervention would entail additional resources would depend on its scale and the integration of such tools. A first step may be to pilot such an intervention with publications that have a permission to self-archive the Version of Record.

This approach may also inform interventions to provide access to publications at other levels beyond that of a given institution. For example, an intervention at the level of protocols published in toll-access journals could be particularly meaningful, given that being able to reproduce the methods described in these publications depends on having access to them. Shareyourpaper provides the date at which an embargo will elapse (if applicable) and supports deferred deposits for embargoed articles in Zenodo. Therefore, such interventions could in principle promote future self-archiving already at the time of publication. This would spare researchers the hassle of finding the version of their publication that can be shared long after publication and could help embed self-archiving as an integral phase of the publication process. Finally, while arguably the priority is to make closed-access publications openly accessible, this approach could also inform efforts to increase self-archiving of OA publications to strengthen the long-term preservation of research outputs.

Data Availability Statement: All the code associated with this study is available in GitHub under an open license at: https://github.com/delwen/oa-archiving-permissions (accessed on 17 December 2022) and https://github.com/quest-bih/unpaywallR (accessed on 17 December 2022). The data presented in this study are openly available in Zenodo at: https://doi.org/10.5281/zenodo.7455305 (accessed on 19 December 2022) [35].

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Classification of OA types as defined in UnpaywallR.

OA Type	Description	Combination of Fields
Closed	The publication is not openly accessible.	is_oa = FALSE OR oa_locations is NULL
Gold	The publication is free to read in a fully OA journal.	is_oa = TRUE AND host_type = "publisher" AND journal_is_oa = TRUE
Hybrid	The publication is free to read under an open license in a non-fully OA journal.	is_oa = TRUE AND host_type = "publisher" AND journal_is_oa = FALSE AND license in oa_locations starts with "cc-"
Bronze	The publication is free to read without a clearly identifiable license in a non-fully OA journal.	is_oa = TRUE AND host_type = "publisher" AND journal_is_oa = FALSE
Green	The publication is free to read in a repository (institutional, subject-specific, or other type of repository).	is_oa = TRUE AND host_type = "repository"

Description of the OA types used in this study, along with the combination of fields to derive them in UnpaywallR.
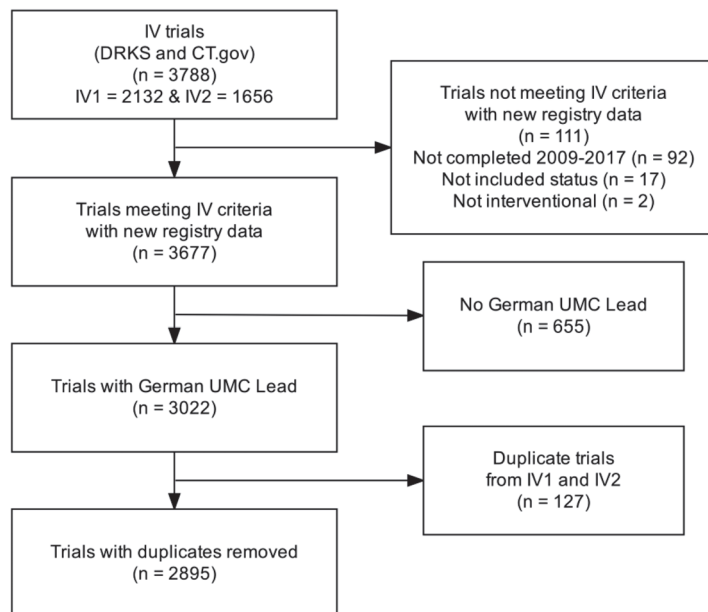
Figure A1. Flow diagram of the trial screening. Abbreviations: CT.gov: ClinicalTrials.gov; DRKS: German Clinical Trials Register; IV: IntoValue; UMC: university medical center.
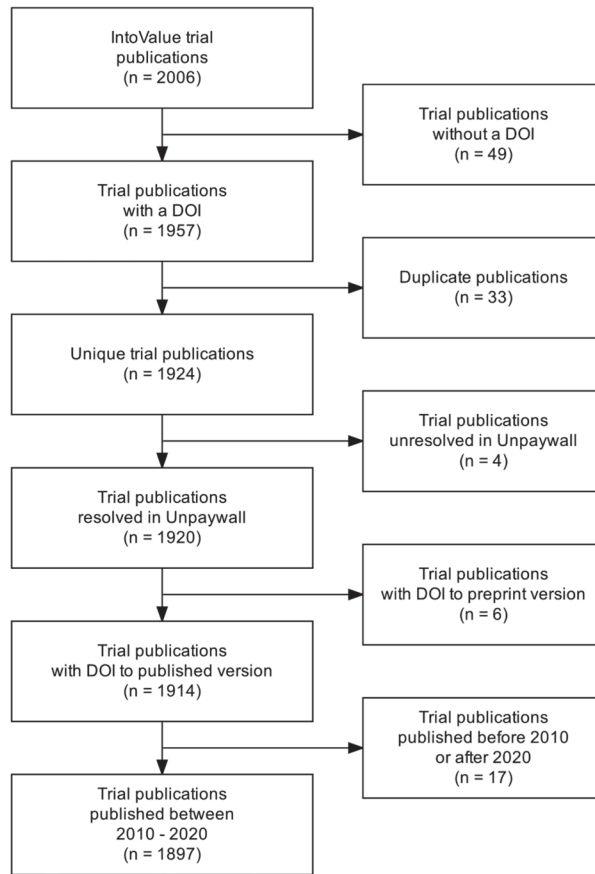
Figure A2. Flow diagram of the publication screening. Abbreviations: DOI: digital object identifier.
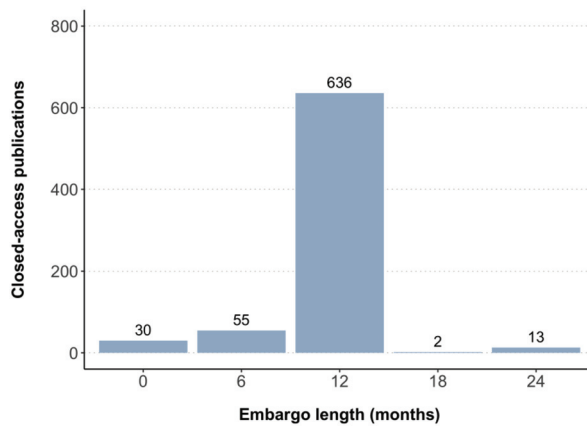

Figure A3. Embargo periods associated with self-archiving permissions found for closed-access publications.

References

1. Suber, P. *Open Access*; MIT Press: Cambridge, MA, USA, 2012.
2. Piwowar, H.; Priem, J.; Larivière, V.; Alperin, J.P.; Matthias, L.; Norlander, B.; Farley, A.; West, J.; Haustein, S. The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles. *PeerJ* **2018**, *6*, e4375. [CrossRef] [PubMed]
3. Hobert, A.; Jahn, N.; Mayr, P.; Schmidt, B.; Taubert, N. Open Access Uptake in Germany 2010–2018: Adoption in a Diverse Research Landscape. *Scientometrics* **2021**, *126*, 9751–9777. [CrossRef]
4. Gargouri, Y.; Larivière, V.; Gingras, Y.; Carr, L.; Harnad, S. Green and Gold Open Access Percentages and Growth, by Discipline. *arXiv* **2012**, arXiv:1206.3664. [CrossRef]
5. Archambault, É.; Amyot, D.; Deschamps, P.; Nicol, A.; Provencher, F.; Rebout, L.; Roberge, G. Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. In *Copyright, Fair Use, Scholarly Communication, etc.*; University of Nebraska-Lincoln: Lincoln, NE, USA, 2014.
6. Laakso, M.; Björk, B.-C. Anatomy of Open Access Publishing: A Study of Longitudinal Development and Internal Structure. *BMC Med.* **2012**, *10*, 124. [CrossRef]
7. Adcock, J.; Fottrell, E. The North-South Information Highway: Case Studies of Publication Access among Health Researchers in Resource-Poor Countries. *Glob. Health Action* **2008**, *1*, 1865. [CrossRef]
8. UNESCO. UNESCO Recommendation on Open Science. *Zenodo* **2021**. [CrossRef]
9. Björk, B.-C.; Laakso, M.; Welling, P.; Paetau, P. Anatomy of Green Open Access. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 237–250. [CrossRef]
10. Blasetti, A.; Golda, S.; Göhring, D.; Grimm, S.; Kroll, N.; Sievers, D.; Voigt, M. Smash the Paywalls: Workflows und Werkzeuge für den grünen Weg des Open Access. *Informationspraxis* **2019**, *5*, 52671. [CrossRef]
11. Shieber, S.; Suber, P. Good Practices for University Open-Access Policies—Harvard Open Access Project. Available online: https://cyber.harvard.edu/hoap/Good_practices_for_university_open-access_policies (accessed on 26 September 2022).
12. Zweitveröffentlichungsrecht—BMBF Digitale Zukunft. Available online: https://www.bildung-forschung.digital/digitalezukunft/de/wissen/urheberrecht/urheberrecht-in-der-wissenschaft/zweitveroeffentlichungsrecht/zweitveroeffentlichungsrecht.html (accessed on 26 September 2022).
13. Sondervan, J.; Schalken, A.; de Boer, J.; Woutersen-Windhouwer, S. Sharing Published Short Academic Works in Institutional Repositories after Six Months: The Implementation of the Article 25fa (Taverne Amendment) in the Dutch Copyright Act. *LIBER Q. J. Assoc. Eur. Res. Libr.* **2021**, *31*, 1–17. [CrossRef]
14. Laakso, M. Green Open Access Policies of Scholarly Journal Publishers: A Study of What, When, and Where Self-Archiving Is Allowed. *Scientometrics* **2014**, *99*, 475–494. [CrossRef]
15. Swan, A.; Brown, S. Open Access Self-Archiving: An Author Study. Available online: https://eprints.soton.ac.uk/260999/ (accessed on 28 September 2022).
16. Creaser, C.; Fry, J.; Greenwood, H.; Oppenheim, C.; Probets, S.; Spezi, V.; White, S. Authors' Awareness and Attitudes Toward Open Access Repositories. *New Rev. Acad. Librariansh.* **2010**, *16*, 145–161. [CrossRef]
17. Harnad, S.; Brody, T.; Vallières, F.; Carr, L.; Hitchcock, S.; Gingras, Y.; Oppenheim, C.; Stamerjohanns, H.; Hilf, E.R. The Access/Impact Problem and the Green and Gold Roads to Open Access. *Ser. Rev.* **2004**, *30*, 310–314. [CrossRef]
18. Troll Covey, D. Self-Archiving Journal Articles: A Case Study of Faculty Practice and Missed Opportunity. *Portal Libr. Acad.* **2009**, *9*, 223–251. [CrossRef]
19. Weisheit, S. Dasselbe in Grün: Analysetool zur Automatisierten Identifizierung von Publikationen Mit Zweitveröffentlichungspotenzial. In Proceedings of the Conference Presentation 8. Bibliothekskongress, Leipzig, Germany, 31 May–2 June 2022.
20. Repanovici, A.; Barsan, I.M.; Repanovici, A.; Barsan, I.M. Integration of SHERPA/RoMEO in Institutional Digital Repositories to Address the Uncertainty of Copyright. *Malays. J. Libr. Inf. Sci.* **2015**, *20*, 1, 29–45.
21. Smith, A. Unlocking Research. A Fast-Track Route to Open Access. Available online: https://unlockingresearch-blog.lib.cam.ac.uk/?p=2536 (accessed on 3 October 2022).
22. OA.Works. Shareyourpaper.Org for Libraries out Now. Available online: https://blog.oa.works/shareyourpaper-org-for-libraries-out-now/ (accessed on 2 October 2022).
23. Modjarrad, K.; Moorthy, V.S.; Millett, P.; Gsell, P.-S.; Roth, C.; Kieny, M.-P. Developing Global Norms for Sharing Data and Results during Public Health Emergencies. *PLoS Med.* **2016**, *13*, e1001935. [CrossRef] [PubMed]
24. Dahn, B.; Mussah, V.; Nutt, C. Opinion | Yes, We Were Warned About Ebola. The New York Times. Available online: https://www.nytimes.com/2015/04/08/opinion/yes-we-were-warned-about-ebola.html (accessed on 3 October 2022).
25. World Health Organization Joint Statement on Public Disclosure of Results from Clinical Trials. Available online: https://www.who.int/news/item/18-05-2017-joint-statement-on-registration (accessed on 15 June 2021).
26. Riedel, N.; Wieschowski, S.; Bruckner, T.; Holst, M.R.; Kahrass, H.; Nury, E.; Meerpohl, J.J.; Salholz-Hillel, M.; Strech, D. Dataset for the IntoValue 1 + 2 Studies on Results Dissemination from Clinical Trials Conducted at German University Medical Centers Completed between 2009 and 2017. *Zenodo* **2021**. [CrossRef]
27. Kein Durchbruch: 5 Jahre Zweitveröffentlichungsrecht für Wissenschaftliche Zeitschriftenbeiträge. Available online: https://irights.info/artikel/kein-durchbruch-5-jahre-zweitveroeffentlichungsrecht-fuer-wissenschaftliche-zeitschriftenbeitraege/29822 (accessed on 26 September 2022).

28. Hartmann, T. Zwang Zum Open Access-Publizieren? Der Rechtliche Präzedenzfall ist Schon da! *LIBREAS Library Ideas* **2017**. Available online: https://libreas.eu/ausgabe32/hartmann/ (accessed on 26 September 2022).
29. Verfassungsgericht Prüft Open-Access-Pflicht. Available online: https://irights.info/artikel/verfassungsgericht-prueft-open-access-pflicht/28820 (accessed on 26 September 2022).
30. World Health Organization (WHO). Glossary. Available online: https://www.who.int/clinical-trials-registry-platform/about/glossary (accessed on 21 December 2021).
31. Wieschowski, S.; Riedel, N.; Wollmann, K.; Kahrass, H.; Müller-Ohlraun, S.; Schürmann, C.; Kelley, S.; Kszuk, U.; Siegerink, B.; Dirnagl, U.; et al. Result Dissemination from Clinical Trials Conducted at German University Medical Centers Was Delayed and Incomplete. *J. Clin. Epidemiol.* **2019**, *115*, 37–45. [CrossRef]
32. Riedel, N.; Wieschowski, S.; Bruckner, T.; Holst, M.R.; Kahrass, H.; Nury, E.; Meerpohl, J.J.; Salholz-Hillel, M.; Strech, D. Results Dissemination from Completed Clinical Trials Conducted at German University Medical Centers Remained Delayed and Incomplete. The 2014–2017 Cohort. *J. Clin. Epidemiol.* **2022**, *144*, 1–7. [CrossRef]
33. Unpaywall: An Open Database of 20 Million Free Scholarly Articles. Available online: https://unpaywall.org/ (accessed on 19 December 2022).
34. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
35. Franzen, D. Dataset from: Leveraging Open Tools to Increase the Potential of Self-Archiving to Increase Discoverability: A Cohort Study in Clinical Trials. *Zenodo* **2022**. [CrossRef]
36. Smith, E.; Haustein, S.; Mongeon, P.; Shu, F.; Ridde, V.; Larivière, V. Knowledge Sharing in Global Health Research—The Impact, Uptake and Cost of Open Access to Scholarly Literature. *Health Res. Policy Syst.* **2017**, *15*, 73. [CrossRef]
37. Harnad, S.; Brody, T. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Mag.* **2004**, *10*, 6. [CrossRef]
38. Huang, C.-K.; Neylon, C.; Hosking, R.; Montgomery, L.; Wilson, K.S.; Ozaygen, A.; Brookes-Kenworthy, C. Evaluating the Impact of Open Access Policies on Research Institutions. *eLife* **2020**, *9*, e57067. [CrossRef] [PubMed]
39. European Commission, Directorate General for Research and Innovation; Chan, T. *Open Research Policies in the United Kingdom*; Publications Office of the European Union: Luxembourg, 2019.
40. Memorandum for the Heads of Executive Departments and Agencies: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. Available online: https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf (accessed on 20 December 2022).
41. Cochrane; The Collaboration for Research Integrity and Transparency (CRIT); Transparency International (TI); TranspariMED. Clinical Trial Transparency: A Guide for Policymakers. 2017. Available online: https://community.cochrane.org/sites/default/files/uploads/Clinical%20Trial%20Transparency%20-%20A%20Guide%20for%20Policymakers.pdf (accessed on 9 July 2022).
42. Salholz-Hillel, M.; Franzen, D.; Müller-Ohlraun, S.; Strech, D. Clinical Trial "Report Cards" to Improve Transparency at Charité and BIH: Survey and Intervention. Available online: https://osf.io/dk6gm (accessed on 14 July 2022).
43. Sanford, H. The State of Unpaywall: Analyzing the Consistency of Open Access Data. In Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada, Spain, 7–9 September 2022.

Automatic XML Extraction from Word and Formatting of E-Book Formats: Insight into the Open Source Academic Publishing Suite (OS-APS)

Carsten Borchert [1], Roberto Cozatl [2], Frederik Eichler [1], Astrid Hoffmann [3] and Markus Putnings [3,*]

1. SciFlow GmbH, Altensteinstraße 40, 14195 Berlin, Germany
2. University and State Library Saxony-Anhalt, Martin-Luther University, Halle-Wittenberg, 06098 Halle, Germany
3. Friedrich-Alexander-Universität Erlangen-Nürnberg, Universitätsbibliothek Erlangen-Nürnberg, Universitätstr. 4, 91054 Erlangen, Germany
* Correspondence: markus.putnings@fau.de; Tel.: +49-9131-8527835

Abstract: Due to resource constraints, most Diamond Open Access journals publish fewer than 25 articles per year, and 75% of journals are not able to provide their content in XML and HTML, primarily providing only PDFs. In order to keep up with larger commercial publishers, a high degree of automation and streamlining of processes is necessary. The Open Source Academic Publishing Suite (OS-APS) project, funded by the German Federal Ministry of Education and Research, aims to achieve this. OS-APS automatically extracts the underlying XML from Word manuscripts and offers optimization and export options in various formats (PDF, HTML, EPUB). The professional corporate design, e.g., of the PDFs, is managed automatically using templates or creating one's own using a Template Development Kit. OS-APS will also connect to scholarly-led and community-driven publishing platforms such as Open Journal Systems (OJS), Open Monograph Press (OMP), and DSpace: the software will be able to be integrated into a wide range of publication processes, whether at small, low-resource commercial Open Access Publishers, or institutional and Diamond Open Access Publishers.

Keywords: automatic typesetting; media-neutral publishing; open access; open source; scholarly publishing; XML/HTML conversion

1. Introduction

The 2021 OA Diamond Journals Study [1] has compiled a representative overview of Diamond Open Access journal operators in its "Part 1: Findings". For example, 53% of journals are operated by fewer than one full-time equivalent (FTE), and 60% of journals rely heavily on volunteers. Due to these resource constraints, most Diamond Open Access journals publish fewer than 25 articles per year, and 75% of journals are not able to provide their content in XML and HTML, primarily providing only PDFs.

In order to keep up with larger commercial publishers and their professionalized content offerings, a high degree of automation and streamlining of processes is necessary. The Open Source Academic Publishing Suite (OS-APS, https://os-aps.de/en/, accessed on 6 December 2022) project funded by the German Federal Ministry of Education and Research aims to achieve this. For this purpose, open source software is to be developed by means of research (especially requirements analysis) and development, with which

1. XML is automatically extracted from (e.g., *.docx Word) author manuscripts;
2. The XML can be processed and, e.g., supplemented with semantic information;
3. Various e-journal or e-book output formats (e.g., XML, HTML, EPUB, and PDFs) can be generated in the respective publisher's preferred design by selecting a template (ready-made or customized) before exporting the file.

Machine learning (ML)/artificial intelligence (AI) is not used here, although it has been considered especially in step 2 (e.g., recognizing headings even if they are not tagged as headings but only made bold) and could play a role in the future of the software, when the basic development is completed.

In addition to the three aforementioned points, OS-APS will also connect to widely-used, scholarly-led and community-driven publishing platforms such as Open Journal Systems (OJS), Open Monograph Press (OMP), and Open Access repositories (e.g., DSpace). The software will be able to be integrated into a wide range of publication processes, whether at small, low-resource commercial Open Access Publishers, or institutional and Diamond Open Access Publishers.

To understand the requirements of these heterogeneous publishers, a practical advisory board and scientific advisory board with representatives from the different publication sectors accompany the OS-APS project. In addition, an extensive survey [2] was conducted across various publishing houses and demo days with corresponding feedback opportunities are held regularly (https://os-aps.de/demo/, accessed on 6 December 2022).

The project is also in line with the recommendations of the OA Diamond Study and its urgent call for cOAlition S Funders and Infrastructures: "Support the development of generic tools to generate structured content in XML and HTML" [3]. This will also be a prerequisite for creating new, dynamic and machine-processable media formats, for example in terms of accessibility and screen readers.

The Open Source software could be thus a significant improvement for smaller, independent Open Access Publishers. It offers the possibility to increase the effectiveness and efficiency of their processes to create, for example, new e-journal article or e-book formats such as HTML and EPUB. These developments contribute to a higher bibliodiversity and may help independent OA publishers to become more viable and sustainable in the long term.

2. Materials and Methods

2.1. Materials

In terms of materials, the OS-APS project team has thus far produced insights into the project's progress via presentations, posters, and articles [2,4–8], various software development sprints documented on GitLab [9], and a demo [10] to provide hands-on testing and feedback on the developments to date.

2.2. Methods

Methodologically, the project work is divided into four milestones. In the first, the requirements for the software were analyzed. In the second, all technology components, interfaces and intended workflows for connecting, e.g., OJS, OMP and DSpace were developed on this basis. In the third, existing journals and book series at the publishing services of the project partners will be iteratively converted to the Open Source Academic Publishing Suite production workflow for the purpose of practical testing and proof of implementation. In the fourth, a release of all open source software development results will take place; the OS-APS software can be downloaded free of charge and installed on the publishers' own servers (all components are browser-based).

The following sections describe the methods used within the milestones in more detail.

The entire OS-APS project is accompanied by two advisory boards, which consist of publishers as well as institutions and libraries active in publishing. The Scientific Advisory Board is responsible for strategic and methodological advice and the User Advisory Board for discussions on practical procedures and requirements.

2.2.1. Software Requirements Analysis

Interviews were conducted with the publishers and publishing services of both the OS-APS project partners and advisory boards. The various publishing operations were mapped graphically via miro (https://miro.com, accessed on 6 December 2022). Miro

provides a range of different options and templates for software development purposes (https://miro.com/templates/developers/, accessed on 6 December 2022). Subsequently, a review and statistical evaluation and classification of the workflows took place, whether they are, e.g., Word-, InDesign- or LaTeX-based, and which export formats are generated.

Through these interviews, a structured overview of various publishing processes could be obtained. Subsequently, a broader online survey based on the interviews was designed in order to reach more experts from the publishing community. This online survey was methodologically designed with Typeform (https://www.typeform.com, accessed on 6 December 2022), which is an online software specialized in creating dynamic surveys with logic flows.

The survey was sent to the mailing lists of the German AG Universitätsverlage (https://ag-univerlage.de, accessed on 6 December 2022), The Association of European University Presses (https://www.aeup.eu, accessed on 6 December 2022), the Enable! community (https://enable-oa.org, accessed on 6 December 2022), Peergroup Produktion of IG Digital (https://www.boersenverein.de/interessengruppen/ig-digital/die-peergroups-der-ig/#accordion-23919, accessed on 6 December 2022), GeSIG (https://gesig.org, accessed on 6 December 2022), Library Publishing Coalition (https://librarypublishing.org, accessed on 6 December 2022), Association of University Presses (https://aupresses.org, accessed on 6 December 2022), Open Access Scholarly Publishing Association (https://oaspa.org, accessed on 6 December 2022), ACUP/APUC—Association of Canadian University Presses (http://acup-apuc.ca, accessed on 6 December 2022), The Association of Japanese University Presses (https://www.ajup-net.com, accessed on 6 December 2022) as well as to cooperation partners of OA-STRUKTKOMM (https://oa-struktkomm.htwk-leipzig.de, accessed on 6 December 2022), DEval Communication and Publications Office (https://www.deval.org/de/publikationen, accessed on 6 December 2022), Center for Digital Systems Berlin (https://www.cedis.fu-berlin.de, accessed on 6 December 2022) and in forums such as the Open Access Books Network (https://openaccessbooksnetwork.hcommons.org, accessed on 6 December 2022) and the German PKP Community Forum (https://forum.pkp.sfu.ca/c/regional-networks/german-topics/13, accessed on 6 December 2022).

The results were evaluated, processed in a structured manner [2] and had a significant impact on some project decisions. Thus, it was decided to initially focus on Word manuscripts for XML extraction, since, e.g., LaTeX or other manuscript format submissions were rather rare (Figure 1).
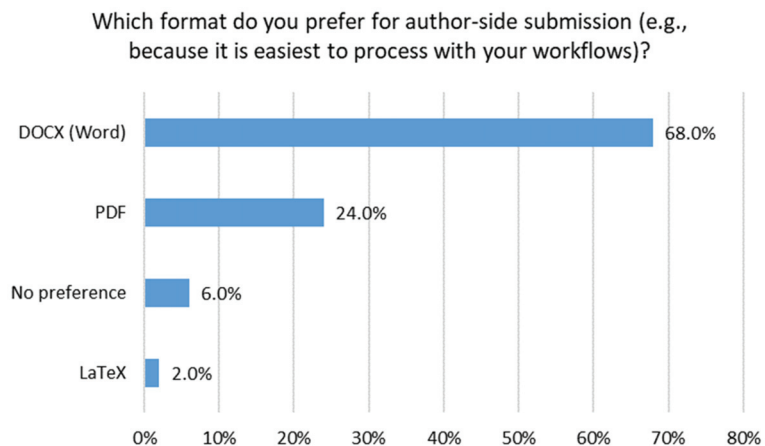

Figure 1. Manuscript acceptance preferences of the surveyed publishers (own representation, translated from [2]).

2.2.2. Open Source Development of the Technology Components

The developing team aimed to build on already existing open source software wherever possible. In several cases it was also possible to build on existing code of the project partner SciFlow, which offers an online platform for collaborative scientific writing and automatic formatting according to the format specifications of renowned academic publishers (cf. https://www.sciflow.net/en/sciflow-free-researchers, accessed on 6 December 2022). SciFlow has extracted the relevant components from its platform and made them available as open source. Additional software development parts in the project context were that

- Word *.docx manuscripts can be fed in the OS-APS browser-based importer; it does not matter which version of MS Word the author uses as long as the file is in *.docx format (hence, all MS Word versions released after the year 2003 should be compatible);
- The XML is automatically extracted from them and is displayed in an editable browser interface; for this, many components were reused from SciFlow's collaborative writing and editing platform;
- Options for optimizing and semantically enriching the XML can be provided;
- Corporate design templates depending on the publisher or its content, for example, for different book or journal series, can be used;
- the publishing user should be able to control this corporate design "look" himself using the planned template development kit.

The Open Source software is currently based on Pandoc, Docker, paged.js, and components extracted by SciFlow from their own platform: https://gitlab.com/sciflow/development/-/milestones (accessed on 6 December 2022).

Pandoc (https://pandoc.org, accessed on 6 December 2022) is a free, GPL-licensed (https://www.gnu.org/licenses/gpl-3.0.html, accessed on 6 December 2022) converter and parser software. It is used to convert one document-based markup and file format to another.

Docker (https://www.docker.com, accessed on 6 December 2022) is an open platform for the running of applications. In this project, it is used to streamline the development for our OJS, OMP and DSpace platforms and to ease the deployment of ready-to-go code from our test environments onto our production systems.

Paged.js (https://pagedjs.org, accessed on 6 December 2022) is an open source library for displaying paginated content in the browser and then creating PDFs and their designs using, e.g., HTML and CCS.

2.2.3. Proof of Implementation and Application of the Open Source Academic Publishing Suite

At the University and State Library in Sachsen-Anhalt (ULB-SA) a testbed was created which can serve the purpose of implementing a number of the software tools developed in the course of the project. The library team supports several publishers' teams in their efforts to publish a wide range of journals spanning across topics such as social geography, transnational economic law, ecology, geosciences. Out of this selection of journals, monographs and series, it has been possible to choose specific examples which have allowed us to not only test specific modules of the OS-APS developed tools but also the connection and integration of our publication tools OJS and OMP systems to a DSpace based publications repository.

In the first case, one journal, the "*Hallesches Jahrbuch für Geowissenschaften*" (*the Yearbook of Geosciences in Halle*) and the ULB-SA's own series "Schriften zum Bibliotheks- und Büchereiwesen in Sachsen-Anhalt" (series on librarianship studies in Saxony-Anhalt) have been selected to be enhanced and given new layouts via the usage of the OS AP suite. In this particular case, Word *.docx templates are uploaded into the OS-APS environment and specific output formats can be generated for importing into the OJS of the ULB-SA. This process streamlines the template generation process of editorial teams, increases its level of automatization, and generally contributes to an increase in citation rates and visibility. These actions are in line with specific Open Science principles which aim at improving the accessibility and reusability of research outputs in fields where these issues may still need attention such as in some areas of the digital humanities. Scholars in these

fields have recognized these endeavors as key components that can promote new research opportunities and can have a great societal value impact [11].

Regarding our connection to our publication tools, a number of journals and series (see for example MLU Human Geography Working Paper Series and the Policy Papers on Transnational Economic Law) are now fully integrated into our OJS/OMP publication cluster and have been exported to our DSpace repository. In this process, all articles have been issued with persistent identifiers (DOIs) and have thus gained a higher visibility and findability given the high data discoverability advantages that the DSpace platform offers. An ongoing migration is taking place so that a total of 13 journals will be migrated in the scope of this project.

As for the technical connection, it has been performed in a way that modular scripts are independently available to suit the different needs of our prospective end users. This means that the developed scripts can be implemented as a full set of scripts or just individually depending on the specifications of the environment where the tools are to be deployed. This modular approach has also meant that our developments do not compromise the native code and functionality of the publication tools in a way that further system upgrades or updates are compromised.

2.2.4. Release of the Open Source Software Development Results

The open source software will be downloadable from https://os-aps.de (accessed on 6 December 2022) and a suitable repository, presumably GitLab, after the end of the project (31 December 2022, if necessary the project will be extended cost-neutrally, then possibly also later in spring 2023). Addition-ally, SciFlow will offer an optional, commercial hosting and support service. Accompanying documentation of the software is of course also provided. The OJS and OMP to DSpace connection scripts and a series of quality control and validation scripts as well as documentation on how these publishing tools have been setup under Docker will be fully available as open source code as part of the project's integral code materials. As part of our project commitments towards open science and transparency and reproducibility, we have already published some of the scripts (in a none-finalized and openly available for scrutiny and feedback version) over the Github repository of the University and State Library in Sachsen Anhalt (explore for instance, our OJS/OMP2DSpace connecting script, and our scripts for the dockerisation of OJS and OMP).

3. Results

The workflow extracted from the perceptions and requirements of the surveyed publishing group is shown in Figure 2 (see also Section 2.2.2).
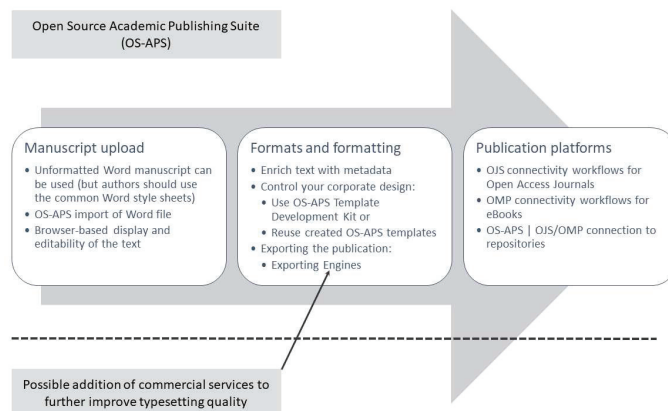

Figure 2. Overview over the Open Source Academic Publishing Suite functionalities.

3.1. OS-APS Importer and Editor

Manuscripts can be imported into the programmed OS-APS editor (Figure 3). By extracting XML structures, elements such as column titles, page breaks, tables, etc. are recognized. In the editor, it is possible to change the text as well as the formatting, if elements were not recognized correctly. If necessary, more metadata (e.g., with regard to accessibility) and semantic references can be added.
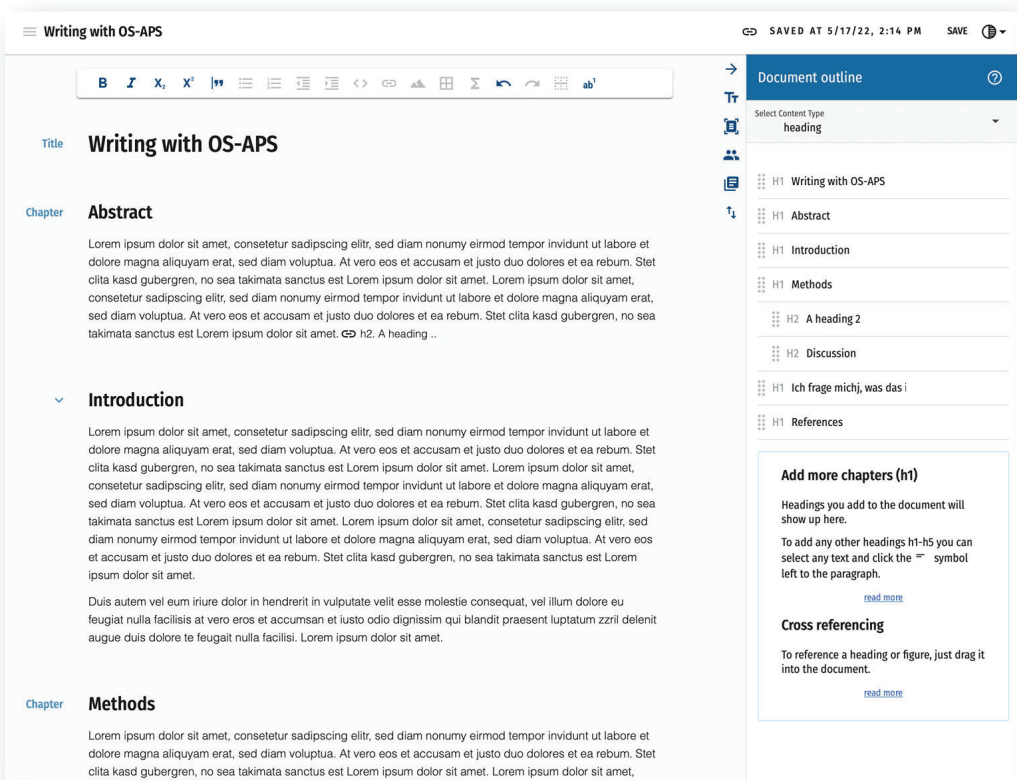

Figure 3. View of the OS-APS editor.

3.2. Template Development Kit and Re-Usable Templates

During export, the corporate design of the respective publisher is mapped via templates. Various standard templates are provided and can be reused.

Further templates and exports can be developed using the Template Development Kit. This is particularly interesting for publishers who have very clear format specifications and do not want to deviate from them.

With the help of the Template Development Kit, individual parameters in ready-made templates can be easily changed. It is also possible to create completely new templates, although this requires prior technical knowledge (esp. web programming). New exports can also be programmed in this way. The Template Development Kit is based on the open source software Pandoc and on SciFlow's own development.

Commercial non-open-source based tools can also be integrated during export for typesetting optimization, specifically, e.g., Prince XML (https://www.princexml.com, accessed on 6 December 2022).

3.3. Connection to OJS, OMP and Repositories Such as DSpace

The OJS and OMP applications are deployed via a Docker environment; the OJS and OMP systems are connected to a DSpace repository (specifically a DSpace 6.3. version, in the case of our project partners). As part of the intended workflow, OJS and OMP data will be exported to DSpace with subsequent return of DOI information. The corresponding publications are displayed in OJS and OMP as well as in DSpace. In general, OJS and OMP are intended as presentation platforms and DSpace for long-term archiving.

The connection scripts as well as documentation on how these publishing tools have been setup will be fully available as open source code as part of the project's integral code (see Section 2.2.4).

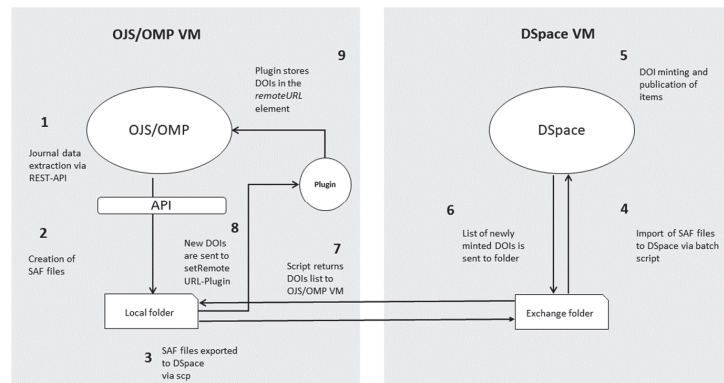Figure 4 gives an overview of the interfaces and data paths.

Figure 4. Schematic representation of the connection of OJS, OMP and DSpace to OS-APS.

3.4. Test Possibility of the Current Results

Every first Wednesday of the month, a public "Demo Day" in form of a video conference (for more information see https://os-aps.de/demo/, accessed on 6 December 2022) takes place, where interested parties are invited to test the current state of the software and give feedback. The input from the "Demo Day" participants is taken into account with regard to the development of OS-APS.

For the final release of the software and documentation, see the methodological announcements in Section 2.2.4.

4. Discussion

4.1. Possible Necessary Exceptions to the OS-APS Workflow

The OS-APS software development project is currently on schedule with its planned milestones. The basic objectives described in the introduction are being achieved. However, the tests conducted so far show that not all special cases that might occur in manuscripts can be implemented graphically in, e.g., PDFs in an ideal way.

This applies primarily but not exclusively to art volumes in which various figures must have exactly the same arrangement as in the original manuscript, grouped figures (e.g., as a block of four or six) with one caption, large, rotated tables, nested tables with multiple content types (e.g., images in different cells of the table), Word text fields or images originally drawn in Word itself with multiple image elements, and much more.

In addition, there may be quality requirements from both publishers and authors that necessitate very thoughtful, small-scale, manual typesetting in InDesign, for example. Examples could include art and exhibition volumes. Here, too, the fully automated approach may not meet these individually high-quality requirements.

4.2. Discussion about Embedding the New Output Formats

What publishers or platforms do with the new output formats remains deliberately open and up for discussion. Those who previously only distributed PDFs via OJS, OMP or repositories (e.g., the university repository, in the case of university presses), must think about how and where they integrate the HTML or EPUB files when using OS-APS, e.g., whether they provide viewers or corresponding plug-ins and whether they also archive them over the long term (or continue to only archive PDF/A). Making publications available as HTML on publisher's or journal websites can be very useful: HTML is a mobile-compatible, easily accessible, indexable, and human- and machine-readable format. These features are also valuable for bibliometric analysis.

In addition, they have to think about URL, DOI and, for eBooks, ISBN registration with regard to the new output formats. In the case of repositories and the use of one front door under which all formats hang, a single DOI could still be used, for example. However, according to the German "Verzeichnis lieferbarer Bücher" (https://vlb.de, accessed on 6 December 2022) as ISBN agency, each different e-book format needs its own ISBN.

4.3. Possible OS-APS Platform Extensions in the Future

The OS-APS platform was developed as open source software. In addition, however, the project partner SciFlow will offer hosting; then for a fee, for those who do not want to set up their own server to run the software or do not want to worry about support.

In the context of this support, further extensions are conceivable, for example with regard to special viewers, such as for EPUB or HTML, depending on the existing information infrastructure, or in terms of accessibility support. The project team is happy to enter into discussions.

5. Practical Application of OS-APS

Regarding the application of OS-APS, there is not yet any finished user case because the software is still being developed. The tests with manuscripts from the ULB-SA do not map the entire workflow, but refer to individual components as well as the connections with OJS, OMP and DSpace.

However, when comparing OS-APS at its current state with other XML conversion tools, a few major differences can already be detected. Firstly, OS-APS is supposed to be a lean tool for the browser, so that it can be used without large initial efforts. Secondly, it does not aim to establish new standards, but instead relies on open standards that are already established in publishing houses (e.g., connection to OJS and OMP as widely used open source tools). This should make it easier for publishers to switch to a workflow that includes the use of OS-APS.

Thirdly, a unique feature of the software is that it aims to require as little technical know-how in the field of IT to work with it as possible. This way, user groups without much or any knowledge in XML or programming (this includes, e.g., most authors and editors) should be able to work comprehensively with OS-APS. The XML stays in the background behind an easy-to-use interface. The only case in which IT knowledge, in particular skills in web programming, would be needed is for the programming of entirely new templates or export formats.

The combination of these three aspects distinguishes it from other tools, e.g., those that focus purely on individual aspects such as XML editing (e.g., XML Copy Editor, https://xml-copy-editor.sourceforge.io, accessed on 6 December 2022, or Oxygen XML Editor, https://www.oxygenxml.com/xml_editor.html, accessed on 6 December 2022 as a more powerful tool) or those that support more holistic media-neutral publication processes with an XML extraction, editing and typesetting system, but rely on strong embedding in local processes and require in-depth technical know-how; e.g., in Germany Heidelberg Monograph PublishingTool (https://github.com/withanage/heimpt, accessed on 6 December 2022) or the XML-first typesetting system of OA-STRUKTKOMM (https://oa-struktkomm.htwk-leipzig.de/forschungsprojekt/publikationsserver/, accessed on

6 December 2022). Internationally, Kotahi (https://kotahi.community, accessed on 6 December 2022) could be functionally comparable to some extent, but it is less streamlined and with more functionalities, e.g., regarding peer review, while OS-APS can be used flexibly and as needed in or apart from existing publishing processes (e.g., use of OS-APS for normal, text-heavy manuscripts, while art volumes are set with InDesign in a classical way).

6. Conclusions

Preparing manuscripts for various formats such as HTML or EPUB can pose challenges for small- and medium-sized, as well as non-commercial (e.g., university) academic publishers. A high level of professionalism often requires extensive technical expertise as well as the use of cost-intensive XML content management systems.

The third-party funded project "Open Source Academic Publishing Suite (OS-APS)" provides relief in this area. It is intended to enable academic publishers to publish in a media-neutral way using XML-based workflows. The XML is automatically extracted from Word manuscripts and the corporate design of the exported PDFs can be controlled via templates. Institutions or publishers using OJS or OMP can also reuse the workflows and connections documented in the project. OS-APS is thus closely integrated into the open science landscape.

Author Contributions: Conceptualization, C.B., F.E., A.H. and M.P.; methodology, C.B., F.E., A.H. and M.P.; software, C.B. and F.E., developers at the ULB-SA; validation, C.B., F.E. and M.P.; formal analysis, C.B., F.E., A.H. and M.P.; investigation, C.B., F.E., A.H. and M.P.; resources, C.B., F.E. and M.P.; data curation, C.B., R.C. and F.E.; writing—original draft preparation, C.B., R.C., F.E., A.H. and M.P.; writing—review and editing, C.B., R.C., F.E., A.H. and M.P.; visualization, C.B., R.C., F.E., A.H. and M.P.; supervision, C.B., R.C., F.E., A.H. and M.P.; project administration, C.B., R.C., F.E., A.H. and M.P.; funding acquisition, C.B., F.E. and M.P. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data and software presented in this study are or will be made available on https://os-aps.de (accessed on 6 December 2022) within the specified project deliverable times.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Jeroen, B.; Jan Erik, F.; Bianca, K.; Pierre-Carl, L.; Vanessa, P. OA Diamond Journals Study. Part 1: Findings. *Zenodo* **2021**. [CrossRef]
2. Borchert, C.; Eichler, F.; Söllner, K.; Putnings, M.; Hoffmann, A.; Berghaus-Sprengel, A.; Brenn, D. Open Source Academic Publishing Suite (OS-APS): OA-Publikationen medienneutral mit automatisiertem Corporate Design erstellen. BIB-OPUS. 2022. Available online: https://nbn-resolving.org/urn:nbn:de:0290-opus4-178739 (accessed on 6 December 2022).
3. Arianna, B.; Jeroen, B.; Lars, B.; Jan Erik, F.; Bianca, K.; Pierre-Carl, L.; Pierre, M.; Vanessa, P. OA Diamond Journals Study. Part 2: Recommendations. *Zenodo* **2021**. [CrossRef]
4. Putnings, M.; Borchert, C.; Eichler, F. Project Open Source Academic Publishing Suite (OS-APS). 2021. Available online: https://librarypublishing.org/lpf21-lightning-talks/ (accessed on 6 December 2022).
5. Konstanze, S.; Markus, P.; Birgit, H.A.; Anke, B.-S.; Daniel, B.; Carsten, B.; Frederik, E. Open Source Academic Publishing Suite (OS-APS): Medienneutrales OA-Publizieren Im Eigenen Corporate Design. *Zenodo* **2021**. [CrossRef]

6. Söllner, K.; Putnings, M.; Hoffmann, A.; Berghaus-Sprengel, A.; Cozatl, R.; Brenn, D.; Borchert, C.; Frederik, E. Publikationen medienneutral und automatisiert gemäß den eigenen Stilrichtlinien erstellen mit der Open-Source-Software "OS-APS". 2021. Available online: https://dini.de/fileadmin/jahrestagungen/Jahrestagung_2019/DINI_2021_Poster_OS-APS.pdf (accessed on 6 December 2022).
7. Söllner, K.; Putnings, M.; Hoffmann, A.; Berghaus-Sprengel, A.; Borchert, C.; Eichler, F. Open Source Academic Publishing Suite (OS-APS): Simple, Media-Neutral OA Publishing with Automatic Typesetting. *SCS* **2021**, *4*, 1–8. [CrossRef] [PubMed]
8. Putnings, M.; Borchert, C.; Cozatl, R. Ein Einblick in Das BMBF-Projekt Open Source Academic Publishing Suite (OS-APS). *ABI Tech.* **2022**, *42*, 166–173. [CrossRef]
9. SciFlow. Milestones. 2022. Available online: https://gitlab.com/sciflow/development/-/milestone0073 (accessed on 6 December 2022).
10. OS-APS Demo. Available online: https://os-aps.de/demo/ (accessed on 6 December 2022).
11. Führ, F.; Bisset Alvarez, E. Digital Humanities and Open Science: Initial Aspects. In Proceedings of the International Conference on Data and Information in Online, Florianópolis, Brazil, 10–12 March 2021; Bisset Álvarez, E., Ed.; Springer International Publishing: Cham, Switzerland, 2021; pp. 154–173.

![publications logo]

Article

The Landscape of Scholarly Book Publishing in Croatia: Finding Pathways for Viable Open Access Models

Iva Melinščak Zlodi

Faculty of Humanities and Social Sciences, University of Zagreb, 10000 Zagreb, Croatia; imelinsc@ffzg.hr

Abstract: (1) Background: Open access to scholarly works is globally recognized as a goal to be achieved as soon as possible; however, there is not yet a general understanding of how to achieve open access for books. In considering the most appropriate models of transition, an accurate and detailed insight into national and regional specifics can be of great importance. The aim of this research is to show the current state of scholarly book publishing in Croatia: recognising the key stakeholders, their characteristics, and the current level of open access to scholarly books. (2) Methods: The existing data from two different sources were used: the data about the public subsidies for book publishers by the Ministry of Science and Education and the data on published books from the Croatian Scientific Bibliography CROSBI, both for the period from 2018 to 2021. (3) Results: In the four-year period, 224 Croatian publishers were awarded subsidies to publish 2359 book titles. The majority of the publishers received support for only a small number of titles and relatively low amounts of subsidies. More than half of the titles are published by small private commercial publishers. However, the uptake of digital publishing among commercial publishers is very modest. Open access to scholarly books is almost entirely in the domain of non-commercial publishers. Most open access titles are available on the websites of their publishers. (4) Conclusions: The analysis of the data from these two sources have resulted in an overview of the current state of book publishing in Croatia. Such an overview provides a good basis for designing future measures and creating a national open science plan and can also be a useful contribution to international discussions.

Keywords: Croatia; open access books; scholarly book publishing

1. Introduction

Open access (OA) to scholarly works is globally recognized today as a goal to be achieved as soon as possible. Currently, pathways for achieving OA for journals (in spite of multiple obstacles still being present) are far clearer than for books. Even organizations firmly oriented towards OA to all results of publicly funded research, such as cOalition S, acknowledge the complexity of book publishing and recognize that OA to books will require more complex models of realization over a longer period [1]. The central role of books in scholarly communication in humanities and social sciences has resulted in intensified discussions on possible models for achieving OA in recent years, especially within organizations such as OPERAS [2], Science Europe [3], SPARC Europe or Open Access Book Network [4]. These documents and discussions portray European book publishing as a fragmented space, with many smaller nationally oriented markets in which there is no domination of several large publishers. It is clear that the models of transition to OA will not be unique for the whole European area. Diverse mechanisms and sustainable business models will be appropriate for different contexts.

There are many studies and extensive research into the evaluation of books that have shown the importance of long-form publications (monographs and edited volumes) for scholarly communication, especially in some fields of scholarship. In the introductory overview of national landscapes studies, Giménez-Toledo et al. show that in social sciences and humanities, a substantial share of research outputs is published in monographs or

edited books, at least in several European countries (Norway, Belgium, UK, Spain, Denmark) [5]. However, as they point out, there is an absence of comprehensive international databases covering long-form publications, probably due to "an intrinsic heterogeneity of scholarly books themselves (e.g., disciplines, languages, formats, peer review and other editorial standards, etc.)", and such absence has prompted several European countries to develop their own custom-built information systems for the registration of scholarly books and publishers [5]. In a subsequently published paper resulting from the ENRESSH activities, the evaluative systems for 19 European countries were analysed, where 8 of them rely on categorization or a ranked list of book publishers: Denmark, Finland, Belgium, Norway, Poland, Slovakia, Slovenia, and Spain [6]. Such lists could possibly offer a good source of insights into the types of publishers of nationally relevant book publishers and possible models for achieving OA for books.

Although on the international level, there is a strong and, in some areas, well-coordinated effort to transition book publishing to OA, it is not equally applicable to all national landscapes, especially in the areas that are not dominated by major international publishers, and where books are published mostly in national languages.

To date, several studies have shed some light on the national landscapes of scholarly book publishers. A key study in this area was the one performed within the Knowledge Exchange organization on the topics of the inclusion of OA monographs in OA policies, funding streams to support OA monographs, and business models for publishing OA monographs [7]. The study presents a very clear and detailed overview of developments in the OA books arena, and includes county studies for countries that were, at the time, members of the Knowledge Exchange group (Denmark, Finland, Germany, Netherlands, the United Kingdom, and France) with the addition of Norway and Austria. Importantly, Ferwerda et al. recognize the significance of national contexts ("country size, language(s) of publication, presence of multinational corporations and socio-economic cultures of countries") [7]. However, not all parts of Europe are portrayed in this study, as it completely lacks the countries from Eastern Europe. Within regions that are portrayed, some countries with important OA books development are not covered (for instance, Sweden, from the Nordic region). Within the countries covered, significant differences were observed with respect to the key stakeholders involved, incentives for OA, available public funding and available joint infrastructures, types of books published and the audience expected, as well as peer review practices. Publishers are never a homogenous group, not even on the county level. In some countries, especially if there is a funder's mandate and available funds for book publishing charges (BPCs), the commercial publishers will take a lead in OA, whereas in others, the learned societies or institutions could have a larger role. The study ended in 2017, and some important developments took place after that (most notably, under the influence of Plan S).

Few additional studies have provided more detailed or more recent insights into the country-specific developments in Sweden [8,9] or Finland [10,11]. The Finnish studies are especially interesting, as they enable a comparison of national journal publishing and national book publishing.

Similarly, a study by Horvat and Velagić on the Croatian publishing landscape for the period 2012–2018 stresses the difference between journals and books [12]. According to the authors, both journals and books in Croatia are largely dependent on public subsidies, but all journals funded by the government are available in OA, whereas books are dominantly not. Moreover, public subsidies for journals are overwhelmingly granted to public institutions or associations, while most recipients of book subsidies are private publishers. In the studied period, only 1.25% of book titles were available in digital format, which clearly indicated that publishers had not seen e-books as a viable business. Furthermore, the authors notice a lack of expertise in peer review, database indexing and OA availability among private publishers. They conclude with the observation that the current subsidy system for books does not promote development and is not successful in enhancing the availability of scholarly e-monographs.

Both at the national and international levels, many important elements of the scholarly OA books landscape are still unknown, not just those related to business models, but also to the prevalence of OA books, their visibility, discoverability and preservation. These were the issues addressed by the recent study *Open access books through open data sources: Assessing prevalence, providers, and preservation* [13]. As Laakso notices, there is no single data source that could comprehensively collect and expose metadata on all OA books, and combining or deduplicating records from multiple sources is difficult for various reasons (most notably, the inconsistent use of persistent identifiers, and using multiple ISBNs for different manifestations of books). One of the results of the study is the insight into the distribution of web domains that offer full-text access to OA books, deduced from the available DOIs and the URLs that they resolve to. According to Laakso, there are several dominant domains, followed by a long tail of smaller websites, including some "clearly volatile services" such as institutional webpages or Dropbox and Google Drive [13]. Moreover, it needs to be pointed out that these are the results of investigating books that have DOIs assigned by their publishers, whereas the results for books without DOIs would likely show us an even more worrisome distribution of hosting domains, with clear implications on the discoverability, quality and preservation of OA book content. Although there are significant international developments, and important OA book infrastructures are already in place [14], they are not equally accepted and employed throughout Europe.

In considering the most appropriate models of transition to OA book publishing, an accurate and detailed insight into individual national and regional specifics can be of great importance. The aim of this research is to show the current state of scholarly book publishing in Croatia: recognise the key stakeholders, their characteristics, and the current level of OA to scholarly books. The existing data from two different sources will be used for this purpose.

This study addresses the following research questions: (1) What type of publishers publish scholarly books in Croatia and what are their shares in overall scholarly book production?; (2) How prevalent are OA books and who are the publishers of OA books?; and (3) What is the preferred model for OA books and where are they hosted?

2. Methods

The main source of funding to cover the costs of scholarly book publishing in Croatia are the direct subsidies from the Ministry of Science and Education (MSE). Given that data on grant recipients are publicly available, it is possible to gain insight into who publishes scholarly books and what are the main types of book publishers. The analysis was based on data from a recent period (2018–2021). The insights from the previously published analysis of the same funding scheme conducted by F. Horvat and Z. Velagić [12] in the earlier period were also considered.

All of the results of public calls for subsidies are publicly available on the website of the Ministry of Science and Education [15]. The documents are published in a pdf format, containing the information on publisher, book title, authors or editors, amount requested and amount approved. For the purpose of this research, the files were downloaded and converted to spreadsheets. The names of publishers were sometimes used inconsistently; therefore, they were cleaned and unified in order to get a list of unique values.

There are some limitations to this dataset (*MSE dataset*):

- There is information on co-publishers for only nine records. However, from the other dataset used (CROSBI, the Croatian Scientific Bibliography), it is visible that co-publishing (often between for-profit and institutional or society publishers) is a very common practice. In the analysis of the MSE data, the co-published books were regarded as published by the first indicated publisher.
- From the available data, for some books, it is not clear if the responsible persons are authors, editors, translators or series editors.
- Furthermore, sometimes it is not possible to discern the type of books (monographs, edited volumes, university textbooks, critical editions, reference works, works for the

popularization of science or translations). Therefore, all of the books on the lists were included in the analysis.

The publishers in the MSE dataset are grouped according to the typology of publishers based on previous studies and adapted to the Croatian publishing landscape. In *A Landscape Study on Open Access and Monographs*, a distinction was made between for-profit and non-profit publishers, and between traditional university presses, new university presses, and academic-led presses [7]. In a study by Late et al., the publishers were divided into the following types: learned societies, universities and university presses, other research organizations, commercial publishers, and other publishers [10]. Horvat and Velagić used the distinction between for-profit private publishers ("privately owned legal entities registered as companies, crafts, or cooperatives") and public publishers (institutions, associations, art organizations, local government and religious organizations) [12].

In the course of this study, the following types or groups of publishers were recognized in the sample and used for further analysis:

1. small or medium-sized private companies, for-profit commercial publishers (SME);
2. higher education institutions: universities, faculties, academies (HEI);
3. research institutes (RES INST);
4. public academic organizations: academies, centres, non-governmental organisations with academic character (PUB ACAD);
5. professional and scholarly associations or learned societies (SOC);
6. others: public bodies, religious organisations, archives, museums, libraries, non-academic public institutions, or associations (OTH).

Another useful source of data on the books published by Croatian authors can be found in CROSBI, the Croatian Scientific Bibliography [16] (*CROSBI dataset*). The data on scientific books (monographs and edited books) published in the same period (2018–2021) were reviewed and from the results, it was possible to obtain additional information about book publishers who are currently active in Croatia, but also about the existence of e-editions, and especially e-editions in OA. The information on different models of OA was particularly useful: OA books on publishers' platforms ("gold" model) and in open repositories ("green" model). Based on this analysis, we could gain insight into the preferred mode of OA for different types of publishers. For open books available on publishers' platforms, we could find out to what extent they meet some of the standards of digital publishing (use of persistent identifiers, standardized metadata and discoverability).

All metadata from CROSBI are publicly available, under the terms of the Creative Commons BY-NC-SA licence, in multiple formats. For the purpose of this study, two csv files were downloaded in September 2022, containing the records on authored books and edited volumes published in the period 2018–2021. The records in CROSBI are deposited by authors–researchers, and (lightly) controlled and edited by the administrators. Extensive cleaning of the downloaded records was performed to eliminate, as much as possible: books published outside Croatia (based on the place of the publisher), books of abstracts, translations, books labelled as 'non-scholarly', exhibition catalogues without peer review and brochures (with less than 30 pages).

It is important to note that the two datasets, MSE and CROSBI, were not combined, but were used to answer different research questions. Furthermore, although both sets refer to the period 2018–2021, the MSE dataset contains titles that were planned in those years (some of the listed titles from recent years have not yet been published), while CROSBI contains records on books already published.

Both sets of the collected data are available as open datasets.

3. Results

3.1. Publishers of Publicly Subsidised Books

Academic book publishing in Croatia has traditionally been largely dependent on public subsidies and continues to be so. There are diverse sources of public funding available to publishers and authors (county or city offices for culture, public and private

endowments, Ministry of Culture, and others), but the most significant one is the Ministry of Science and Education (MSE). The Ministry regularly announces annual calls for grants to interested publishers (all types of publishers are eligible). The subsidies are awarded to publishers for individual book titles (scholarly books and higher education textbooks), based on the pre-published quality criteria, and according to the assessment performed by the Ministry's Committee on Scholarly Publishing. All publishers who apply should provide at least two independent peer review reports with each manuscript. Subsidies are awarded to publishers of print and digital (or hybrid) editions, without distinction. The Ministry's quality criteria do not include OA requirements or recommendations. The annual call for financial support by the MSE is an important point in the publishing cycle of all academic publishers in Croatia, and the MSE's application criteria have a strong influence on the publishers' practices.

3.1.1. Awarded Public Subsidies

The data published by the MSE provide insight into the amounts and the pattern of public investment. In the 2018–2021 period, the total amount of MSE subsidies for books varied from approximately EUR 1 to 1,5 million (a significant decline occurred in 2020), which is a continuation of the period from the 2015 onwards [12].

Between 550 and 610 titles per year were subsidized (Table 1).

Table 1. Total amount of all awarded subsidies and a number of subsidized titles in the period 2018–2021, according to MSE dataset. (at the time of the analysis, EUR 1 equals HRK 7.5345.)

Year	Awarded Subsidies	No of Subsidized Titles
2018	HRK 11,949,706.00	602
2019	HRK 11,338,645.00	550
2020	HRK 7,172,131.00	597
2021	HRK 11,808,206.00	610

The average amount awarded per title is HRK 17,918.05; however, it varies considerably, from HRK 1323.00 to HRK 117,117.00.

3.1.2. Size of Publishers (According to Number of Titles and Received Funding)

From the MSE set, we can see that in the four-year period (2018–2021), 224 Croatian publishers were awarded subsidies to publish 2359 book titles. Most of the publishers received support for only a small number of titles (Table 2), while only three received subsidies for 100 or more titles, indicating that most Croatian publishers are either small (in terms of number of published titles), or scholarly books make up only a fraction of their publishing portfolios.

Table 2. Number of publishers according to the number of subsidized titles (from the MSE dataset) in the period 2018–2021.

No of Subsidized Book Titles	No of Publishers
>100	3
50–99	8
10–49	52
2–9	86
1	75

Eleven publishers who received funding for more than 50 titles are: Filozofski fakultet Zagreb, Medicinska naklada, Hrvatska sveučilišna naklada, Školska knjiga, Srednja Europa, Matica hrvatska, Jesenski i Turk, Naklada Ljevak, Hrvatska akademija znanosti i umjetnosti, TIM press, and Sveučilište u Zadru.

Similarly, most publishers received relatively low amounts of subsidies, while a small minority received more than HRK 1 million (Table 3)

Table 3. Number of publishers according to the total amount of awarded subsidies (from the MSE dataset) in a period 2018–2021. (at the time of the analysis, EUR 1 equals HRK 7.5345.)

Total Awarded Subsidies	No of Publishers
HRK > 2,000,000.00	3
HRK 1,000,000.00–2,000,000.00	7
HRK 100,000.00–1,000,000.00	72
HRK 10,000.00–100,000.00	119
HRK < 10,000.00	23

Ten publishers who received more than HRK 1,000,000.00 in subsidies are: Filozofski fakultet Zagreb, Medicinska naklada, Hrvatska sveučilišna naklada, Školska knjiga, Matica hrvatska, Hrvatska akademija znanosti i umjetnosti, Srednja Europa, Književni krug Split, Hrvatski institut za povijest, and Fakultet hrvatskih studija.

The same three publishers who received the largest amounts in subsidies were the ones with the largest number of planned titles.

3.1.3. Types of Publishers

When we look at the publishers who were receiving subsidies from the Ministry and group them according to the types (as they are listed and defined in the methodology section), we can see that small or medium-sized private companies are the largest group (SME), followed by the higher education institutions (HEI) and professional and scholarly associations or learned societies (SOC). A smaller number of publishers are from other types of organizations, including public bodies, religious organisations, archives, museums, libraries, non-academic public institutions or associations (OTH), research institutes (RES INST) and public academic organizations: academies, centres, non-governmental organisations with academic character (PUB ACAD) (Table 4).

Table 4. Number of publishers (from the MSE dataset) according to their types.

Publisher Type	No of Publishers	% of All Publishers
SME	77	34%
HEI	61	27%
SOC	38	17%
OTH	20	9%
RES INST	14	6%
PUB ACAD	14	6%

Although this principle of grouping was considered adequate for the analysis in this study, it could be pointed out that there is further diversification within many of the listed groups. Some of the private publishers are dominantly specialized in textbooks or translations. Others are closer to what is today called 'scholar-led', and very much mission-oriented (as well as less profitable). For some of the learned societies, publishing is their core business, while for many, it is only a supplemental activity.

When we look at the distribution of approved titles and awarded subsidies, we observe even greater shares of private commercial publishers (SMEs) compared to other types. More than half of all titles and financial grants received are in the domain of private publishing (Tables 5 and 6).

Table 5. Number of titles (and shares in total number, from the MSE dataset) according to the publisher types.

Publisher Type	No of Subsidized Tiles	% of All Subsidized Titles
SME	1250	53%
HEI	568	24%
RES INST	192	8%
PUB ACAD	153	6%
SOC	138	6%
OTH	58	2%

Table 6. Amounts of awarded subsidies (and shares in the total amount awarded, from the MSE dataset) according to the publisher types.

Publisher Type	Awarded Subsidies	% of Total Amount
SME	HRK 21,866,528.00	51.73%
HEI	HRK 9,379,605.00	22.19%
RES INST	HRK 3,963,749.00	9.38%
PUB ACAD	HRK 3,157,243.00	7.47%
SOC	HRK 2,924,828.00	6.92%
OTH	HRK 976,735.00	2.31%

3.2. Open Access for Scholarly Books in Croatia

The MSE data enabled us to gain a certain overview of the main stakeholders in publicly supported book publishing in Croatia and the patterns of public spending. However, this dataset does not contain any information on the format of the published books, whether they are available in print, online or both. In addition, if they are available online, there is no information on whether they are available as paywalled books (via subscription or purchase) or as books in OA.

More information on the format is available in the CROSBI dataset. Among 2674 records for books published in the period 2018–2021, there were 188 records with information in the field 'URLS' (intended to provide URLs for editions other than OA). However, on further inspection, it was established that the majority (186) of these URLs were either not valid anymore, were resolved to descriptive-only landing pages (without full text), or were in fact OA editions.

For 427 records, the URLs for OA editions were provided.

3.2.1. Commercial Publishers and Open or Paywalled e-Books

The analysis of the MSE dataset uncovered the importance of commercial private sector in the scholarly book publishing in Croatia. However, given the overall context of book publishing in a small national circle with limited readership, it could not have been expected that digital innovation would be taking place among private publishers on a significant scale or that it could be fostered by their commercial interests.

The CROSBI dataset confirms that the uptake of digital publishing among commercial publishers is very modest. Only two such publishers offer paywalled scholarly e-books. One is the publisher of books specialized in business and economics that offers institutional subscriptions to e-books in e-pub format, but without the usual access control mechanisms (based on IP ranges, shibboleth protocol or proxy access). The other one offers only the option of individual purchases (predominantly titles in popular science).

The evidence of OA models for books among Croatian private publishers is even scarcer. In the CROSBI sample, only two titles in OA were recorded. One is from a small niche publisher: there is a pdf document on the publisher's website, but without associated metadata. The other is from a larger publisher specialized for textbooks: there is a copy in an institutional repository self-archived by the editor. In both cases, the initiative for making books open clearly came from the author or the editor, and not the publisher.

3.2.2. Publishers of Open Access Books in Croatia

According to the CROSBI data, OA to scholarly books is almost entirely in the domain of non-commercial publishers: higher education institutions, scholarly societies, research institutes and other public organizations (government offices and NGOs).

Four publishers with higher number of recorded titles in OA (if the editions co-published by several organizations were excluded) are:

1. Filozofski fakultet Sveučilišta u Zagrebu (35 titles on two platforms, 'FF Open Press' and 'Stilistika');
2. Sveučilište u Zadru (21 titles on the platform: 'Morepress');
3. Filozofski fakultet Sveučilišta u Rijeci (14 titles on the website of 'Izdavaštvo FFRI');
4. Sveučilište Jurja Dobrile u Puli (13 titles deposited in the institutional repository).

All four are higher education institutions (faculties or universities).

In addition to looking at who is providing OA books, it is also interesting to look at the models of delivery (Table 7).

Table 7. Models of providing OA to books recognized in the CROSBI dataset.

OA Host	Number of Records	% of Total No
Web site	280	66%
Publishing platform	46	11%
Scholarly networking site	42	10%
Repository	38	9%
General hosting service	15	4%
Digital library	6	1%

In some of these models, OA is accomplished through the effort of the publisher; in others, the author or the editor will take that responsibility.

Only 11% of titles in the sample are published in OA on the specialised publishing platforms maintained by their publishers. Both such platforms (Morepress and FF Open Press) are developed by using the open-source software PKP Open Monograph Press, and they enable efficient dissemination of full text content and standardized metadata through established discovery channels.

An additional 10% of titles are available in digital libraries (1%) and institutional repositories (9%) that provide interoperability of standardized metadata. However, it is not clear from the data if the books are self-archived to the repositories by their authors/editors, or they are deposited by the publishing institution (examples of both cases are known to the author of this paper).

A non-negligible share of books (10% in the sample) is available in OA due to their authors sharing them via scholarly networking sites, such as Academia.org or ResearchGate. Although it is disputable whether this is an actual OA model, it is indicative of the need of the authors to provide wider access and secure readership and visibility.

Fifteen titles (4%) are made available on several general hosting services, free or paid (Google Drive, Issuu, Scribd, Dropbox).

However, most titles are available on the institutional websites of their publishers. Generally, that means that there are no standardized landing pages, metadata dissemination or compliance with interoperability standards.

Further evidence of slow acceptance of professional standards of digital publishing is the low number of Digital object identifiers (DOIs) registered by Croatian publishers: in our sample we found only 32 records with DOIs from 3 institutional publishers (one publisher had 2 titles with DOIs, one had 5 and one had 25).

However, it should be recalled that this analysis of OA titles and their hosts is based on a limited sample of 427 records for which the OA URLs were provided in the CROSBI dataset.

4. Discussion

The fact that most of the Croatian scholarly book publishing is happening in the private sector has a considerable impact on the way the future OA books' infrastructure should be designed and implemented. The existing and available national scholarly infrastructure is currently primarily intended for academic institutions, and some mechanisms (for instance, the authentication based on the eduroam identities) prevent the inclusion of non-academic stakeholders. However, the Croatian repositories infrastructure could be, with some effort, adapted for the needs of book publishing. Functionalities that would need to be added are, for instance, the existence of tables of contents and relations from book title to book chapters, or the possibility to distinguish between books published by the institution from the institution's research output in the institutional repositories.

Unlike large international publishers who are incentivised to innovate and operate in a digital environment, Croatian private commercial publishers are still very traditional and slow in embracing newer technologies. Several reasons for that could be identified: a fragmented market with many small players (and each of them without individual capacities and expertise), low collaboration, and a small market (limited readership of academic content in the national language). Currently, several initiatives are found among higher education institutions that are actively and thoughtfully approaching OA book publishing. However, such OA publishing, entrenched in the public sector and dependent on in-kind contributions and public subsidies, has its own challenges and limitations, as described in the collection of case studies detailing the business models of a range of OA academic book presses [17].

The pattern of governmental subsidies has shown both stability and dedication (the amount awarded has stayed in the same range over years) as well as the volatile nature of the system (a single drop of more than 1/3 of the total amount in one year shows that the scheme can easily collapse due to a change in government or a financial crisis).

It is important to note that financial difficulties are not the only challenge for the sustainability of academic book publishing. Expertise, knowledge, and experience in all areas of digital scholarly publishing are lacking among most Croatian book publishers. That becomes even more obvious from the information on hosting domains, or from the prevalence of permanent identifiers. The inability to comply with the established interoperability standards is seriously compromising the visibility and discoverability of book content, despite its open availability. Collaboration and developing joint national infrastructure and support systems seem to be the only ways forward for overcoming the existing shortcomings.

Many international debates on OA to scholarly books focus on new business models, book publishing charges (BPCs) or innovative (often collective) funding mechanisms, such as crowdfunding, "opening the future" or library memberships. There is no evidence of any Croatian academic publishers trying out any of the mentioned business models. Although some of the experiences from those models could surely be instructional even for Croatian publishers, the overall landscape of scholarly communication (including the way the research activities and libraries are funded, as well as the dominant language and readership of books) makes it highly unlikely that they could be fully adopted and relied upon. The current system of public subsidies still appears to be the most rational and efficient way of supporting academic books, and the MSE's requirements for awarding subsidies are the main area where improvements could be sought.

One very important distinction of the Croatian landscape, as opposed to the countries that have already achieved a more successful transition to OA books (such as the cases of Austria, Sweden, or the UK), is a lack of a clear policy from the funder (either from the research funder, the Croatian Science Foundation, or the grantor of public subsidies to publishers, the Croatian Ministry of Science and Education). Currently, there is no national open science or OA policy, and no requirement from the main national research foundation that would be binding for book authors. Hopefully, that will change in the near future, as the result of the work of the Croatian Open Science Cloud Initiative *(HR-OOZ)* [18].

The HR-OOZ working group is currently defining a future Croatian national OS plan that would also address books.

The results of this research provided us with some important insights into the landscape of book production in Croatia, especially regarding the publishing of OA books. However, there are also certain limitations of the chosen approach, which primarily relate to the features of the data available for analysis. The research used already existing publicly available data that could provide an overview and typology of the main stakeholders and ways of achieving OA; however, these were insufficient to reveal the motives and challenges faced by individual publishers or groups of publishers. Since knowing the motivations and barriers to achieving OA among publishers can help us identify the most effective ways of transition, it would be extremely important to conduct research in the future that would examine the views of publishers using a survey, qualitative interviews or focus groups.

5. Conclusions

By analysing data from two sources, the MSE and CROSBI, we formed an overview of the current state of book publishing in Croatia. It has become clear that most of the books subsidized by the government are published by small and medium-sized for-profit publishers. At the same time, private publishers hardly ever engage in publishing OA books; therefore, OA to scholarly books is almost entirely in the domain of non-commercial publishers: higher education institutions, scholarly societies, research institutes and other public organizations. Most of the OA books (66%) are hosted on the institutional websites without standardized metadata. A smaller number is available on specialised publishing platforms (11%) or hosted in institutional repositories (9%). Such an overview gives us a good basis for defining future measures and designing a national open science plan that is feasible and realistic with regard to scholarly books. It could also be a useful contribution to international discussions.

The information gathered has shown us the importance of public policies and conditional funding that will require or at least reward OA. To be sustainable, the public funding model needs to promote development, to encourage change, foster collaboration and incentivise adoption of the international standards.

Furthermore, this landscaping exercise made apparent the weaknesses of book publishing infrastructure and showed us another area where major improvements could be envisaged and planned. The importance of public, available and reliable publishing infrastructure needs to be considered. The development of common national infrastructure for OA book publishing could use and adapt the models developed elsewhere in Europe, or even use some of the lessons from the development of the Croatian journal publishing platform, Hrčak. The need to be included in the wider international networks and infrastructures, such as OAPEN or DOAB, will be the key in securing visibility and discoverability.

References

1. COAlition S Statement on Open Access for Academic Books. 2021. Available online: https://www.coalition-s.org/coalition-s-statement-on-open-access-for-academic-books/ (accessed on 1 May 2022).
2. Stone, G.; Błaszczyńska, M.; Lebon, C.; Morka, A.; Mosterd, T.; Mounier, P.; Proudman, V.; Speicher, L.; Melinščak Zlodi, I. *Collaborative Models for OA Book Publishers*; Zenodo: Geneve, Switzerland, 2021. [CrossRef]
3. Science Europe. *Briefing Paper on Open Access to Academic Books*; Zenodo: Geneve, Switzerland, 2019. [CrossRef]

4.   A Plan S for Books? *Voices from the Open Access Books Community: Sessions Summary*. 2021. Available online: https://hcommons. org/deposits/item/hc:41205/ (accessed on 1 May 2022).

5.   Giménez-Toledo, E.; Mañana-Rodríguez, J.; Engels, T.C.E.; Ingwersen, P.; Pölönen, J.; Sivertsen, G.; Verleysen, F.T.; Zuccala, A.A. Taking scholarly books into account: Current developments in five European countries. *Scientometrics* **2016**, *107*, 685–699. [CrossRef]

6.   Giménez-Toledo, E.; Mañana-Rodríguez, J.; Engels, T.C.E.; Guns, R.; Kulczycki, E.; Ochsner, M.; Pölönen, J.; Sivertsen, G.; Zuccala, A.A. Taking scholarly books into account, part II: A comparison of 19 European countries in evaluation and funding. *Scientometrics* **2019**, *118*, 233–251. [CrossRef]

7.   Ferwerda, E.; Pinter, F.; Stern, N. *A Landscape Study on Open Access and Monographs: Policies, Funding and Publishing in Eight European Countries*; Zenodo: Geneve, Switzerland, 2017. [CrossRef]

8.   Lawrence, D.; Bernhardsson, K.; Eriksson, J.; Henning, K.; Hertil Lindelöw, C.; Neidenmark, T.; Pelling, M.; Svensson, A. A National Consortium for Open Academic Books in Sweden: An Investigation Funded by National Library of Sweden, Swedish Research Council and Riksbankens Jubileumsfond. 2013. Available online: http://urn.kb.se/resolve?urn=urn:nbn:se:liu: diva-94273 (accessed on 19 November 2022).

9.   National Library of Sweden. Open Access to Books: A Study within the Remit of the National Library of Sweden's National Coordination Mandate for Open Access to Scholarly Publications (Study 3). 2021. Available online: https://urn.kb.se/resolve? urn=urn:nbn:se:kb:publ-83 (accessed on 17 December 2022).

10. Late, E.; Korkeamäki, L.; Pölönen, J.; Syrjämäki, S. The role of learned societies in national scholarly publishing. *Learn. Publ.* **2020**, *33*, 5–13. [CrossRef]

11. Pölönen, J.; Laakso, M.; Guns, R.; Kulczycki, E.; Sivertsen, G. Open access at the national level: A comprehensive analysis of publications by Finnish researchers. *Quant. Sci. Stud.* **2020**, *1*, 1396–1428. [CrossRef]

12. Horvat, F.; Velagić, Z. Supply-Side Model of Academic Publishing in Croatia (2012–2018). *Knygotyra* **2020**, *75*, 66–91. [CrossRef]

13. Laakso, M. *Open Access Books through Open Data Sources: Assessing Prevalence, Providers, and Preservation*; Zenodo: Geneve, Switzerland, 2022. [CrossRef]

14. Stern, N. A brief saga about open access books. *Nord. Perspect. Open Sci.* **2021**, *6*. [CrossRef] [PubMed]

15. Republika Hrvatska Ministarstvo Znanosti i Obrazovanja. Znanstveno-Izdavačka Djelatnost. 2022. Available online: https: //web.archive.org/web/20221220180509/https://mzo.gov.hr/istaknute-teme/znanost/znanstvena-infrastruktura/znanstveno-izdavacka-djelatnost/146 (accessed on 20 December 2022).

16. Institut Ruđer Bošković. CROSBI Hrvatska Znanstvena Bibliografija. 2022. Available online: https://www.bib.irb.hr (accessed on 1 May 2022).

17. Morka, A.; FF Open Press. Business Models for Open Access Books. 2022. Available online: https://web.archive.org/web/2022 1220180236/https://oabooksbusinessmodels.pubpub.org/pub/ff-open-press/release/1 (accessed on 16 December 2022).

18. Hrvatski Oblak za Otvorenu Znanost (HR-OOZ). 2022. Available online: https://www.srce.unizg.hr/hr-ooz (accessed on 20 December 2022).

MDPI

Article

Characteristics of European Universities That Participate in Library Crowdfunding Initiatives for Open Access Monographs

Mirela Roncevic

Faculty of Humanities and Social Sciences, University of Zagreb, 10000 Zagreb, Croatia; mirelaroncevic@gmail.com or mironcevi@ffzg.hr

Abstract: The aim of the study was to identify the traits of 100 European universities across 26 countries that did or did not support one particular library crowdfunding initiative for open access (OA) monographs over the past few years. By relying on the rankings of four sources, including THE, ARWU, QS, and Leiden, the study identifies some of the traits of the universities that have shown strong interest in the model by already taking part in an established library crowdfunding initiative, as well as those that may play a vital role in its sustainability. The study's results show that the institutions that are likely to participate in library crowdfunding initiatives for OA monographs may be defined as highly ranked and produce research in quantity, quantity, relevance, and timeliness. The study's key revelation is the high academic standing of the institutions that rarely participate in one crowdfunding initiative. These institutions may not be as "international" in their outlooks, but they stand out for their high-quality and significant research output. As such, they may accelerate the model's adoption with more consistent participation in library crowdfunding.

Keywords: open access publishing; open access monographs; open access scholarly books; library crowdfunding; open access business models; sustainability of open access business models; sustainability of open access monographs

1. Introduction

Since its beginnings, the open access (OA) movement has been growing in many ways. It was designed to remedy the "perceived failings" of the broken publishing system, which mainly concerned the rising prices of journal subscriptions [1]. For this reason, the story of OA began with journals and institutional repositories (IRs), with some scholars still associating the story of OA with journals exclusively. However, other types of scholarly content have since been published OA, including, for example, conference proceedings, scholarly videos, scholarly blogs, and monographs.

In 2015, Pinfield identified the dominant themes in the analysis of OA publishing: uncertainties around the green vs. gold OA possibilities; the development of evidence to inform OA discussions; researchers' disinterest in and skepticism about OA; policies or so-called "mandates" that encourage OA; OA infrastructure (i.e., repositories); the emergence of OA journals; the OA-related challenges for institutions; and the impact of OA content beyond citation scores [2]. One notices the absence of academic books (i.e., monographs) after examining these early themes. While studies were already underway that considered the possibilities of OA monographs at the time [3,4], they were in their early stages. The idea of publishing monographs OA, in fact, is still considered to be in the early stages [5].

Therefore, academic journals and IRs were the first "delivery vehicles" that dominated the distribution of OA content and the low-hanging fruit of the OA movement from its start [6]. This was because journal authors had little to lose from publishing OA. Unlike article authors, book authors receive advance payments and royalties for monographs. Although these payments may not be as high as the advances and royalties earned in commercial publishing, book authors do not want to lose these royalties. The fact that

authors do not receive financial compensation for publishing articles means that there is little interest in protecting any income by restricting access to articles. This, however, is the key trait of print monograph publishing [7]—access must be restricted only to those that purchase it or buy rights to access it, such as libraries. In addition, journal articles are much shorter than monographs, which makes them less risky investments. Conversely, monographs involve significant assets [8]. Owing to the sheer number of stakeholders involved in the production of monographs, the business models that would make the publishing of OA monographs possible needed an entirely different approach [9].

1.1. The Emergence of OA Monographs

The monograph—which may be described as "a long, academic and peer-reviewed work on a single topic normally written by a single author, and also extended to include peer-reviewed edited collections by multiple authors" [10]—is a vital medium through which the humanities achieve impact [11]. Authored books have an important place in the humanities and social sciences (HSS) [12]. Further, the publishing of monographs through university presses has formed an essential component of the tenure evaluation for scholars [5]. Moreover, the monograph has remained "the" vehicle for articulating arguments from extensive research. As Cheshire puts it, "If the academic monograph is no longer valued, why do we require an 80,000-word thesis from doctoral students?" [13]. In short, publishing a thesis as a printed monograph is still "the proxy for being recognized as serious researchers" [14].

Notwithstanding its relevance in the HSS community, the monograph has been in a state of crisis since the turn of the century [5], or since the OA movement began to shift the direction of academic publishing. This is one reason why OA monographs have taken a while to catch up to OA journals [15,16]. The landscape of the business models for OA monographs is now expanding owing to the demand from scholars, funders, and the public to make scholarship widely available [17]. Today, OA publishing has become an integral part of the businesses of publishers—corporate and nonprofit alike—including university presses. In fact, new kinds of university presses have begun to emerge that are committed to OA and "born digital" content [18].

As the Ithaka S + R Faculty Survey recently showed, scholars now recognize that the importance of the print monograph has declined, and that monographs in digital format remain essential for their teaching and research [19]. According to a recent study published in the *Journal of Librarian and Information Science*, academics' awareness of OA increased significantly during the pandemic [20]. In mid-2019, the Directory of Open Access Books (DOAB) listed fewer than 20,000 OA books, compared with some 86,000 monographs published internationally every year [21]. In late 2022, the DOAB listed 62,500 OA books, which is a significant jump in a relatively short period. A strong argument could be made that the limiting conditions of the pandemic contributed to the increase in the awareness of the need to accelerate OA publishing, including monographs.

There is a growing interest in European countries in investing in OA monographs. The Plan S initiative for OA publishing is a good example. Launched in September 2018 by cOAlition S, a global consortium of research funding organizations, Plan S advocates that scientific publications that are funded by public money must be published OA. As stated on the European Science Foundation's website, cOAlition S's main principle is tied to the following statement: "With effect from 2021, all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo" [22].

In its current statement on open access for academic books, the coalition states that it "recognizes that academic book publishing is very different from journal publishing. Our commitment is to progress towards full open access for academic books as soon as possible, in the understanding that standards and funding models may need more time to

develop" [22]. Further, the UK Research and Innovation funding council announced its OA policy in the summer of 2021. It states that academic books, including monographs and edited chapter books, must be published OA from January 2024, with a permissible one-year embargo. This development has resulted in UK researchers questioning the ramifications of OA book publishing [16].

Although their OA funds have primarily focused on OA journals, librarians have generally supported the progress of OA monographs. The OAPEN-UK 2014 librarian survey was among the first that revealed how positive librarians were about OA monograph publishing in its nascent phase [4]. The same survey showed that 80% said that they would support OA monograph publishing as a matter of principle, and that they were willing to fund the publishing of OA monographs in the face of uncertainties, lack of experience, and pressures imposed on them by the rising costs of monographs. A 2022 survey of European librarians also confirmed their interest in supporting OA monographs as a matter of principle [23].

1.2. Business Models for OA Monographs

Various business models have been tested to determine how to publish OA books in financially viable ways for authors, publishers, and researchers, and that do not call into question the integrity of either authors or publishers [24]. Collins et al. [3] first identified the following business models for OA monographs: the author payment model (the author pays a fee to the publisher, known as a book processing charge, or BPC); the selective open access model (other activities of the press subsidize monograph publishing); the collaborative underwriting model (libraries join forces to meet the price of a publisher set for a title to become OA and share the cost); the crowdfunding model (publishers pitch a title and seek funding from the "crowd", which can include individuals or institutions); the embargo/delayed OA model (a monograph is released OA after a publisher has had time to gain revenue from the sale of the title); the new university press model (new university presses emerge at institutions and receive subsidies for publishing); the freemium model (the basic version of the monograph is free, while the premium version, with more content and features, is sold in a different format).

Speicher et al. [25] also divided the OA models into several groups, some of which overlap with Collins' approach: the author processing charge (APC) model (authors pay publishing fees); the freemium model (one simple version of the work is free, while others are not free); the collaboration model (institutions join forces to open knowledge globally); the community model (researchers in specific disciplines join forces with the common goal of making the literature in their field OA); the library model (libraries cover the cost of OA publishing).

To date, no model has become dominant, as each faces unique challenges [25]. The APC model, for example, recognizes the costs behind quality publications that need to be financed, but little funding is available to HSS scholars. The freemium model works to generate extra revenue for the publisher, but it remains unclear if it is beneficial in the long run. The collaboration model brings together communities with similar views and goals, but it must prove its sustainability. The community model brings publishing to the academic community, but funding and resources remain an issue. If budgets are not an issue, then the library model works well with the existing library workflows and distributes funds similarly to how funds are allocated for subscriptions. However, it can only succeed if libraries continuously set aside funds for OA monographs.

Informed by these two breakdowns, OAPEN published its list of updated business models for OA book publishing in 2022 [26]: book processing charges (BPCs); the freemium model; institutional subsidies/new university presses; library membership; library consortiums (institutional crowdfunding); subscribe to open (libraries subscribe to collections of closed-access books, and the subscription fees are used to fund OA for new books); crowdfunding (individuals pledge fees to make a book OA). "Library crowdfunding"—the term used in this paper—was also identified as a specific revenue model two years earlier

in a COPIM report and defined as an intermediating platform "connecting publishers to 'unlock' or 'unlatch' a title" [27].

1.3. The Library Crowdfunding Model

The scholarly community has, over the years, reiterated the importance of collaboration in the publishing ecosystem [28]. Because collaboration is about joining forces and sharing, the library crowdfunding model encourages libraries to share the cost of publishing peer-reviewed OA monographs to take the financial burden off researchers [29]. For this reason, the model has been considered innovative and possibly sustainable long term [30]. In this model, libraries worldwide join forces to "open" a number of scholarly monographs (or collections of monographs) every year because they have similar beliefs regarding how scholarly content should be published and made available [31]. The money collected from the participating institutions is then distributed to publishers and authors to cover the cost of the BPCs.

The crowdfunding model's advantage is that the funds collected from the participating libraries are used to cover author fees. When enough funds are collected, selected monographs are published OA with various Creative Commons (CC) licenses assigned to them. They then become available to the institutions that fund them, and to any user online [32]. Several OA initiatives rely on institutional crowdfunding to finance the publishing of OA monographs, both frontlist titles (never-before-published titles) and backlist titles (older books that already exist in print and are being permanently "flipped" to OA), including, for example, Reveal Digital, Unglue.it, and Knowledge Unlatched [33].

As explained on its website, Reveal Digital "collaborates with libraries to produce OA primary source collections from under-represented voices." Libraries can participate in three ways: via a one-time contribution model (to support a single collection or project), via multiyear funding to a publishing program, and by contributing content to collections [34]. Unglue.it is rooted in the idea that "the gifts from many readers can free e-books from the DRM [Digital Rights Management] chains that bind them" [35]. Once a funding goal for a title is set, the organization collects pledges from individuals or institutions and distributes the collected funds to the rights holder, who can publish an OA title under a CC license. The model started by primarily focusing on commercial backlist titles, but it eventually included frontlist monographs [36]. Unlike Unglue.it, Knowledge Unlatched (KU)—the focus of this study—aims to provide open access to books from established scholarly publishers by inviting libraries to support large batches of books [33]. The libraries that participate each year form a global consortium that collectively funds the publishing of OA monographs via KU.

Many factors influence an institution's decision to participate in library crowdfunding initiatives, including, for example, it's budget, the local significance of the content published, the author's affiliation with the institution, the reputation of the crowdfunding initiative, etc. Previous research has indicated that the budgets of institutions are a strong reason why libraries participate in crowdfunding initiatives for OA monographs. Surveys have also shown that one of the main reasons that institutions participate is the belief of librarians in the principle of OA. Further, librarians are more likely to support OA content that is closely tied to their institution's research interests [23].

1.4. Knowledge Unlatched Crowdfunding Initiative

The COPIM report identified KU as the initiative that pioneered the library crowdfunding model [27], and OAPEN's most recent list of available models for OA monographs also identifies it as the key example of "institutional crowdfunding" [26]. Since the 2013 pilot, KU has been "unlatching" new collections of OA monographs every year and has facilitated the publishing of about 4000 monographs, with over 670 institutions participating in the crowdfunding to date [37]. It is best known to libraries for KU Select HSS Books, a multidisciplinary collection of peer-reviewed monographs in English that cover various HSS disciplines. The titles to be "unlatched" are selected each year by the KU Selection

Committee, which comprises over 260 librarians worldwide. The committee members participate each year in a democratic voting process online [37].

After the titles have been selected, libraries worldwide are invited to make "pledges", with the option of supporting the complete KU Select package (frontlist and backlist), Frontlist Only package, or individual subject packages within KU Select (e.g., Anthropology, History, Politics, etc.). KU collects funds from libraries and passes them on to publishers, who then use the funds to publish OA monographs and pay the authors. After the pledging cycle closes, the crowdfunding results are assessed, and the "unlatching" process begins. The more funding KU receives from libraries, the more books from the "planned" KU Select HSS Books collection are published OA.

COPIM's 2021 report on collective funding models [38] identified positive and negative sides to KU's annual library crowdfunding campaign. On the positive side, KU was perceived as providing a large amount of varied content from reputable publishers and given credit for involving librarians in the selection process. Librarians decide which titles will receive support from the crowdfunding efforts, and they may join the Selection Committee regardless of whether their institution participates in the crowdfunding. KU was also perceived as providing libraries with clear assurances around digital preservation and long-term access, as well as being transparent about their earnings. In contrast, KU's transition from a nonprofit entity to a for-profit organization has drawn criticism, disappointing some librarians in the process, who fear that KU's for-profit status—and, most recently, Wiley's acquisition of KU, which was announced in December 2021—might "monetize" the OA movement and compromise its core values [38].

1.5. Rankings as Tools for Identifying University Traits

In the past two decades, various ranking sources have sprung up that claim to offer reliable quantifications of the achievements of universities worldwide, including, among many others, the Times Higher Education (THE) World University Rankings and Academic Ranking of World Universities (ARWU). The methodologies of these sources have been criticized over the years, and mainly the focus on research performance and reputation, which may not accurately reflect the attributes that make up a "quality" university [39,40]. Studies have pointed out that universities in English–speaking countries dominate university rankings [40,41]. Likewise, concerns have been raised about the inconsistencies in the methodologies applied to rank universities, leading scholars to question whether the rankings underserve many institutions, and particularly non-Western ones [40]. While such concerns are valid and the shortcomings of university rankings should be emphasized when used as research tools, they may still provide relevant information to researchers, students, and funders, among other stakeholders, when used responsibly [42]. They may offer a useful international comparative perspective and make an institution's strengths visible [42].

University ranking sources have also been studied for their accuracy [43], stability over time [39], and usefulness in improving research [41]. In OA publishing, they have been studied in the context of OA journal publishing and whether the citation impact of the research articles published OA contributes to improving an institution's ranking [44]. There appear to be no studies that use university rankings to profile institutions that support a business model for OA publishing, and particularly the library crowdfunding model for OA monographs—the focus of this paper.

Because the aim of this study was to identify the traits of the universities most or least interested in one specific business model for OA monographs—KU's library crowdfunding model—the following four ranking sources were chosen for this research: Times Higher Education, Academic Rankings of World Universities, Quacquarelli Symonds World University Rankings, and Leiden Ranking.

Times Higher Education (THE) evaluates research-intensive universities across their core missions: teaching, research, knowledge transfer, and international outlook. The performance indicators are divided into five areas: teaching (the learning environment);

research (the volume of research produced, income, and reputation); citations (the universities' role and impact in spreading new knowledge and ideas); international outlook (staff, students, and research); industry income (knowledge transfer) [45].

The Academic Rankings of World Universities (ARWU) uses six indicators to rank universities, including the number of alumni and professors that have won Nobel Prizes and other awards; the number of highly cited researchers (according to Clarivate Analytics); the number of articles published in *Nature* and *Science*; the number of articles indexed in the "Science Citation Index—Expanded and Social Sciences Citation Index;" the per capita performances of universities [46].

The Quacquarelli Symonds (QS) rankings are based on six key indicators: academic reputation based on its academic survey, which collates the opinions of over 130,000 individuals (40%); employer reputation based on the QS Employer Survey (10%); the faculty/student ratio (20%); citations per faculty (20%); the international faculty ratio (5%); the international student ratio (5 percent) [47].

What sets the Leiden Ranking apart from the other three sources is that it does not use any data obtained directly from universities (e.g., via surveys). Leiden's exclusive focus is on the university performance. It is also the only one of the four ranking sources that examines the impact of the institutions' OA research, measuring the volume and proportion of OA publications [48], which makes it an important—perhaps the most important—addition to this study. Although the data collected by Leiden are focused on OA journal publications and not monographs (the exclusive focus of this study), this is a unique indicator that helps to explain an institution's current commitment to publishing its research OA.

1.6. The Study's Aim and Direction

This paper is a reworked, modified, and updated version of a portion of the author's doctoral thesis ("Sustainability of the crowdfunding model for Open Access academic e-books"), defended in 2021 in English at the Department of Information and Communication Sciences, the Faculty of Humanities and Social Sciences, the University of Zagreb. The tables included were updated in the summer of 2022 to reflect the most recent rankings, and the methodology was revised (the Leiden ranking was not part of the original research) to investigate the library crowdfunding model's sustainability from the perspective of institution rankings to determine the possible characteristics that may be assigned to the European institutions that participate or that are likely to participate in one specific library crowdfunding initiative for OA monographs. The study explores the institutions' traits in the context of only this business model, and only one initiative that is based on the model—KU Select HSS Books—for several reasons:

○ KU is the longest-running library crowdfunding campaign, and it is identified as a pioneer of the model [27], with enough history and data to warrant relevant results;

○ Owing to KU's cooperation, obtaining accurate information on the annual participation of libraries in KU's initiatives has not presented challenges;

○ KU Select HSS Books is a multidisciplinary collection of monographs in the HSS fields most relevant to long-form scholarship;

○ KU Select HSS Books is international in its approach, featuring a wide range of international authors and global publishers;

○ The "internationality" and collaborative aspects of KU Select HSS Books significantly support the argument that OA must function as a movement that supports international rather than national interests.

The study relies on the data supplied by KU to assess the characteristics of the institutions that participated in KU's annual crowdfunding and, therefore, financially supported the multidisciplinary collection between 2016 and 2021. While KU runs several library crowdfunding initiatives each year, this study only focuses on KU's "in-house" collections: KU Select HSS Books and KU Focus Collection (a "topical" extension of KU Select HSS Books).

The study seeks to answer how various universities "score" according to the four university ranking sources, and how their rankings relate to their participation in KU's crowdfunding campaign over the past six years (the participation data supplied by KU show which European institutions supported KU's legacy multidisciplinary collection over a six-year period between 2016 and 2021). The ranking data used in the study were collected from the websites of the four sources, focusing on five scores/indicators:

1. World ranking—determines an institution's overall reputation on the world stage;
2. Citation impact—determines an institution's role in spreading new knowledge and ideas, and the global influence of its scholars;
3. Research impact—determines the volume of an institution's published research, as well as the reputation and perceived excellence of the published research;
4. Research published open access—determines the proportion of an institution's research that is published OA;
5. Internationality—determines an institution's commitment to international collaboration and its investment in attracting international students and faculty.

To remain impartial and mindful of the constraints of ranking sources, the study adhered to the following principle highlighted in a 2017 blog post on the Leiden's Center for Science and Technology's website [42] regarding the use of university rankings: "An exclusive focus on the ranks of universities should be avoided; the values of the underlying indicators should be taken into account." For this reason, specific indicators (and subindicators)—qualitative and quantitative—were selected from the four ranking sources to highlight the institutions' strengths and characteristics.

As mentioned, no studies have used university ranking sources to determine the characteristics of the institutions that participate in library crowdfunding initiatives for OA monographs. The goal of the study is to add to the growing body of literature focused on OA monographs, and particularly with an understanding of the acceptance of the library crowdfunding model in its early stages, and the determination of the types of institutions that are likely to support it in the future. Although the study focuses on one type of crowdfunding initiative, it aims to give insight into the kinds of institutions that are likely to support similar initiatives with the aims of interdisciplinarity, internationality, and global collaboration.

As surveys confirm, librarians have recognized the role of their institutions in helping to sustain the publishing of OA monographs [23]. While their motives are likely to draw attention, they are not the focus of this study. The study's goal is to serve as a starting point from which initial conclusions can be drawn about the types of institutions that have shown interest in supporting global library crowdfunding by participating in a widely spread and well-known initiative of this kind.

1.7. Research Questions and Hypotheses

By examining the various rankings of European universities, and by focusing on one library crowdfunding initiative for OA monographs, KU Select HSS Books, the study seeks to answer the following:

1. What types of European universities participate in well-established global library crowdfunding initiatives involving various disciplines, authors, and publishers?
2. How do the institutions that participate in library crowdfunding for OA monographs via one library crowdfunding initiative (KU Select HSS Books) differ from those that do not participate (or that rarely participate) in terms of their world rankings, citation impacts, general research output, open access research output, and internationality?

The hypotheses include the following:

1. Research-intensive universities that rank high in their overall reputations and research output will most likely participate in large multinational library crowdfunding campaigns featuring international scholarly content, authors, and publishers;

2. Institutions that do not currently participate (or that rarely participate) in library crowdfunding on a global level are ranked lower overall, have a smaller research output, do not invest in OA research, and are less focused on building international academic communities.

2. Materials and Methods

This research traces the rankings of 100 European institutions in 26 countries that supported or did not support KU's annual crowdfunding initiative for OA monographs, known as KU Select HS Books, over the course of six years (2016–2021). The sample includes institutions from various countries, which ensures a broad scope, variety, and internationality.

2.1. Institutions' Participation Data and Measurements

Each institution in the study was given a Support Score, which ranged from 0 to 6, to reflect the number of times that the institution participated in KU's initiative, starting in 2016 and ending in 2021 (Tables A1 and A2). (These data were provided by Knowledge Unlatched (KU) and reflect each institution's participation on an annual basis for six consecutive years, including 2016, 2017, 2018, 2019, 2020, and 2021. The listing of institutions that participated in KU's crowdfunding initiatives since the pilot is available on KU's website (https://knowledgeunlatched.org/library-partners/ [37] accessed on 1 September 2022). If an institution received a 0, then it participated zero times; if it received a 1, then it supported KU's collection only once in the past six years, which makes it a "rare supporter". If an institution received a 6, it participated in this crowdfunding initiative every year since 2016, making it a "consistent supporter". It is important to note that this study does not consider the funds that each institution contributed to the initiative. Although pricing is an important factor to consider, the study is not concerned with the amount of funding set aside by each institution, but rather solely with the institution's participation and, consequently, its interest in supporting the publishing of OA monographs via a specific business model.

To establish the necessary averages, the study divided the 100 institutions into 4 distinct groups based on the Support Scores: those that do not support (Score 0); those that support rarely (Scores 1 and 2); those that support often (Scores 3 and 4); those that support the most (i.e., those that participated five or six times in the past six years) (Scores 5 and 6). Those that "support rarely" only supported the initiative once or twice in six years, while those that "support often" supported three or four times. These four groups were further broken down as follows (Table A1):

- Institutions that support the most (Scores 5 and 6): 9 institutions;
- Institutions that often support (Scores 3 and 4): 15 institutions;
- Institutions that rarely support (Scores 1 and 2): 21 institutions;
- Institutions that do not support (Score 0): 55 institutions.

The objective was to collect the participation data for the institutions according to the four distinct groups, profile them by relying on four university ranking sources—chosen for their prominences, diverse methodologies, and transparent rankings—and establish the averages in each ranking category. Table A2 lists the 100 institutions by country.

2.2. University Ranking Data and Measurements

The four ranking sources used in the study differ in how they calculate various indicators and are broken down as follows:

THE indicators for the year 2022:

○ World ranking (combining the scores for teaching (30%); research (30%); citations (30%); international outlook (7.5%); industry income (2.5%));

○ Citation impact (the average number of times a university's published work is cited by scholars in the literature globally);

○   Research impact (including three indicators: a university's reputation for research excellence based on the institutions' responses to the THE's reputation survey (18%), research volume (6%), and research income (6%));

○   International outlook (including three indicators: the proportion of international students (2.5%), proportion of international staff (2.5%), and proportion of a university's total relevant publications that have at least one international coauthor (2.5%)).

ARWU indicators for the year 2022:

○   *World ranking* (combining the scores for award-winning alumni (10%); award-winning staff (10%); highly cited researchers (20%); papers published in *Nature* and *Science* (20%); papers indexed in the Science Citations Index and Social Science Citation Index (20%); per capita academic performance (10%));

○   Highly cited researchers (Clarivate Analytics' list issued in November 2021 was used to calculate the "Hi-Ci" indicator for 2022. Only the primary affiliations of highly cited researchers are considered);

○   Research output (the number of papers indexed in the Science Citation Index—Expanded and Social Science Citation Index).

QS indicators for the year 2023 (which were published in 2022 and are therefore comparable to the THE, ARWU, and Leiden rankings):

○   World ranking (combining the scores for academic reputation (40%); employer reputation (10%); the faculty/student ratio (20%); citations per faculty (20%); the international faculty ratio (10%); the international student ratio (10%));

○   Citations per faculty (the university's research quality is measured with a citation-per-faculty metric, which takes the total number of academic citations in papers produced by an institution in five years);

○   International faculty ratio (the ratio of international faculty staff to overall staff);

The Leiden indicator for the year 2022 (published in 2022):

○   The proportion of articles published open access in all disciplines (in the period 2017–2020).

Although some information provided by the four university ranking sources used in this research—THE (the data were collected from the following website in the summer of 2022 and leading up to 1 September 2022: https://www.timeshighereducation.com/world-university-r%C3%A0nkings/2022); ARWU (the data were collected from the following website in the summer of 2022 and leading up to 1 September 2022: https://www.shanghairanking.com/rankings/arwu/2022); QS (the data were collected from the following website in the summer of 2022 and leading up to 1 September 2022: https://www.topuniversities.com/university-rankings/world-university-rankings/2023); Leiden (the data were collected from the following website in the summer of 2022 and leading up to 1 September 2022: https://www.leidenranking.com/ranking/2022/list)—is similar and may overlap, each of these sources offers insight into at least one unique aspect that the others do not. To provide the most accurate data that help to "profile" institutions, only the institutions ranked by all four ranking sources for at least one of the five chosen indicators were included in the analysis. Below is the breakdown of the examined indicators in the order in which they are presented in the Results section of this paper:

- World ranking—from three sources: THE, ARWU, and QS;
- Citation impact—from three sources: THE (average citations per publication), ARWU (highly cited researchers), and QS (citations per faculty);
- Research impact—from two sources: THE (research volume, reputation, and income) and ARWU (volume of papers indexed);
- Research published open access—from one source: Leiden;
- Internationality—from two sources: THE (international outlook) and QS (international faculty ratio).

3. Results

The tables in this section show how the 100 institutions rank in general in each category (i.e., indicator) and when divided into 4 distinct groups (do not support, support rarely, support often, and support the most). The tables break down the numbers and averages per group for consistency and clarity, always stating the overall average in the first row. In all cases, the groups are presented in the following order: all institutions; institutions that support the most; institutions that support often; institutions that support rarely; institutions that do not support.

3.1. World Ranking

The first factor considered is the world university rankings according to three sources: THE, ARWU, and QS.

The THE world ranking is made up of scores for teaching (30%), research (30%), citations (30%) (as indexed in Scopus), international outlook (7.5%) (including the proportion of international students and staff, as well as the institution's international collaboration), and industry income (2.5%) (how much research income an institution earns from the industry).

Table 1 shows the THE average world rankings for all institutions combined and for each group. The average THE ranking for all 100 institutions is 277.6. When we calculate the average ranking within each group, the institutions that support the most (5 or 6 times out of six) receive the highest ranking (i.e., the lowest number): 148.1. The institutions that did not support KU crowdfunding over the past six years receive the lowest ranking (i.e., the highest number): 370.2. In other words, while the THE ranking of the institutions that participate in some capacity (either the most, often, or rarely) is well above the overall average, the average ranking of the institutions that do not support is significantly below the overall average.

Table 1. THE average rankings of institutions.

Institutions	THE Avg. Ranking
All institutions (100)	277.6
Support the most (9)	148.1
Support often (15)	162.1
Support rarely (21)	173.1
Do not support (55)	370.2

Unlike the THE's world ranking approach, the ARWU places the most emphasis on faculty achievements. Award-winning alumni, award-winning staff, and highly cited researchers make up 50 percent of the overall score. Table 2 shows the ARWU average world rankings for all institutions combined and for each group. The average ARWU ranking for all 100 institutions is 257.5. When we examine the average ranking within each group, we see that the highest ranking (i.e., the lowest number) is given to those institutions that support the most: 188.1. The lowest ranking is given to the institutions that do not support KU's library crowdfunding initiative: 297.9. The average ARWU ranking of the institutions that participate in some capacity (the most, often, or rarely) is above the overall average, while the average ranking of the institutions that do not support is below the overall average.

Table 2. ARWU average rankings of institutions.

Institutions	ARWU Avg. Ranking
All institutions (100)	257.5
Support the most (9)	188.1
Support often (15)	205.0
Support rarely (21)	218.9
Do not support (55)	297.9

Like THE, QS relies on survey data to determine each institution's reputation among academics and employers. Half of its overall ranking score is based on opinions and not on calculable data. The other major indicator for QS is the faculty–student ratio, which points to the quality of the teaching and learning environment and makes up 20% of the overall score (compared with THE, for which the faculty–student ratio makes up only 4.5 percent of the overall score).

Table 3 shows the QS average world rankings for all institutions combined and for each group. The average overall QS ranking for all institutions is 284.2. When we examine the average ranking of each group, we see that this time the highest ranking is not given to those that support the most but to those that rarely support (190.2), while the lowest ranking is given to the institutions that do not support KU crowdfunding (351.1). While all the institutions that participate in some capacity rank higher than the overall average, those that rank the highest this time are those that rarely support.

Table 3. QS average rankings of institutions.

Institutions	QS Avg. Ranking
All institutions (100)	284.2
Support the most (9)	197.0
Support often (15)	222.7
Support rarely (21)	190.2
Do not support (55)	351.1

The more institutions that support the KU crowdfunding initiative for OA monographs, the higher their world rankings. The institutions that support the most (five or six times in six years) on average rank higher than the institutions that support often (three or four times in six years), rarely (one or two times in five years), or never (no participation in six years). However, as the QS ranking shows, the most supportive institutions do not consistently rank higher than those that rarely support (three or four times in six years). What is evident is that the institutions that support in some capacity (always, sometimes, or rarely) rank higher, while the institutions that do not support at all rank the lowest, and lower than the overall average of all institutions.

One way to explain the discrepancy in QS's ranking is by comparing the sources' methodologies. Of the three sources, QS is the most dependent on survey results. Half of its overall score rests on academic and employer reputation survey results. In comparison, the ARWU develops its scores by relying solely on quantifiable data, with great emphasis placed on the academic achievements of faculty. This may also explain why its average score for all institutions is the lowest of the three sources. Lastly, the THE's score combines reputation survey results (33%) and other quantifiable parameters. In other words, the QS scores may deviate from the pattern because the QS methodology relies on survey data, which may not always reflect accurate and reliable answers and opinions, and may even reflect subjective thoughts and unmotivated participants.

What is also noticeable when comparing Tables 1–3 is that the three groups that support in some capacity are much closer in score than the overall average ranking or the

group that does not support. We can also see that the institutions that support rarely have high average world rankings. Consequently, while we may conclude that institutions that support KU crowdfunding are ranked higher than average, it is also accurate to conclude that institutions that support rarely rank high.

3.2. Citation Impact

The citation impact analysis led to similar conclusions to those of the analysis of the world rankings. This category assesses the quality of the faculty, how much their research is shared, and to what extent it influences the academic community. THE examines this influence by capturing the average number of times that a university's published work is cited by scholars globally. Its bibliometric data supplier, Elsevier, examined over 108 million citations to 14.4 million journal articles, article reviews, conference proceedings, books, and book chapters published over five years. The data include 24,600-plus academic journals indexed by Elsevier's Scopus, and all indexed publications between 2015 and 2019 [46].

Table 4 shows that the average number of citations per work for the most supportive group is the highest, 86.2, while it is the lowest for the group that does not support, 68.9. Upon closer examination, it becomes apparent how close the numbers are for the three groups that support crowdfunding in some capacity—the most, often, or rarely—leading to the conclusion that the institutions that support crowdfunding to some degree employ researchers whose works are cited significantly more than those of the institutions that do not participate in this type of library crowdfunding.

Table 4. Average citations per publication: THE.

Institutions	THE Avg. Citations
All institutions (100)	76.0
Support the most (9)	86.2
Support often (15)	83.9
Support rarely (21)	84.7
Do not support (55)	68.9

The ARWU focuses on the number of "highly cited" researchers of an institution selected by Clarivate Analytics, considering only the primary affiliations of highly cited researchers. Table 5 shows that the highest number of highly cited researchers belongs to the institutions that support the most (19.7), while the lowest belongs to the institutions that do not support (12.8). However, the institutions that support rarely do not trail far behind the institutions that support the most, and they score higher (19) than the institutions that support often (17). The only group with a significantly low score in this category is the group of institutions that, to date, have not supported crowdfunding (12.8). The results in Table 5 point to the closeness in the scores between the "Support the most" and "Support rarely" groups compared with the others. Again, the data reveal that the institutions that do not support KU crowdfunding score low in the number of highly cited researchers—lower than the overall average for all institutions. However, the "Support rarely" group ranks higher than the "Support often" group.

Table 5. Average number of highly cited researchers: ARWU.

Institutions	ARWU Avg. Score for Highly Cited Researchers
All institutions (100)	15.3
Support the most (9)	19.7
Support often (15)	17.0
Support rarely (21)	19.0
Do not support (55)	12.8

While the THE citation score captures the average number of times that a university's published work is cited by scholars globally and the ARWU captures the number of highly cited researchers at an institution, the QS measures the research quality by using the citation-per-faculty metric, taking the number of citations in papers produced by a university in a five-year period. Table 6 shows that the highest number of citations per faculty belongs to the "Support often" group (55.2), compared with the "Do not support" group (33.5). Table 6 reveals that the highest scores for the number of citations per faculty belong to the three groups that support in some capacity (the most, often, or rarely). The numbers for these three groups are not far apart, whereas the difference between the institutions that support in some capacity vs. those that do not support at all is notably greater (33.5 vs. 51.2–55.2).

Table 6. Citations per faculty: QS.

Institutions	QS Average Citations per Faculty
All institutions (100)	42.2
Support the most (9)	53.2
Support often (15)	55.2
Support rarely (21)	51.2
Do not support (55)	33.5

Although the way that the citation impact is measured differs from source to source—with THE focusing on the average citations per published work, the ARWU focusing on the number of highly cited researchers, and QS focusing on the number of citations per faculty—all three sources give insight into the influence of the institution's produced research and researchers. It can be concluded that the institutions that support KU crowdfunding the most have the greatest citation impact overall, while the institutions that support often and rarely have above-average citation impacts when compared with the average for all institutions. This again points to the quality of the institutions that have not yet fully embraced the crowdfunding model for OA monographs, but that have not entirely ignored it either.

3.3. Research Impact

The next examined indicator is the research output and reputation, as assessed by two sources: THE and the ARWU. Because the two organizations do not apply the same criteria for evaluating this indicator, their numbers do not reflect the same metric, but when taken together, they give insight into each institution's research output and the perception of that output in the scholarly community.

According to the THE website, the most prominent indicator in this category looks at the university's reputation for research excellence among its peers, based on the responses to the THE's annual Academic Reputation Survey (18%); however, the category also takes into account the number of journal publications indexed by Elsevier's Scopus database per scholar, scaled for institutional size (6%) and research income (6%). Table 7 shows that the average THE research score of all institutions is 40.5. The average research score for the "Support the most" group is 47, while the average score for the "Do not support" group is below the overall average (35.7), which is consistent with previous findings. However, the "Support rarely" group scores higher in this category (48.2) than the "Support the most" group (47). This finding again points to the high performance of the "Support rarely" group—higher than the other two groups that support the KU initiative the most—while the overall average for the "Do not support" group (grey) is lower than the overall average for all institutions.

Table 7. Average scores for research volume, income, and reputation combined: THE.

Institutions	THE Avg. Research
All institutions (100)	40.5
Support the most (9)	47.0
Support often (15)	43.3
Support rarely (21)	48.2
Do not support (55)	35.7

The ARWU examined the number of papers indexed in Clarivate's Science Citation Index—Expanded and Social Science Citation Index in 2021 (identified on the ARWU website as the PUB indicator). As explained, "to distinguish the order of author affiliation, a weight of 100 percent is assigned for corresponding author affiliation, 50 percent for first author affiliation (second author affiliation if the first author affiliation is the same as corresponding author affiliation), 25 percent for the next author affiliation, and 10 percent for other author affiliations" (ARWU Methodology, 2022). Table 8 shows that the average score for all institutions for the PUB indicator (i.e., the number-of-papers index in 2021) is 43, while the average PUB score for the institutions that do not participate in crowdfunding is below average: 42.4. The scores for the "Support rarely" and "Support the most" groups are similar, with a slightly higher score for the "Support rarely" group (44.5 vs. 44.4). What is surprising here is that the score for the "Support often" group is lower than the overall average (42.2 vs. 43), as well as the average of the "Do not support" group (42.2 vs. 42.4). In other words, the institutions that support rarely have the highest number of papers indexed in the two citations indexes, which again points to the productivity of the research of these institutions.

Table 8. ARWU papers indexed in Science Index and Social Science Index.

Institutions	ARWU Avg. PUB Score
All institutions (100)	43.0
Support the most (9)	44.4
Support often (15)	42.2
Support rarely (21)	44.5
Do not support (55)	42.4

In summary, the THE and ARWU indicators that are focused on some aspect of research productivity—including the volume of published articles, volume of indexed articles, income, and reputation—show that the institutions that support KU crowdfunding rarely have the highest performances in terms of their overall research output and impacts, which are lower in the other groups. When we combine this finding with the findings on the citation impact, we conclude that the institutions that do not support KU crowdfunding have the lowest scores in all aspects of the citation and research impact. In contrast, the institutions that support KU in some capacity always perform higher than those that do not in terms of the research volume, research excellence, and research influence. Further, while the institutions that support KU crowdfunding the most may be defined as producing the highest number of the most influential ("highly-cited") researchers, the institutions that rarely support stand out for their research volume and the perception of their research in the scholarly community.

3.4. Research Published Open Access

If we next consider the extent to which these same institutions produce their research OA—including gold, hybrid, green, or bronze OA journals (the focus of the Leiden

ranking)—then we obtain a better understanding of these institutions' current commitment to OA (even if it is not specific to monographs).

Table 9 shows the percentages of OA journal publications at the 100 institutions, with the overall average for all institutions at 71%. This means that, on average, 71% of the publications that these institutions produce are published OA. For the institutions that support the most, the number goes up to 80.6%, while for those that do not support, the number goes down to 65.5%. What the data in Table 9 reveal is that the institutions that support the most and the institutions that support often publish over 80 percent of their research (in journal format) OA, which is significantly higher than the OA output of the institutions that do not support it.

Table 9. Proportion of OA publications: Leiden.

Institutions	P.P. (OA)
All institutions (100)	71.0%
Support the most (9)	80.6%
Support often (15)	80.4%
Support rarely (21)	74.5%
Do not support (55)	65.5%

When we combine the results of the Leiden ranking for OA with the results focused on the research impact, we conclude that those institutions that do not support KU crowd-funding rank below the overall average regarding their overall research output and the percentage of their research published OA. In addition, the more institutions that participate in crowdfunding initiatives for OA monographs, the greater the percentages of their current research that is published OA in journals.

3.5. Internationality

The last indicator considered is not related to an institution's rank or reputation, or to the volume or quality of its research output. Instead, it looks at the "internationality" of each institution (i.e., the extent to which each institution is international in its structure and approach to building diverse cultures and encouraging cross-cultural and cross-university collaborations—one of the pillars of the OA movement).

The THE's international outlook score (which makes up 7.5 percent of the overall ranking score given to an institution) comprises three indicators: the proportion of international students, proportion of international staff, and international collaboration. The third indicator is especially relevant because it shows how much an institution collaborates with other institutions and promotes various "collaborative" endeavors. Here, THE calculates the proportion of a university's publications with at least one international coauthor in five years [46]. The three indicators give each institution a score that determines its "international outlook".

Table 10 breaks down the scores per group, showing that the institutions that do not support KU crowdfunding scored 63, which is below the average for all institutions (72.2). The institutions that support the most and those that support rarely received the same score: 81.8, while the institutions that support often (but not the most) received the highest score: 86.8.

QS's international faculty ratio indicator compares the international staff ratio to the university's overall staff. The term "international" is determined by the faculty's citizenship, and in the case of dual citizenship, the deciding criteria are the citizenship obtained through birth or the first passport in possession. This QS indicator may complement the THE's international outlook indicator, which seeks to determine an institution's internationality by looking at the proportion of international students and staff and the "international collaboration". The QS's international faculty ratio zooms in only on the staff that have contributed to academic research or teaching for at least three months.

Table 10. International outlook scores: THE.

Institutions	THE Average International Outlook
All institutions (100)	72.2
Support the most (9)	81.8
Support often (15)	86.8
Support rarely (21)	81.8
Do not support (55)	63.0

Table 11 shows the international faculty ratio results, which are similar to the "international outlook" results. The lowest international faculty ratio is assigned to the "Do not support" group (40.3%), while the highest ratio is assigned to the "Support often" group (88.9%).

Table 11. International faculty ratios: QS.

Institutions	Average International Faculty Ratio
All institutions (100)	59.3
Support the most (9)	80.8
Support often (15)	88.9
Support rarely (21)	78.6
Do not support (55)	40.3

The analysis of the universities' "internationality" shows that the most "international" of all the universities are not those that support the most but those that support often, and that although the institutions that support rarely show strong rankings in the categories of research output in general, they appear not to be as "international" as the universities that support KU crowdfunding often or the most. However, they are still well above the overall average even in this category, and they are above the average when compared with the "Do not support" group.

4. Discussion

By taking a closer look at several university ranking indicators given by four ranking sources—THE, ARWU, QS, and Leiden—we arrive at some observations about the types of institutions that are presently most inclined to support KU's crowdfunding initiative for OA monographs (and thus possibly also the library crowdfunding approach in general). We also arrive at some observations about the types of institutions that are currently not considered to be drivers of the success of KU's library crowdfunding model because they do not support it consistently, but do support it, which may result in stronger future support by these institutions and thus contribute to the sustainability of the model. Lastly, the results reveal the institutions that have remained uninterested in this type of "international" library crowdfunding.

4.1. Key Findings

The following are the study's key findings:

○ Institutions that financially support OA monographs through crowdfunding in some capacity (rarely, often, and the most vs. never) have higher overall world rankings than those that do not financially support. On average, the institutions that support crowdfunding the most (i.e., that supported five or six times in the past six years) obtained the highest overall world ranking scores compared with other institutions;

○ Institutions that have not yet participated in KU's crowdfunding initiative consistently rank below average in every category examined. This is particularly noticeable when comparing their "internationality" to that of all the other institutions;

○ Institutions that support crowdfunding in some capacity (rarely, often, and the most) perform higher than those that do not in terms of research volume, research excellence, and citation influence;

○ Institutions that support crowdfunding the most may be defined as producing the highest number of the most influential researchers;

○ Institutions that support KU crowdfunding the most also have the greatest citation impacts, while those that support often and rarely have above-average citation impacts;

○ Institutions that support rarely rank higher than institutions that support often in terms of their citation impacts;

○ Institutions that support rarely stand out for their research output and the overall perception of their research in the scholarly community, pointing to their overall research productivity;

○ Institutions that do not support KU crowdfunding rank below the overall average when it comes to their overall research output, as well the percentage of their research published OA, while the institutions that support crowdfunding the most produce the most significant (i.e., the highest) proportion of research OA;

○ The most "international" of all institutions are not those that support the most but those that support often.

In conclusion, the institutions that are most likely to participate in library crowdfunding initiatives for OA monographs, such as the KU initiative, may be defined as highly ranked overall and productive of relevant and timely research in quantity, quantity, relevance, and timeliness. They already publish a significant proportion of their research OA in article/journal format, and they are highly invested in advancing HSS scholarship and maintaining their international outlooks.

The study confirmed the hypothesis that research-intensive universities that are ranked highly tend to be the most inclined to support library crowdfunding initiatives. The study also revealed that institutions that support rarely rank significantly higher than those that do not support at all in terms of their research output and influence. Therefore, the most significant revelation of this study is the high academic standing of the institutions that support rarely (i.e., the institutions that have participated in KU crowdfunding but have not shown consistent commitment). These institutions may not be as "international" in their outlooks, and they may not be as productive on the OA front as the institutions that support the most, but they stand out for high-quality and significant research output, and particularly when compared with the institutions that do not participate. This finding points to the fact that the library crowdfunding business model for OA monographs is still maturing, as many high-quality institutions worldwide have not embraced it; at best, they have tested it or encountered it in some small capacity over the past few years.

Further, given their high academic standing, the "Support rarely" institutions may be recognized as the type of institutions that could significantly contribute to the sustainability of the KU crowdfunding initiative (and thus library crowdfunding for OA monographs), as their future participation is needed to ensure the success of international library crowdfunding campaigns, such as KU. It may be argued that if these institutions participate in more significant numbers, then the KU crowdfunding model for OA monographs would become more widely accepted and eventually lead to more acceptance among the institutions that have not yet embraced it. It may also be argued that the consistent participation of some institutions may have led to the proliferation of other types of library crowdfunding campaigns for monographs in recent years—not only by KU, but also by various other entities and publishers that now embrace the model (which are not covered in this study).

4.2. Limitations of the Study

This study was guided by the assumption that the institutions ranked by all four sources put significant effort into cooperating with the ranking organizations to ensure that they appear on various ranking lists. However, this does not imply that the institutions not

ranked by all four sources (and not included in this study) do not warrant a closer analysis in the context of their support for OA monographs and the library crowdfunding model (through KU or otherwise). There could be a number of reasons why an institution is not ranked by one or all four sources. While the study aimed to diversify the sample as much as possible by including institutions that represent a wide range of European countries, only those institutions ranked by all four sources were examined. Likewise, other ranking sources that were not the subject of this study may yield different results.

The study relied on the data of only one of several existing library crowdfunding initiatives for OA monographs. It is possible and to be expected that the institutions that do not participate in KU's crowdfunding initiative support other initiatives (including but not limited to crowdfunding) for OA monographs, and that the results of this study may not give a complete picture of their interest in library crowdfunding for OA monographs. Given the controversies surrounding KU's for-profit status, it is also to be expected that some institutions may not support KU not because they do not support the idea of library crowdfunding through such a scheme, but because they do not wish to support the model if it is not rooted in nonprofit ideals.

The study also did not focus on the amount of funding set aside by each institution for each year that it participated; the focus was on the institutions' ongoing commitment or lack thereof. Library budgets are important in determining an institution's ability to support OA publishing. If library budgets are limited, and particularly those that may be allocated for OA monographs—as often may be the case in the institutions that receive lower ranking scores overall—then they may be the main reason that these institutions do not participate.

Further, the study attempted to build on the existing knowledge of the possible indicators that point to the sustainability of an innovative business model for OA monographs. While the findings indicate some of the characteristics of the institutions most likely to participate in global library crowdfunding, they do not give a complete picture of the likelihood of support in the future. They may, however, point us in the right direction.

Lastly, the study did not consider the academic focuses of the 100 institutions. While many institutions comprise a range of HSS and STEM programs, some may have a more pronounced focus on and investment in STEM programs. Given that monographs are closely tied to HSS disciplines, they are expected to be most supported by the institutions that are most heavily invested in the humanities and social sciences.

4.3. Recommendations for Further Research

When examining the traits of institutions that embrace the global library crowdfunding model for OA monographs, future studies should go beyond KU Select HSS Books (one of several initiatives run by KU), consider other library crowdfunding initiatives by other players in the OA ecosystem, and compare their findings with this study. Future studies should also evaluate additional indicators that help identify the institutions most likely to keep the library crowdfunding model going, including, for example, their OA budgets. It is common knowledge that libraries worldwide are under pressure to keep investing in OA on tight budgets. The questions that naturally arise from this study are whether the institutions that support the most also have the most OA monographs, and what proportions of their budgets are allocated to OA monographs, compared with those that support rarely or never.

Future studies should also examine the relationship between the ranking of an institution in a specific academic discipline (rather than the overall ranking) and its participation in crowdfunding initiatives that aim to publish OA monographs in that same discipline to determine the likelihood of an institution supporting the fields most relevant for its community.

A future survey with librarians that explores the reasons for their institutions' support of OA monographs—and particularly on the types of crowdfunding models supported and their positions on supporting for-profit vs. nonprofit initiatives—would complement this study. Lastly, an assessment of the kinds of library crowdfunding initiatives currently

supported by the institutions would provide further insight into the current prevalence of the KU Select HSS Books initiative compared with other library crowdfunding initiatives for OA monographs.

Appendix A

Table A1. Breakdown of institutions and their scores.

	Support Score
Total number of institutions examined in the study: 100	–
Institutions that supported the most in the given period (2016–2021): 9/100	Score 5 (supported 5 times in 6 years) Score 6 (supported 6 times in 6 years)
Institutions that supported often in the given period (2016–2021): 15/100	Score 3 (supported 3 times in 6 years) Score 4 (supported 4 times in 6 years)
Institutions that rarely supported in the given period (2016–2021): 21/100	Score 1 (supported 1 time in 6 years) Score 2 (supported 2 times in 6 years)
Institutions that did not support in the given period (2016–2021): 55/100	Score 0 (did not support)

Table A2. Institutions included in the study, by country, including Support Score for each.

Country	Institution	Support Score [†]
Austria	Johannes Kepler University Linz	0
Austria	University of Innsbruck	0
Austria	University of Vienna	0
Belgium	Ghent University	0
Belgium	Hasselt University	0
Belgium	University of Antwerp	0
Belgium	Vrije Universiteit Brussel	0
Croatia	University of Zagreb	0
Czechia	Charles University, Prague	0
Denmark	Aarhus University	0
Denmark	University of Copenhagen	0
Denmark	University of Southern Denmark	0
Denmark	Aalborg University	1

Table A2. *Cont.*

Country	Institution	Support Score [†]
Estonia	University of Tartu	1
Finland	University of Eastern Finland	0
Finland	University of Helsinki	0
Finland	University of Turku	1
Finland	Tampere University	2
France	Grenoble Alps University	0
France	Paris-Saclay University	0
France	Sorbonne University	0
France	University of Bordeaux	0
France	University of Lorraine	0
France	University of Strasbourg	0
Germany	Goethe University Frankfurt	0
Germany	University of Munich	0
Germany	University of Freiburg	1
Germany	University of Hamburg	1
Germany	University of Bonn	2
Germany	Bielefeld University	4
Germany	University of Cologne	6
Greece	National and Kapodistrian Univ. of Athens	0
Greece	University of Crete	0
Hungary	Eötvös Loránd University	0
Ireland	University College Dublin	0
Ireland	Trinity College Dublin	1
Ireland	University College Cork	2
Ireland	University of Limerick	2
Italy	Sapienza University of Rome	0
Italy	University of Bologna	0
Italy	University of Genoa	0
Italy	University of Milan	0
Italy	University of Padua	0
Italy	University of Siena	0
Lithuania	Vilnius University	0
Netherlands	Maastricht University	0
Netherlands	Utrecht University	0
Netherlands	Erasmus University Rotterdam	2
Netherlands	University of Amsterdam	2
Netherlands	University of Groningen	3
Netherlands	Vrije Universiteit Amsterdam	5
Netherlands	Leiden University	6
Norway	University of Bergen	0
Norway	Norwegian Univ. of Science and Technology	3
Norway	University of Oslo	4
Norway	UiT The Arctic University of Norway, Tromsø	6
Poland	Adam Mickiewicz University, Poznań	0
Poland	Jagiellonian University	0
Poland	University of Warsaw	0

Table A2. *Cont.*

Country	Institution	Support Score [†]
Portugal	NOVA University of Lisbon	0
Portugal	University of Lisbon	0
Portugal	University of Porto	0
Romania	Babeș-Bolyai University	0
Serbia	University of Belgrade	0
Slovakia	Comenius University, Bratislava	0
Slovenia	University of Ljubljana	0
Spain	Autonomous University of Barcelona	0
Spain	Autonomous University of Madrid	0
Spain	Complutense University of Madrid	0
Spain	University of Barcelona	0
Spain	University of Navarra	0
Sweden	Lund University	0
Sweden	University of Gothenburg	0
Sweden	Uppsala University	0
Sweden	Stockholm University	5
Switzerland	University of Geneva	2
Switzerland	University of Basel	4
Switzerland	University of Bern	4
Switzerland	University of Zurich	6
United Kingdom	University of Oxford	0
United Kingdom	University of Birmingham	1
United Kingdom	University of Cambridge	1
United Kingdom	University of Glasgow	1
United Kingdom	Loughborough University	2
United Kingdom	UCL, University of London	2
United Kingdom	University of Edinburgh	2
United Kingdom	University of Liverpool	2
United Kingdom	University of Southampton	2
United Kingdom	Durham University	3
United Kingdom	Queen Mary University of London	3
United Kingdom	University of Bath	3
United Kingdom	University of Exeter	3
United Kingdom	University of Kent	3
United Kingdom	University of Sheffield	3
United Kingdom	King's College London	4
United Kingdom	University of Nottingham	4
United Kingdom	University of Surrey	4
United Kingdom	University of Reading	5
United Kingdom	Lancaster University	6
United Kingdom	University of Manchester	6

[†] Notes: For each year that an institution participated, the institution received 1 point, totaling 6 points for the institutions that participated every year between 2016 and 2021. If an institution received a 0, then it did not participate in KU crowdfunding in the period between 2016 and 2021.

References

1. Bailey, C.W., Jr. Open Access and Libraries. *Collect. Dev.* **2007**, *32*, 351–383. [CrossRef]
2. Pinfield, S. Making Open Access work. *Online Inf. Rev.* **2015**, *35*, 604–636. [CrossRef]

3.   Collins, E.; Milloy, C.; Stone, G. Guide to Open Access Monograph Publishing for Arts, Humanities and Social Science Researchers. Research Report. AHRC/JISC Collections. 2015. Available online: https://issuu.com/carenmilloy/docs/guide_to_open_access_monograph_publ (accessed on 1 September 2022).

4.   Collins, E.; Milloy, C. OAPEN-UK Final Report: A Five-Year Study into Open Access Monograph Publishing in the Humanities and Social Sciences. OAPEN-UK. JISC. 2016. Available online: https://issuu.com/carenmilloy/docs/oapen-uk_final_report_single_page_v (accessed on 1 September 2022).

5.   Altman, M. Exploring the Public Evidence on Open Access Monographs. MIT Libraries, DSpace@MIT. 2021. Available online: https://hdl.handle.net/1721.1/129690 (accessed on 15 August 2022).

6.   Suber, P. *Open Access*; "Essential Knowledge" Series; MIT Press: Cambridge, MA, USA, 2012.

7.   Hagerlid, J. Open Access to Monographs in the Humanities and Social Science. Available online: https://journals.lub.lu.se/sciecominfo/article/download/4913/4404/ (accessed on 1 September 2022).

8.   Mongeau, P. 'The Future is Open?': An Overview of Open Access Monograph Publishing. *Ijournal* **2018**, *3*, 1–10.

9.   Adema, J. Overview of Open Access Models for eBooks in the Humanities and Social Sciences. OAPEN Project Report. 2010. Available online: https://www.academia.edu/273622/Overview_of_Open_Access_Models_for_E_Books_In_the_Humanities_and_Social_Sciences (accessed on 15 September 2022).

10.   Ferwerda, E.; Pinter, F.; Stern, N. A landscape study on open access monographs: Policies, funding and publishing in eight European countries. *Knowledge Exchange* **2017**. [CrossRef]

11.   Elliott, M. The Future of the Monograph in the Digital Era: A Report to the Andrew W. Mellon Foundation. *J. Electron. Publ.* **2015**, *18*. [CrossRef]

12.   Crossick, G. Monographs and open access. *Insights* **2016**, *29*, 14–19. [CrossRef]

13.   Cheshire, J. Open access monographs: A humanities research perspective. *Insights* **2014**, *27*, 17–20. [CrossRef]

14.   Bargheer, M.; Dogan, Z.M.; Horstmann, W.; Mertens, M.; Rapp, A. Unlocking the Digital Potential of Scholarly Monographs in 21st Century Research. *LIBER Q.* **2017**, *27*, 194–211. [CrossRef]

15.   Adema, J. Towards a Roadmap for Open Access Monographs: A Knowledge Exchange Report, 2017. Knowledge Exchange Stakeholder Workshop. Available online: https://www.ouvrirlascience.fr/wp-content/uploads/2019/06/KE_Towards-a-Roadmap-for-Open-Access-Monographs-May-2019_2.pdf (accessed on 15 August 2022).

16.   Fathallah, J. Open Access Monographs: Myths, Truths and Implications in the Wake of UKRI Open Access Policy. *LIBER Q. J. Assoc. Eur. Res. Libr.* **2022**, *32*. [CrossRef]

17.   European Commission, Directorate-General for Research and Innovation. *Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission*; Publications Office: Luxembourg, 2019. Available online: https://data.europa.eu/doi/10.2777/836532 (accessed on 14 September 2022).

18.   Adema, J.; Stone, G. Changing Publishing Ecologies: A Landscape Study of New University Presses and Academic-Led Publishing. Jisc. 2017. Available online: https://repository.jisc.ac.uk/6666/1/Changing-publishing-ecologies-report.pdf (accessed on 15 August 2022).

19.   Blankstein, M. Ithaka S+R US Faculty Survey 2021. University of Nebraska—Lincoln. 2022. Available online: https://sr.ithaka.org/publications/ithaka-sr-us-faculty-survey-2021/ (accessed on 15 December 2022).

20.   Turgut, Y.E.; Aslan, A.; Denizalp, N.V. Academicians' awareness, attitude, and use of open access during the COVID-19 pandemic. *J. Librariansh. Inf. Sci.* **2021**, *54*, 350–362. [CrossRef]

21.   Grimme, S.; Holland, C.; Potter, P.; Taylor, M.; Watkinson, C. The State of Open Monographs: An Analysis of the Open Access Monograph Landscape and Its Integration into the Digital Scholarly Network. Digital Science, Dimensions & Altmetric Report. 2019. Available online: https://openresearch.community/documents/59412-the_state_of_open_monographs_report-2 (accessed on 15 December 2022).

22.   European Science Foundation Plan S: Making Full and Immediate Open Access a Reality: What Is cOAlition S? Available online: https://www.coalition-s.org/coalition-s-statement-on-open-access-for-academic-books/ (accessed on 1 September 2022).

23.   Roncevic, M.; Spiranec, S. The sustainability of the crowdfunding model for scholarly open access books from the perspective of librarians. *Vjesn. Bibl. Hrvat.* **2022**, *65*, 199–224. [CrossRef]

24.   Gatti, R.; Mierowsky, M. Funding open access monographs: A coalition of libraries and publishers. *Coll. Res. Libr. News* **2016**, *77*, 456–459. [CrossRef]

25.   Speicher, L.; Armando, L.; Bargheeer, M.; Eve, M.P.; Fund, S.; Leao, D.; Mosterd, M.; Pinter, F.; Souyioultzoglou, I. OPERAS Open Access Business Models White Paper. *Zenodo* **2018**. [CrossRef]

26.   OAPEN. Business Models for Open Access Book Publishing. OAPEN. 2022. Available online: https://oabooks-toolkit.org/lifecycle/10944589-planning-funding/article/10432084-business-models-for-open-access-book-publishing (accessed on 15 December 2022).

27.   Penier, I.; Eve, M.P.; Grady, T. COPIM—Revenue Models for Open Access Monographs 2020. *Zenodo* **2020**. [CrossRef]

28.   Deegan, M. The Academic Book and the Future Project Report: A Report to the AHRC and the British Library. 2017. Available online: https://bl.iro.bl.uk/concern/reports/97e5819f-5def-4b69-b31a-6d4f3fe5fb91?locale=en (accessed on 1 September 2022).

29.   Reinsfelder, T.; Pike, C.A. Using Library Funds to Support Open Access Publishing through Crowdfunding: Going Beyond Article Processing Charges. *Collect. Manag.* **2018**, *43*, 138–149. [CrossRef]

30. Leach-Murray, S. Knowledge Unlatched—Free access to scholarly content for every reader across the world. *Tech. Serv. Q.* **2017**, *34*, 219–221. [CrossRef]
31. Bushong, S.; Cleveland, S.; Cox, C. Crowdfunding for Academic Libraries: Indiana Jones Meets Polka. *J. Acad. Librariansh.* **2018**, *44*, 313–318. [CrossRef]
32. Pinter, F. Open Access for Scholarly Books? *Pub. Res. Q.* **2012**, *28*, 183–191. [CrossRef]
33. Bulock, C. Crowdfunding for Open Access. *Ser. Rev.* **2018**, *44*, 138–141. [CrossRef]
34. Reveal Digital—Jstor Website. Available online: https://about.jstor.org/revealdigital/get-involved/ (accessed on 1 September 2022).
35. Sigal, A. Unglue.it, an e-Books Site That Functions Like a True Participatory Democracy. No Shelf Required. 2017. Available online: http://www.noshelfrequired.com/unglue-it-an-ebooks-site-that-functions-as-a-true-participatory-democracy (accessed on 1 September 2022).
36. Ferwerda, E. Open access monograph business models. *Insights* **2014**, *27*, 35–38. [CrossRef]
37. Knowledge Unlatched Website. Available online: www.knowledeunlatched.org (accessed on 1 September 2022).
38. Gerakopoulou, E.I.; Penier, J.D. The promise of collaboration: Collective funding models and the integration of Open Access books into libraries. *Zenodo* **2021**. [CrossRef]
39. Selten, F.; Neylon, C.; Huang, C.; Groth, P. A longitudinal analysis of university rankings. *Quant. Sci. Stud.* **2020**, *1*, 1109–1135. [CrossRef]
40. Fauzi, M.A.; Tan, C.N.-L.; Daud, M.; Awalludin, M.M.N. University Rankings: A Review of Methodological Flaws. *Issues Educ. Res.* **2020**, *30*, 79–96. Available online: https://search.informit.org/doi/10.3316/informit.085786218299205 (accessed on 1 September 2022).
41. Vernon, M.M.; Balas, E.A.; Momani, S. Are university rankings useful to improve research? A systematic overview. *PLoS ONE* **2018**, *13*, e0193762. [CrossRef]
42. Waltman, L.; Wouters, P.; van Eck, N.J. Ten Principles of the Responsible Use of University Rankings. Center for Science and Technology Studies (CWTS), Leiden University. 2017. Available online: https://www.cwts.nl/blog?article=n-r2q274 (accessed on 16 December 2022).
43. Vidal, J.; Ferreira, C. Universities Under Pressure: The Impact of International University Rankings. *J. New Approaches Educ. Res.* **2020**, *9*, 181–193. [CrossRef]
44. Baldock, C. Citations, Open Access and University Rankings. In *World University Rankings and the Future of Higher Education*; Downing, K., Ganotice, F.A., Jr., Eds.; IGI Global: New York, NY, USA, 2016.
45. THE World University Rankings Methodology Page. Available online: https://www.timeshighereducation.com/world-university-rankings/world-university-rankings-2022-methodology (accessed on 1 September 2022).
46. Academic World University Rankings Methodology Page. Available online: https://www.shanghairanking.com/methodology/arwu/2022 (accessed on 1 September 2022).
47. QS World University Rankings Methodology Page. Available online: https://www.topuniversities.com/qs-world-university-rankings/methodology (accessed on 1 September 2022).
48. Leiden Ranking Indicators. Available online: https://www.leidenranking.com/information/indicators (accessed on 1 September 2022).

Citizen Science in Europe—Challenges in Conducting Citizen Science Activities in Cooperation of University and Public Libraries

Alisa Martek, Dorja Mučnjak and Dolores Mumelaš *

National and University Library in Zagreb, 10000 Zagreb, Croatia
* Correspondence: dmumelas@nsk.hr

Abstract: Citizen science has many definitions but it is commonly known as collaboration between professional scientists and the rest of society. Although there have been cases of its implementation in the past, the term became globally known in 2012. Citizen science activities cover a wide range of academic disciplines and vary widely in what is required of the activity participants in terms of knowledge, time commitment, travel, and the use of technology). For the past ten years, libraries have often introduced citizen science in order to encourage greater interaction between science and society as a form of their services or specially organized activities. The types of libraries that often conduct citizen science are academic, public, and research libraries. Each of these library types has a specific user population; academic libraries have students and scientific and teaching staff; public libraries have the local community; and research libraries have researchers. However, libraries usually carry out CS activities separately, and very rarely in cooperation with other types of libraries. Some collaboration challenges are related to its complexity, the uncertainty regarding research cocreation, and participant retention strategies. Such cooperation is one of the aspects explored by the LIBER project CeOS_SE Project—Citizen-Enhanced Open Science in Southeastern Europe Higher Education Knowledge Hubs. The main goal of the project is to raise awareness of mainstream Open Science and CS practices in Southeastern (SE) Europe. As a project partner, the National and University Library in Zagreb, in cooperation with the University Library of Southern Denmark, conducted a survey that included other European countries in addition to SE Europe to examine and collect good practices of civil engagement in university libraries.

Keywords: CeOS_SE project; citizen science; libraries cooperation; National and University Library in Zagreb; organizational challenges

1. Introduction

As a term, citizen science (hereinafter CS) has many definitions that depend on various factors. Mordechai et. al. stated that CS definitions depend on diverse perceptions, cultural differences, and varied contexts in every world country [1]. In the case of Europe, we can extract the definition from the *Ten Principles of Citizen Science* developed by the European Citizen Science Association, led by the Natural History Museum, London. From these principles, it can be assumed that CS projects include citizens in scientific research who produce new knowledge and have an original scientific outcome where both the scientists and the citizens can benefit [2]. CS has numerous goals such as research, education, and action, imbued with a variety of different individuals, and as such, it has proven to be an excellent opportunity to develop and implement inclusive practices [3].

There are a few CS platforms on the web, but the world's largest platform is SciStarter (scistarter.org, accessed on 14 September 2022). SciStarter was founded in 2011 by Darlene Cavalier, Professor of Practice and Senior Global Futures Scientist at Arizona State University, and it represents an online hub developed to expedite, explore, and popularize CS projects from all around the world. It offers an overview of more than 3000 CS projects,

and brings together people who are interested in science, and offers them various tools and ideas to successfully participate in scientific activities [4]. The platform is significant in education and socialization, but it also encourages the cooperation of various institutions such as libraries, schools, universities, and museums, among others. It has also developed some CS events such as *Citizen Science Month*, which takes place in April [5]. SciStarter has also issued a guide *The Library and Community Guide to Citizen Science for Libraries* so that they can learn more about CS and the integration of such activities into their business, listen to the interests of the local community, connect existing community programs with SciStarter projects, access information sources for conducting CS, etc. [6].

In the USA, there are many examples of CS conducted with the help of all kinds of libraries. In 2015, the Library of Congress launched a crowdsourced transcription program named *By the people* in which all interested individuals from all around the world could participate by making transcripts of historical document (e.g., Wm. Oland Bourne Papers, Branch Rickey Papers, Elizabeth Cady Stanton Papers, Rosa Parks Papers.) [7]. By 1 November 2022, more than 580,000 pages of completed transcriptions have been collected in seven cataloged datasets and more than 32,000 participants have registered accounts [8]. The National Library of Medicine has been conducting a national CS activity since 2018 called the *#CiteNLM Wikipedia Edit-a-thon*. The main goal of this activity is to increase the credibility of Wikipedia content related to medicine and health by adding citations from the National Library of Medicine sources. By spring 2020, more than 300 participants edited more than 500 articles and added 700 references [9]. In 2020, several institutions collaborated in North Carolina's Candid Critters (NCCC) Project including 63 public libraries that helped with distributing camera traps to CS participants in 100 counties in North Carolina, USA. The main aim of this CS activity was to analyze whether large-scale CS camera trapping surveys were useful for gathering wildlife records. The project had a great response, 580 volunteers joined and collected more than 120,000 wildlife records and more than two million photos [10]. A similar activity, but with an emphasis on biodiversity and collective observations of animals and plants, was conducted in 2017 by the San Francisco Public Library. In this way, the aforementioned library supported the San Francisco Biodiversity Policy resolution and gathered citizens to observe and collect data about San Francisco's biodiversity and changes in the environment through the iNaturalist application by the crowdsourcing method [11].

Vohland et al. divided CS activities in Europe into several geographical regions. They established that in Western and Northern Europe, CS relies on learned associations and is very well-developed. In Central and Eastern Europe, there are not as many CS activities, mostly because of the uneven knowledge management, but a great involvement of volunteers was observed. For Southern Europe and the Balkans, there are even fewer CS activities, mostly because of the economic development of countries, but the activities that are carried out are mostly oriented on sensing and monitoring projects [12]. It should be emphasized that most European CS projects relate to natural and technical sciences, while activities related to social sciences and humanities are rarer [13]. In addition, European policymakers still hesitate to utilize data obtained in CS projects with the goal of decision making. What would improve the credibility of CS data is defining its minimum standards [14].

The fundamental concepts we associate with CS are scientific standards, collaboration, open science, communication, and data management. All of these terms can be related to librarianship, so it is not surprising that libraries are stakeholders in the CS implementation, serving not only as a bridge that connects citizens and science, but also as the main bearers of ideas and initiators of activities in the community in which they operate. Ignat et al. suggested a few roles that libraries could play in conducting CS activities such as establishing skills to employ in CS projects, fabricating a toolkit to design CS library projects, forming collections of data forms and educational resources, bid infrastructure, and advertising a positive mindset toward CS and much more [15].

CS activities in European libraries do exist, but unfortunately there have been few case studies in the literature written and available in the English language. More impor-

tantly, some libraries that have implemented such projects are not familiar with the term "citizen science". For example, University College London and their library developed a participatory project named Transcribe Bentham, with the main goal of engaging the citizens and the wider public in the online transcription of Jeremy Bentham's unstudied manuscripts [16]. They did not call it citizen science, but Ignat et al. recognized it as such because the traits of that project are similar to CS activities. The University of Southern Denmark and their library are actively involved in creating CS activities, and they have projects such as A Healthier Funen, Active Living Area, community-driven journalism, and narrative medicine [16]. The ETH Library from Zurich has a successful CS project for their image archive where citizens had to locate areas, date photographs. and identify people and objects, and in this way, they improved the image archive metadata [17]. The Barcelona Network of Public Libraries organized a CS activity named Science and Citizen Action, which was also characterized as a behavioral experiment where data about the housing market simulation were collected [18]. These are just a few examples of the CS practices in European libraries, and considering the names of the activities and their topics, it is possible to conclude that libraries can organize CS activities related to many types of scientific disciplines.

To ensure networking in the implementation of CS, library organizations have begun to implement projects related to CS at the international level. One of them is the Erasmus+ Program under Cooperation Partnerships in Higher Education project *Citizen-Enhanced Open Science in Southeastern Europe Higher Education Knowledge Hubs* led by LIBER (Ligue des Bibliothèques Européennes de Recherche—Association of European Research Libraries). One of the project's goals is to train and upskill the librarians in Southeastern European countries in the field of OS and CS because such activities are lacking in that part of Europe. The project started on 1 January 2022 with the end date of 1 January 2025 (more information on https://ceosse-project.eu/, accessed on 1 December 2022). The project partners are LIBER (Netherlands), University Library of Southern Denmark (Denmark), University of Torino (Italy), University of Patras (Greece), University of Cyprus (Cyprus), University Library "Svetozar Marković" (Serbia), National and University library in Zagreb (Croatia), and the University of Library Studies and Information Technologies (Bulgaria) [19]. Each project partner is in charge of carrying out a part of the project.

The National and University Library in Zagreb's role is PR2 (project results from project package number two) *Report on implementation of citizen-enhanced open science in various open knowledge hubs in SE Europe*. PR2 facilitates transfer and participation in SE Europe between university libraries (project partners) and public libraries (associated partners) for a social purpose by means of citizen engagement in OS. Therefore, PR2 seeks to trigger and build up the dialog and partnership between the mentioned types of libraries as knowledge and innovation hubs in SE Europe that will be carried out together in CS activities [20]. To create the report, the National and University Library in Zagreb conducted several surveys and a guided interview, and it turned out that there have been few examples of the implementation of CS activities organized through the cooperation of universities and public libraries, which requires the development of a deeper study.

It turns out that there are some barriers to conducting library activities in cooperation between different library types. Although cooperation between several library types has a positive effect on their management, libraries can face various barriers in the joint implementation of services, programs, and activities. Regardless of the fact that it may be a matter of collaboration between different types of libraries, certain barriers are more common than others.

Looking at studies regarding the cooperation of American schools and public libraries, Fitzgibbons claimed that there were three main problems in creating joint activities: a lack of staff, insufficient resources, and a lack of unique goals. The author offered some collaboration recommendations such as shared goals, the development of a managing process and its evaluation, strong commitment, good communication channels, decent funding, and adequate staffing [21]. Moreover, LaMaster explained that the greatest barriers

to cooperation are the lack of time, lack of professional training and education, lack of common interest in cooperation, and negative feelings about actual past cooperation. As solutions to overcome the mentioned barriers, the author suggested the establishment of informal networks and an improvement in shared ambitions [22]. A similar nuance was indicated in a study by Masterson, who concluded that the biggest barrier in such cooperation was the lack of time. Other more frequent barriers are a lack of money, lack of interest in collaboration, and a lack of support from superiors [23].

Sarjeant-Jenkins and Walker investigated collaboration between Canadian academic and public libraries and showed that the greatest challenge was a lack of time, followed by the lack of resources, finding compromises, and funding problems [24]. A very comprehensive study of higher education and public library partnerships in England was commissioned and published by the Arts Council England. This study recognized the lack of resources, funding, and capacity as the biggest obstacles in creating partnership activities. Other frequent barriers are different organizational priorities and the lack of a clear idea about the potential activities. To overcome these barriers, the study indicated that both types of libraries have to find the right point of contact, good funding, sufficient resources that will back the activity, and to try and identify cooperation as a sustainable way of saving costs [25]. Lauddusaw and Wilhem explored the cooperation between Texas academic and public libraries, where they referred to the low amount of such cooperation. They set a premise that the cause of this was a reluctance to connect on the part of public libraries, and the fixation of academic libraries on research rather than activity development [26].

While many studies have been written about successful collaborations between different types of libraries, there is a lack of studies that indicate the problems that libraries may encounter when organizing joint activities. It is interesting that studies on cooperation between university and public libraries are very rare. It has previously been mentioned that more frequent co-operational problems such as a lack of unique ideas, time, and funding can be the main cause of this. Libraries probably face the same challenges while creating CS activities as well as the lack of knowledge about what CS is and what its benefits are for the library community and the community in general. Proposed solutions such as firm and proper management, identifying a common vision, and the constant reminder that all this is undertaken for the needs of library users can serve as guidelines for libraries to overcome these barriers.

2. Materials and Methods

This study was based on two data collection methods: online surveys and interviews. Both methods were not fundamentally related to the main topic of this research, but to CS activities in cooperation with university and public libraries. However, it was assumed that there were a few joint mentioned activities, so the research also included a section on organizational barriers. The main research questions for this study are as follows:

(1)   What are the challenges of conducting CS activities in cooperation with university and public libraries?
(2)   What are the possible solutions to those challenges?

The online survey was conducted in partnership with the University Library of Southern Denmark, which, as a project partner, was in charge of implementing PR1 (project results from project package number one):". The main goal of the survey was to identify attitudes toward open science and CS as well as good practices and challenges. The survey consisted of several parts: OS engagement, CS engagement, assessment of involvement (typology), projects, policies, and partnerships (OS and CS), skills, competences and potential barriers, cooperation and partnerships, and public libraries. The National and University Library in Zagreb participated in the creation of survey questions to explore the cooperation between university and public libraries related to CS activities. The part of the survey for which the National and University Library in Zagreb was in charge referred to the segment about public libraries. This joint survey was conducted because the target respondents were the same—European universities and higher education libraries. Only a

few European examples of the mentioned activities were found by studying the literature as well as talking to project partners and by searching the web, which was the reason why part of the survey questions were related exclusively to co-operational barriers. In that survey, we tried to study the first research question—what are the challenges of conducting CS activities in cooperation with university and public libraries? This was investigated in the part of the survey regarding cooperation and partnership with a special section on partnerships with public libraries.

The online survey was pretested by project partners to improve its quality. It was created by the SurveyXact tool, launched on 15 March 2022, and was available for completion until 15 May 2022. The survey took approximately 15 min to conclude. Considering that all the project partners dispatched the survey to university libraries in their country and even the geographical region, it is difficult to say how many surveys were sent. The total number of responses was 82. All of the partners sent out at least two reminders to university libraries. After removing the responses that were not valid or incomplete, we retained a total of 56 responses. It can be assumed that such a small response to filling out the survey was due to a lack of knowledge and the lack of interest in CS topics as well as because of the language barrier (the study was in English). Over and above, 52 university libraries responded that they never had CS activities co-organized with a public library, which was 92.8% of the total responses.

In the survey, university libraries could voluntarily write their name and 46 of them chose to do that. At least 18 different European countries participated in the research. If we use the UN geoscheme [27] for European regions, most respondents (34.7%) were from Southern Europe, immediately followed Northern Europe (30.4%). A smaller number of libraries from Eastern Europe (19.5%) and Western Europe (15.2%) participated in the survey.

The part of the survey related to barriers included four questions. The first question was a Likert-type scale to rate the organization of CS activities in cooperation with the public library and included eight claims for scaling: lack of resources (staff, time, institutional funding), bad previous experiences in organizing joint events, lack of experience in co-organizing events, different work culture in higher education and public libraries, administrative barriers, financial barriers, insufficient technical equipment, and lack of knowledge about CS. The other three questions were of the closed-end type and they wanted to find out whether university libraries believe that it is possible to bypass these barriers with good organization and cooperation, whether university libraries are well-connected with their local community, and whether they are considering cooperation with the national library in creating CS activities in the future. All three questions could be answered with "yes" or "no".

As far as the results about the actual CS activities in cooperation with university and public libraries, university libraries from four countries (Serbia, Denmark, Portugal, and Finland) answered that they had conducted such activities. To explore the examples of successful cooperation more deeply, personalized emails were sent to all the librarians who answered positively, asking if they would like to participate. Three of them answered positively, so interviews via Microsoft Teams were conducted. Languages that were used for the interviews were English (in the case of Denmark), Spanish (in the case of Portugal), and Croatian (in the case of Serbia), and the interviews were recorded, transcribed, and translated later. The interviewed librarians were, among other things, asked how they overcame the challenges in co-organizing joint CS activities and how the data for research question number two (What are the possible solutions to those challenges?) were collected. Their proposed solutions were analyzed and compared in order to create guidelines to prevent challenges in cooperation.

3. Results and Discussion

This paper was based on partial results from the above-mentioned survey, that is, only answers regarding cooperation between university libraries and public libraries in

organizing CS activities were considered as well as the barriers they encountered. According to the answers, we distinguished two groups of university libraries: (G1)—university libraries that have not yet cooperated with public libraries in organizing citizen science activities, and (G2)—university libraries that have thus far cooperated with public libraries in organizing citizen science activities.

There are 56 university libraries in the first group (G1) and there are only four university libraries in the second group (G2). This number alone indicates that cooperation between university and public libraries is not common and that there are certain barriers.

3.1. First Group (G1)

The first group (G1) had to answer four questions related to the cooperation between university and public libraries in organizing CS activities (reference Supplementary Materials).

The first question (On a scale of 1–5, in which 1 means 'strongly agree' and 5 'completely disagree', how would you rate the organization of citizen science activities in cooperation with your library?) was rated using a Likert scale (1—strongly agree; 5—completely disagree) of the following statements:

1. Lack of resources (staff, time, institutional funding);
2. Bad previous experiences in organizing joint events;
3. Lack of experience in co-organizing events;
4. Different work cultures in university and public libraries;
5. Administrative barriers;
6. Financial barriers;
7. Insufficient technical equipment;
8. Lack of knowledge about citizen science.

For the first statement "lack of resources (staff, time, institutional funding)", the majority of libraries said that it was the biggest obstacle. In contrast, only six libraries agreed with the second statement, "bad previous experiences in organizing joint events". Almost 50% of libraries expressed no opinion here. The results led to the conclusion that the collaboration had not yet occurred or it had but had gone well. For the third statement "lack of experience in co-organizing events", the distribution of answers was almost symmetrical. Almost half of the libraries agreed with the fourth statement "different work culture in university and public libraries". It can be said that there was an attitude that there was a difference in business between the two types of libraries, but that a certain number of libraries considered this to be an obstacle and did not engage in cooperation. Answers to the following three statements "administrative barriers", "financial barriers", and "insufficient technical equipment" confirm the first statement of "lack of resources". It is evident that libraries feel that they lack the resources to initiate CS activities. The last statement (*lack of knowledge about citizen science*) showed that libraries were aware that they did not have enough knowledge about the term CS itself.

Analyzing the responses, the conclusion was made that a large number of responses were undecided (neither agree nor disagree) with regard to certain questions as 30–45% of university libraries checked that option, that is, they did not have an expressed opinion. These data indicate that this research has its limitations (i.e., that at least a third of libraries do not have any opinions about barriers or about overcoming these barriers (Figure 1)).
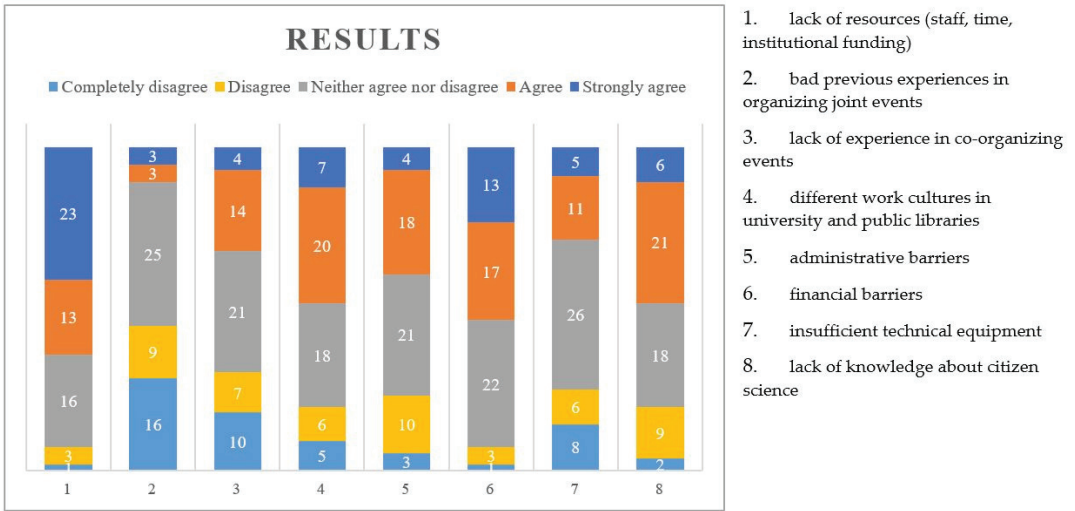
Figure 1. Results.

The analysis of the data led to the conclusion that university libraries see the lack of resources (staff, time, institutional funding) and financial barriers as the biggest obstacle in cooperation with public libraries in organizing CS activities, which is in line with the aforementioned previous research on the co-organization of joint activities of various types of libraries. Previous bad experiences in co-organizing joint activities were rated as the smallest obstacle, which was not the case in previous research (Figure 2).
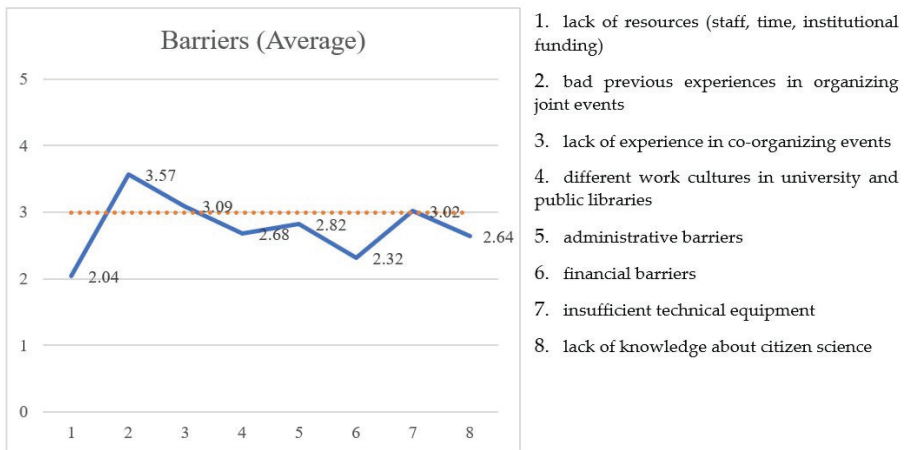
Figure 2. Average score (1 strongly agree; 5 completely disagree).

The second question read: Do you believe that it is possible to bypass these barriers with good organization and cooperation? Only six libraries out of 56 answered negatively, that is, it is evident that the majority of university libraries believe that with certain investments, these obstacles could be overcome.

The third question was: Do you think your library is well-connected with your local community? Twenty-seven libraries answered negatively, and twenty-nine libraries answered positively.

The fourth question read: Are you considering cooperation with a public library in conducting citizen science activities in the future? Twenty-three libraries answered negatively, and thirty-three answered positively.

The analysis of the last two questions led to the conclusion that there was a slight difference in the responses of libraries that were thinking about cooperation with national libraries and those not thinking about cooperation. It shows that university libraries that did not think about cooperation with public libraries answered seven out of eight options from the first question closer to the mean value (3), that is, they had a less pronounced attitude. Only for one option (*bad previous experiences in organizing joint events*) did they express greater agreement than the other group (i.e., libraries that are considering cooperation with public libraries in organizing citizen science activities) (Figure 3)

Legend:
1. lack of resources (staff, time, institutional funding),
2. bad previous experiences in organizing joint events,
3. lack of experience in co-organizing events,
4. different work culture in university and public libraries,
5. administrative barriers,
6. financial barriers,
7. insufficient technical equipment,
8. lack of knowledge about citizen science.
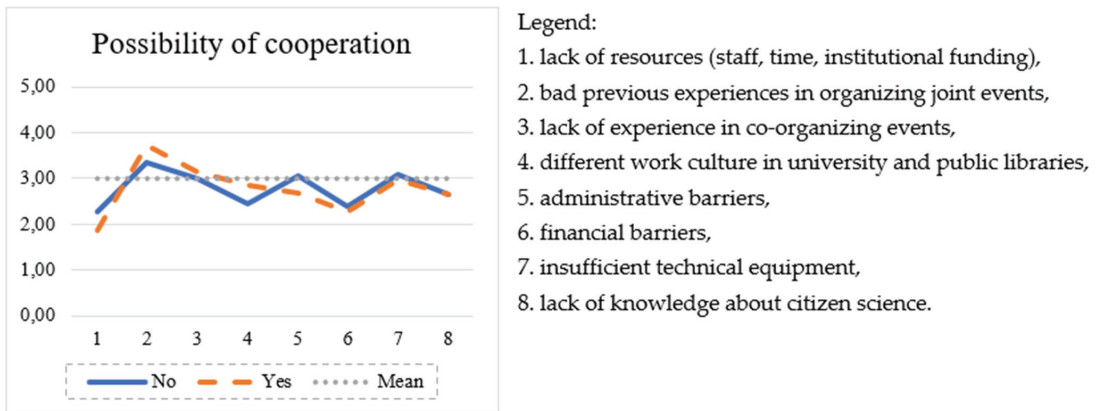
Figure 3. Possibility of cooperation.

In conclusion, libraries from group G1 did not have a strong opinion about cooperation between universities and public libraries in organizing CS activities. However, the lack of resources (staff, time, institutional funding) and financial barriers were cited as some of the more important obstacles. It was assumed that for this reason, university libraries did not even decide to establish cooperation with public libraries. This conclusion is in line with previous research that stated that the lack of resources was the most common obstacle in organizing or co-organizing CS activities.

3.2. Second Group (G2)

The second group (G2) included only four university libraries that have so far cooperated with public libraries in the organization of CS activities (reference Supplementary Materials). One library was in Finland, another from Denmark (D1), the third from Portugal (P1), and the fourth from Serbia (S1). In response to the question from the sent questionnaire that referred to their collaboration with public libraries (*Please evaluate the Citizen Science collaboration between your library and the public library*), they rated the following statements using a Likert scale (1—strongly agree; 5—completely disagree):

1. It was easy to organize citizen science activities in collaboration with the public library;
2. The implementation of citizen science activities in collaboration with the public library was successful;
3. Participants in citizen science activities were satisfied with the activity/activities;
4. We plan to organize more citizen science activities together in the future.

The mean values of the responses were taken for analysis as follows:

1. The mean value for the first statement was 2.00;
2. The mean value for the second statement was 1.75;

3. The mean value for the first statement was 2.25;
4. The mean value for the first statement was 1.75.

The obtained results showed that the libraries evaluated the cooperation with the public libraries relatively positively.

However, in order to obtain more precise results, it was decided to invite the aforementioned libraries that stated having conducted citizen science activities and agreed to be contacted for an interview through which the aforementioned cooperation would be explained in detail, and the obstacles encountered in the organization of these activities would be additionally discussed and potential solutions found.

Unfortunately, the library from Finland did not respond to an additional invitation to participate in an in-depth interview to which the other three libraries agreed. All interviews were conducted during June 2022. The interview with the university library from Portugal (P) and the one from Denmark (D) lasted half an hour, and the interview with the library from Serbia (S) lasted an hour.

The results obtained after the conducted interviews confirmed some of the claims of group G1.

One of the most prominent obstacles was the lack of resources (staff, time, financial funding). Namely, even during the collaborations, it was confirmed that there were certain obstacles in the form of lack of time, staff, etc. in the joint organization of such activities:

P1: "There were approximately 10 organizers who donated their personal time and weekends. Cooperation among colleagues was never a problem because in each region in Portugal, there is a library network where all types of libraries cooperate together."

P1: "Another difficulty was to secure people who would be in charge of security because the activity was held in nature. We told everyone to participate at their own risk, but people still came. Ensuring security was complicated."

P1: "And one difficulty is that all of us have other jobs and we all depend on hierarchical orders to be able to advance and it is slightly difficult to reconcile everything, but we will see."

D1: "Then another barrier is mapping of skills at libraries. In addition, to be honest, I think that could be a barrier in research libraries as well, because I can see from my own library that a lot of the skills our staff have are transferable skills that we use in citizen science projects, and I will bet that at many Danish public libraries the same skills are there."

The interviewed libraries confirmed that one of the obstacles was the lack of financial resources. Namely, it is evident that the activities were often carried out through the enthusiasm of the organizers, but the established system requires secured funds:

S1: "We have no finances."

S1: "On the other hand, we do not have any big expenses to organize it, except travel expenses where we go or people who come to us. Serbia is not a very big country, we have an incoherent system, librarians know each other—a lot is based on personal acquaintances. I know a librarian there, this one knows the director there, this one knows the deputy director there, when a project is being made—we will work with them easily, let us take them, these are interested in digitization, these have manuscript materials, these want to learn. Someone knows someone who knows someone. Here in Serbia, things mostly work on the basis of personal acquaintance, and there is no money."

D1: "In Denmark, public libraries have been cut almost 20% in budgets the last few years. [. . .] So, in the public sector in Denmark, and it is also true. For example, in our own library we have to do more for less money."

P1: "The major barrier was the financing of the activity because the support was given by the University of Beira."

Only one of the libraries stated that insufficient IT equipment was an obstacle when organizing these activities, but this obstacle could be compared to the first two.

S1: "Indecent technical equipment. Good IT support is key."

Through interviews, it was proven that the awareness of CS itself was insufficiently expressed in public libraries and that there was certainly room for the development of awareness.

S1: "We did not even understand that what we do is citizen science."

D1: "I think there is a big barrier in knowledge in public libraries: What is citizen science? Why should we do it? Why is it good or potentially good, and why does it live up to the things we should be doing? Therefore, advocacy is a big barrier."

D1: "They are good of all kinds of things. However, they are not aware that they that it fits into citizen science."

P1: "As for prior knowledge in the organization of citizen science activities, only I and my colleague were familiar with open science."

In the same way, the university libraries confirmed that the cooperation was very fruitful and that the public libraries had resources for cooperation on activities, that is, CS projects. Public libraries have better contact with the local community, good connections with the media, and they have special skills...

S1: "We received participants of those activities. Our users are very limited, we do not have that much contact with citizens. Through public libraries, we come into contact with the local community. The National Library knows its users well, the different types of audiences."

D1: "They are good at doing events. They're doing good at communicating. They are good at doing evaluation, they are too good at doing reading groups. They are good of all kinds of things. However, they are not aware that they that it fits into citizen science."

D1: "I think there is an enormous potential not only in research libraries, but also in public libraries. In addition, when I discuss with library management and also the head of the Danish Public Library Association, why, why, why are there public libraries? Well, it is to enhance knowledge society, it is for democratic conversation to happen. It's that we have free and clear knowledge for everybody. It's available for everyone. It's to mitigate fake news. I mean, those are some of the wise for citizen science. It is exactly the same for public libraries."

Ultimately, the surveyed libraries from group G2 offered certain solutions for cooperation between university and public libraries in organizing CS activities. We can explicitly state the following suggestions:

1.   Mapping of transferable skills in libraries:

D1: "Then another barrier is mapping of skills at libraries."

D1: "I think primarily skills from my own stuff at the research library, they have been fairly trained, they have been on a number of projects. I think their skills will, we always learn something to pick something new up, hmm but we lack the competences that we have at our library for every single employee and make sure that they can live up to that so we don't put people in the wrong position at the wrong product. However, I think something that could easily be worked on is a more systematic skill set for public libraries."

2.   Strategy and advocacy:

D1: "And then there is a truly interesting component. That is leadership and prioritization. Because you cannot go out to an employee at a public library and say we think you should work on this citizen science project. If it is not a part of their strategy, if they're not trained for it, and if they're not told that they should do it, it is not a voluntary exercise. In addition, you need to be very on point. We are prioritizing this just as we do as lending books out, for example."

D1: "I think there is a big barrier in knowledge in public libraries: What is citizen science? Why should we do it? Why is it good or potentially good, and why does it live up to the things we should be doing? Therefore, advocacy is a big barrier."

3. Collegiality:

S1: "The teams we work in are designed so that the people in those teams get along well. Collegiality is crucial."

D1: "Yes, I would say when we do it, it is excellent. I would wish that we do it a lot more. I would wish that we had public libraries as partners in every single project that we had. Right now it is more on when there is a really good fit. In addition, it is usually by libraries we know very well."

In conclusion, university libraries that cooperated with public libraries in organizing CS activities have a positive attitude about cooperation, although they state that certain obstacles exist, which is stated as a lack of resources and financial barriers, but also a lack of awareness and knowledge about what citizen science is. For even better cooperation, they suggest the mapping of transferable skills to accurately understand the human resources in libraries as they believe that better leadership and prioritizing are needed so that CS activities become the basic activities of the library, and they believe that collegiality is crucial to carry out these activities.

These results show that cooperation between different types of libraries has a positive impact and should be undertaken more. Of course, as previously said, more can be done, but this study showed that different types of libraries bring greater value to the projects.

4. Conclusions

Project *Citizen-Enhanced Open Science in Southeastern Europe Higher Education Knowledge Hubs* is a project led by LIBER (Ligue des Bibliothèques Européennes de Recherche—Association of European Research Libraries). One of the project's goals is to train and upskill librarians in Southeastern European countries in the field of open science and CS because such activities are lacking in that part of Europe.

Although cooperation between several library types has a positive effect on their management, libraries can face various barriers in the joint implementation of services, programs, and activities. Regardless of the fact that it may be a matter of collaboration between different types of libraries, certain barriers are more common than others.

While many studies have been written about successful collaborations between different types of libraries, there is a lack of studies that indicate the problems that libraries may encounter when organizing joint activities. It is interesting that studies on cooperation between universities and public libraries are very rare. Co-operational problems such as a lack of unique ideas, time, and funding can be the main cause. Libraries probably face the same challenges while creating CS activities as well as the lack of knowledge about what CS is and what its benefits are for the library community and the community in general. In the literature, there have been some proposed solutions such as firm and proper management, identifying a common vision, and the constant reminder that all this is carried out for the needs of library users that can serve as guidelines for libraries to overcome these barriers.

This study was based on two data collection methods: an online survey and interviews. Both methods were not fundamentally related to the main topic of overall research but to CS activities conducted with the cooperation of universities and public libraries, however, it was assumed that there are a few mentioned joint activities, so the research also included a section on organizational barriers.

The analysis concluded that a very small number of higher education libraries (four out of 56) have had any cooperation with public libraries in organizing CS activities, thus far. Higher education libraries that have not yet cooperated with public libraries in the organization of CS activities state that their biggest obstacle is the lack of resources. Likewise, they believe that insufficient knowledge of the term CS is an obstacle. Additionally, from the results, it can be read that there is a certain prejudice against cooperation with public libraries, mostly due to the opinion that there is a different work culture.

Libraries that have organized CS activities in cooperation with public libraries confirm that the lack of resources and insufficient knowledge of the concept of CS is the largest obstacle. Furthermore, they state that cooperation with public libraries was very fruitful, precisely because of the different work culture, that is, the different roles of public libraries in society. Different skills, different stakeholders, and different user groups all contribute to better and more successful cooperation.

In conclusion, university libraries that cooperated with public libraries in organizing CS activities have a positive attitude about cooperation, although they state that certain obstacles doo exist. They state that it is a lack of resources and financial barriers, but also a lack of awareness and knowledge about what citizen science is. For even better cooperation, they suggest the mapping of transferable skills to accurately understand the human resources in libraries, as they believe that better leadership and prioritizing are needed so that CS activities become the basic activities of the library, and they believe that collegiality is crucial for carrying out these activities, which is in line with previous solutions that the management should be firmer, that the common vision should be identified, and that all of these are conducted for the needs of library users.

These results indicate that greater cooperation between universities and public libraries in organizing CS activities is necessary precisely because of the different roles of these libraries in society. Raising awareness of the concept of CS and its importance needs to be carried out through advocacy among all stakeholders—librarians, management, the public, and financiers—to reduce all of the mentioned obstacles and increase the reach of both libraries.

References

1. Haklay, M.; Dörler, D.; Heigl, F.; Manzoni, M.; Hecker, S.; Vohland, K. What is citizen science? The challenges of definition. In *The Science of Citizen Science*; Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., Wagenknecht, K., Eds.; Springer: Cham, Switzerland, 2021; pp. 13–33.
2. ECSA. *Ten Principles of Citizen Science*; ECSA (European Citizen Science Association): Berlin, Germany, 2021.

3.   Cooper, C.B.; Hawn, C.L.; Larson, L.R.; Parrish, J.K.; Bowser, G.; Cavalier, D.; Dunn, R.R.; Haklay, M.; Gupta, K.K.; Jelks, N.O.; et al. Inclusion in citizen science: The conundrum of rebranding. *Science* **2021**, *372*, 1386–1388. [CrossRef]
4.   Zaken, D.B.; Gal, K.; Shani, G.; Segal, A.; Cavalier, D. Intelligent Recommendations for Citizen Science. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Event, 2–9 February 2021; Volume 35, pp. 14693–14701.
5.   Ben Zaken, D.; Segal, A.; Cavalier, D.; Shani, G.; Gal, K. Generating Recommendations with Post-Hoc Explanations for Citizen Science. In Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, Barcelona, Spain, 4–7 July 2022; pp. 69–78. Available online: https://dl.acm.org/doi/pdf/10.1145/3503252.3531290 (accessed on 7 November 2022).
6.   Cavalier, D.; Nickerson, C.; Stanton, D. The Library and Community Guide to Citizen Science: Understanding, Planning, and Sustaining Ongoing Engagement in Citizen Science at Your Library. Scistarter; Arizona State University. 2019. Available online: https://orrery-media.s3-us-west-2.amazonaws.com/curated/LibraryGuideFebruary.pdf. (accessed on 7 November 2022).
7.   Van Hyning, V.; Algee, L.; Jones, M.; Osborn, C.; Owens, T.; Seroka, L.; Shelton, A. By the People Crowdsourcing Datasets from the Library of Congress. *J. Open Humanit. Data* **2022**, *8*, 5. [CrossRef]
8.   Library of Congress Official Website. About by the People. Available online: https://crowd.loc.gov/about/ (accessed on 7 November 2022).
9.   Cowles, K.; Sheppard, M.; Waltman, E.; Wilson, T.K. Crowdsourcing and collaboration from coast to coast: NNLM's# CiteNLM Wikipedia edit-a-thons. *J. Electron. Resour. Librariansh.* **2020**, *32*, 267–275.
10. Lasky, M.; Parsons, A.; Schuttler, S.; Mash, A.; Larson, L.; Norton, B.; Pease, B.; Boone, H.; Gatens, L.; Kays, R. Candid Critters: Challenges and Solutions in a Large-Scale Citizen Science Camera Trap Project. *Citiz. Sci. Theory Pract.* **2021**, *6*, 1–17. [CrossRef]
11. San Francisco Library Official Webpage. Citizen Science. Available online: https://sfpl.org/locations/main-library/environmental-center/citizen-science (accessed on 6 November 2022).
12. Vohland, K.; Göbel, C.; Balázs, B.; Butkevičienė, E.; Daskolia, M.; Duží, B.; Schade, S. (Eds.) Citizen science in Europe. In *The Science of Citizen Science*; Springer: Cham, Switzerland, 2021; pp. 35–53.
13. Mahr, D.; Göbel, C.; Irwin, A.; Vohland, K. Watching or Being Watched: Enhancing Productive Discussion between the Citizen Sciences, the Social Sciences and the Humanities. In *Citizen Science: Innovation in Open Science, Society and Policy*; Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J., Bonn, A., Eds.; UCL Press: London, UK, 2018; pp. 99–109.
14. Heigl, F.; Kieslinger, B.; Paul, K.T.; Uhlik, J.; Dörler, D. Toward an international definition of citizen science. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 8089–8092. [CrossRef] [PubMed]
15. Ignat, T.; Ayris, P.; Juan, I.L.I.; Reilly, S.; Dorch, B.; Kaarsted, T.; Overgaard, A.K. Merry work: Libraries and citizen science. *Insights* **2018**, *31*, 1–10. [CrossRef]
16. Moyle, M.; Tonra, J.; Wallace, V. Manuscript Transcription by Crowdsourcing: Transcribe Bentham. *Liber Q.* **2011**, *20*, 347–356. [CrossRef]
17. Wiederkehr, S. Open data for the crowd: An account of citizen science at ETH Library. *Liber Q.* **2018**, *29*, 1–10. [CrossRef]
18. Cigarini, A.; Bonhoure, I.; Vicens, J.; Perelló, J. Public libraries embrace citizen science: Strengths and challenges. *Libr. Inf. Sci. Res.* **2021**, *43*, 101090. [CrossRef]
19. The CeOS_SE Project: Citizen-Enhanced Open Science in Southeastern Europe Higher Education Knowledge Hubs. Available online: http://ceosse-project.eu/ (accessed on 17 July 2022).
20. The CeOS_Se Project Documentation. 2022.
21. Fitzgibbons, S.A. School and Public Library Relationships: Deja Vu or New Beginnings. *J. Youth Serv. Libr.* **2001**, *14*, 3–7.
22. LaMaster, J. Collaboration of Indiana public and school media center youth services: A survey analysis of current practices. *Indiana Libr.* **2005**, *24*, 38–41.
23. Masterson, C.A. School and Public Library Collaboration: Identifying Facilitators and Barriers. Master's Thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, April 2012. Available online: https://cdr.lib.unc.edu/concern/masters_papers/0v838441d (accessed on 10 August 2022).
24. Sarjeant-Jenkins, R.; Walker, K. Library Partnerships and Organizational Culture: A Case Study. *J. Libr. Adm.* **2014**, *54*, 445–461. [CrossRef]
25. Pekacar, K. Independent Mind. Higher Education and Public Libraries: Partnerships Research. Report, Arts Council England, Manchester, Great Britain. 2018. Available online: https://www.artscouncil.org.uk/sites/default/files/download-file/Public%20Libraries%20and%20HE%20report.pdf (accessed on 13 August 2022).
26. Laddusaw, S.; Wilhelm, J. Yours, Mine, Ours: A Study of a Successful Academic & Public Library Collaboration. *Collab. Librariansh.* **2018**, *10*, 30–46.
27. UN Statistics Division: Geographic Regions. Available online: https://unstats.un.org/unsd/methodology/m49/ (accessed on 17 July 2022).

MDPI