

DESIGN FOR OPEN ACCESS PUBLICATIONS IN EUROPEAN AREAS FOR SOCIAL SCIENCES AND HUMANITIES

User-driven surveys on open access scholarly communication (may-june 2017)

Grant Agreement number	: 731031
Project acronym	: OPERAS-D
Project title	: Design for Open Access Publications in European areas for Social Sciences and Humanities
Funding Scheme	: INFRASUPP-03-2016
Project's coordinator Organization	: CLEO-CNRS
E-mail address	: pierre.mounier@openedition.org
Website	: http://operas-eu.org



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No731031

Objective	2
Executive summary	2
Description of the surveys	2
Topics addressed	2
Overview	3
Transverse questions	5
R&D set of questions	5
Multilingualism	7
Seamless services	9
Surveys details	12
Survey for the researchers	12
Survey for the publishers	16
Survey for the libraries	17
Survey for the socio-economic actors	19

Objective

In order to identify the services the OPERAS infrastructure will have to develop and implement in the future, OPERAS consortium planned to conduct an online survey during OPERAS-D project. Addressed to the academic community and beyond, the survey was meant to collect information on:

- the current practices regarding Open Access;
- the evaluation of existing services;
- the missing services;
- the level of interest for integrated new services.

Executive summary

Description of the surveys

So as to address the specific needs of each stakeholder in the field, the investigation has been divided in five different surveys:

- Researchers: <https://survey.openedition.org/index.php/831687>
- Publishers: <https://survey.openedition.org/index.php/468227>
- Librarians: <https://survey.openedition.org/index.php/212534>
- Funders: <https://survey.openedition.org/index.php/578782>
- Socio-economic actors¹: <https://survey.openedition.org/index.php/214336>

The surveys took the form of an online questionnaire using the Limesurvey open source software (<https://www.limesurvey.org/>). They were disseminated during 1 month through the OPERAS network of partners (websites, social media, mailing lists).

The questionnaire contains both open and closed questions in order to collect qualitative and quantitative results. In particular, specific questions were triggered when negative answers (level of interest or quality assessment) were given. Mandatory questions ensured the main questions were answered.

Topics addressed

The OPERAS consortium organizes its work through working groups addressing the main challenges of the OPERAS infrastructure. These working groups helped to define the sections of the questionnaires. Here is the list of the working groups and the corresponding questions:

- *Communication*: questions about knowledge and usage of Open Access publications;
- *Tools R&D*: set of questions regarding new or future advanced services (crosslinking, discovery, annotation, etc.);

¹ For communication purposes, this specific survey was disseminated as “Survey for the general public”.

- *Standards*: questions addressing the topic of data and/or metadata management;
- *Business models*: questions on this topic were only for the publishers' survey;
- *Best practises*: various questions to investigate the publishing workflows;
- *Multilingualism*: questions regarding the multilinguistic publishing and metadata;
- *Platforms*: new integrated platforms that will be developed by the OPERAS infrastructure.

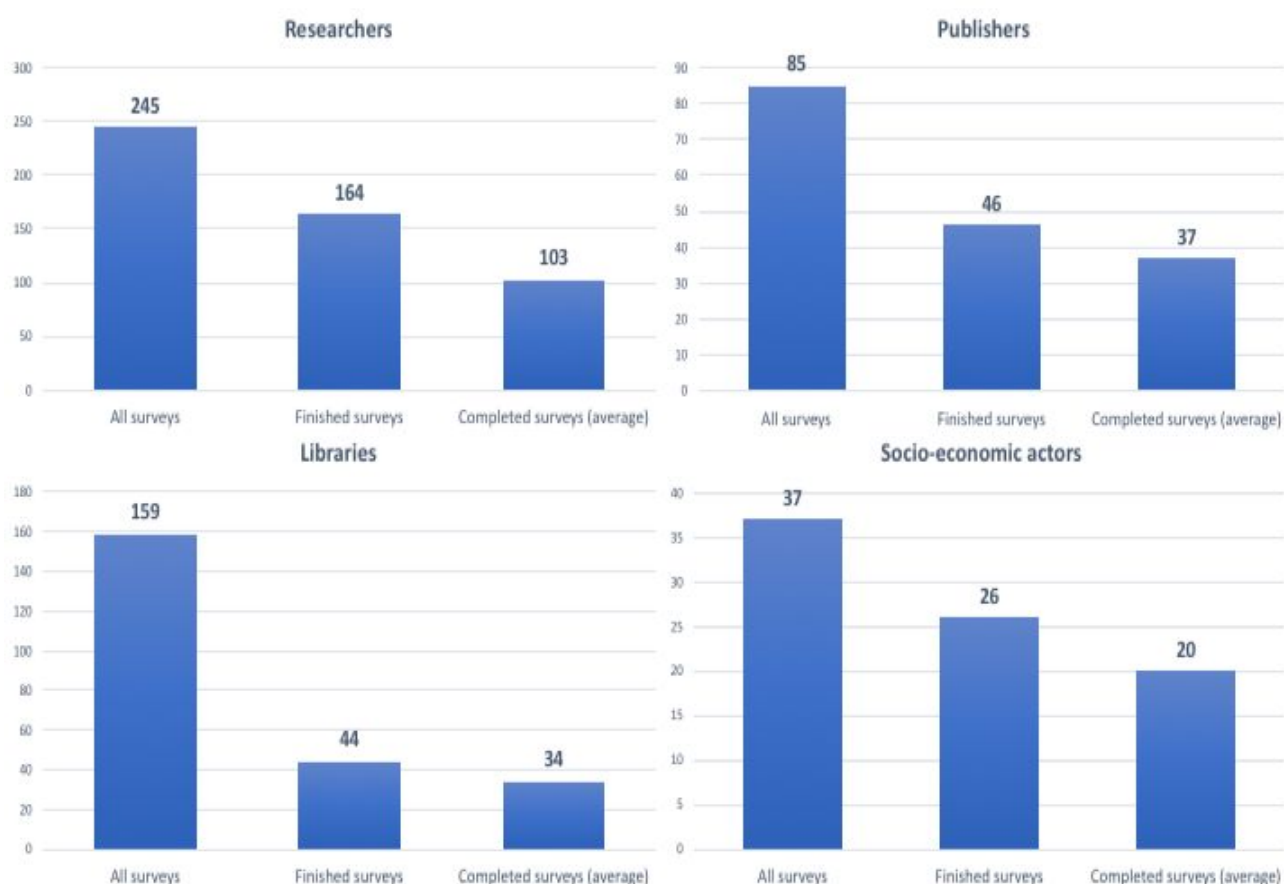
Some of these sections made it possible to establish transverse questions for various stakeholders. The “tools”, “multilingualism” and “platforms” sections, even if with some partial adaptation, have been proposed in different surveys.

The report present these questions in a specific section.

Overview

Overall, the participation was sufficient, especially regarding the main targets of OPERAS infrastructure, researchers, libraries and publishers. The dissemination towards the various stakeholders worked rather well and, whenever the survey is completed, the questions are thoroughly answered. The charts below show the results for the four surveys with useful results². The first column “All surveys” counts every started survey; the second column “Finished surveys” counts every survey consulted till the last page; the third column “Completed surveys” is an average of the surveys with answers to all questions.

² The chart doesn't show the “funders” survey where only 3 answers have been collected, due to a non efficient dissemination process. A new serie of usage surveys more individualized will give us the possibility to correct this issue.



Participation to the OPERAS surveys

The distribution by country of the answers offers an acceptable representativity of the european countries but only partially reflects the countries present in the OPERAS consortium.

Researchers	Publishers	Libraries	Socio-economic actors	Funders
France (35%)	France (35%)	France (27%)	Greece (33%)	Austria (66%)
Belgium (24%)	Germany (16%)	Germany (13%)	France (16%)	Luxembourg (33%)
Germany (8%)	Other (13%)	Norway (9%)	Slovenia (12%)	--
Italy (7%)	Italy (13%)	Greece (9%)	Poland (8%)	--
Portugal (6%)	Portugal (5%)	Sweden (7%)	Germany (4%)	--

Regarding the content of the answers, we can observe that open access publishing services are well known and the satisfaction about the quality of the publications and the services is generally good. Nevertheless, some confusion still persists between open and free access and open access is often related to Article Processing Charges and Book Processing Charges issues. These aspects legitimate the open access advocacy which is part of the “Communication” OPERAS working group as well as the working group dedicated to open access business models.

More directly interesting for the OPERAS infrastructure, questions about the future platforms revealed a great interest on the part of the different stakeholders.

Transverse questions

This section gathers specifically the questions or set of questions which were common to different surveys.

R&D set of questions

The “R&D” or “Tools” section listed various advanced services specific or adapted to digital open access publishing. Some of them are not yet well known and the survey gave the opportunity to provide some information and useful links.

The list of services investigated is mostly based on the HIRMEOS³ H2020 project which is related to the OPERAS infrastructure as its proof-of-concept.

The implementations currently in progress within the HIRMEOS project are:

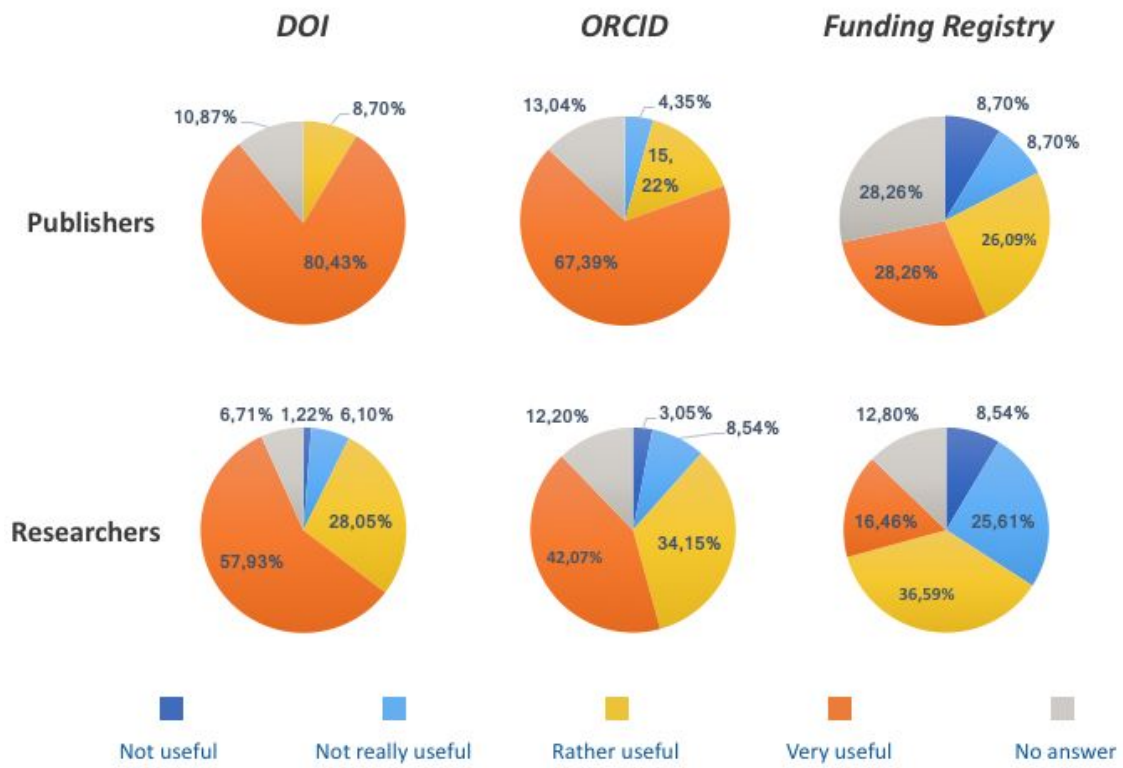
- Identification: DOI, ORCID, Funding registry
- Online annotations
- Entity recognition for indexing
- Enhanced and alternative metrics

This set of questions investigated the level of interest of 3 stakeholders: researchers, publishers and libraries.

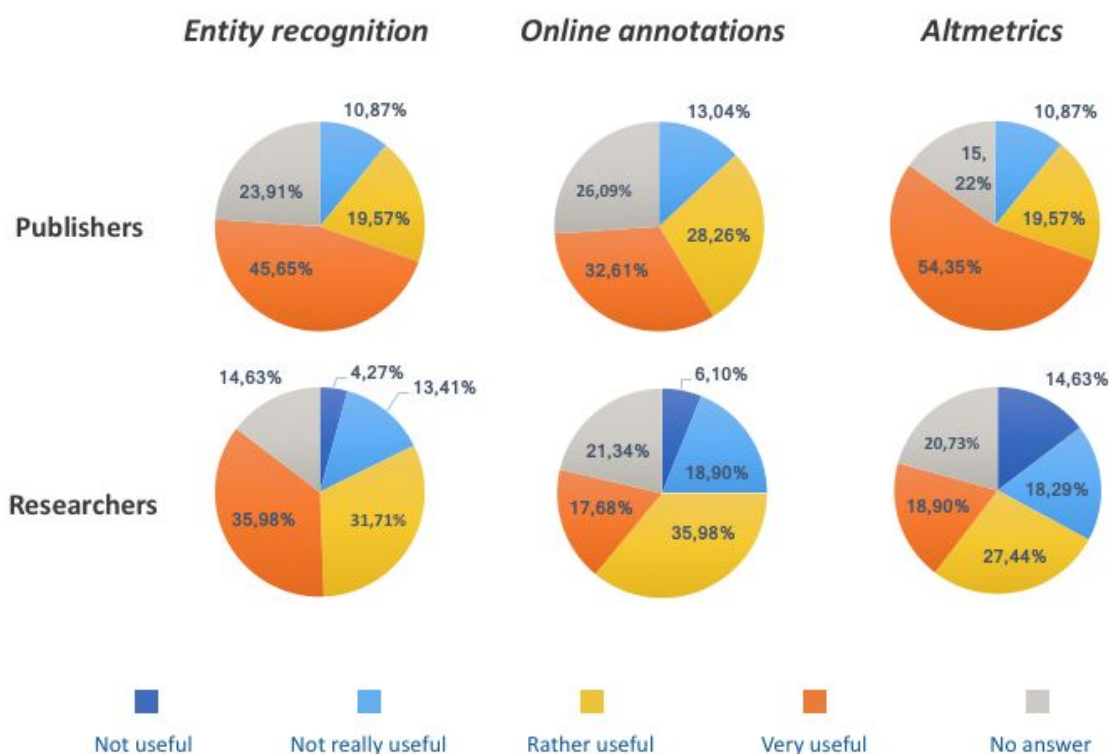
The results show a good level of interest and, if the numbers differ a little, the proportions are comparable from one stakeholder to another.

The charts below show the results for researchers and publishers for the question: “*In the list of new or enhanced services Operas will provide, could you tell us which ones could be useful for your own activity?*”.

³ See: www.hirmeos.eu



Level of interest for new services (1)



Level of interest for new services (2)

As we can see, there is in most cases a majority of positive answers, the amount of which partly varies in relation with the knowledge of each technology. DOIs and ORCID's are rather well known and their implementation corresponds to a real need in the community. The other services, because of their specificity or because they are disruptive require some communication effort but all already arouse interest among the stakeholders.

Multilingualism

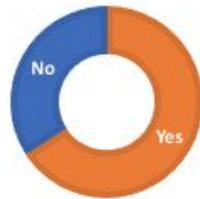
The section "Multilingualism" was present with adaptations in the various surveys. This rather short set of questions aimed at gathering first raw information about the multilingual usage (in publishing and in reading as well). It was also addressing the usage of metadata in several languages.

Authors and publishers engage quite often in multilingual publication, as shown in the figures ⁴ below:

⁴ Blue=No. Orange=Yes.

Question "Do you publish in several languages?"

Researchers:



Publishers:



Libraries:

Question "Do you provide access to OA publications in the SSH in several languages?"

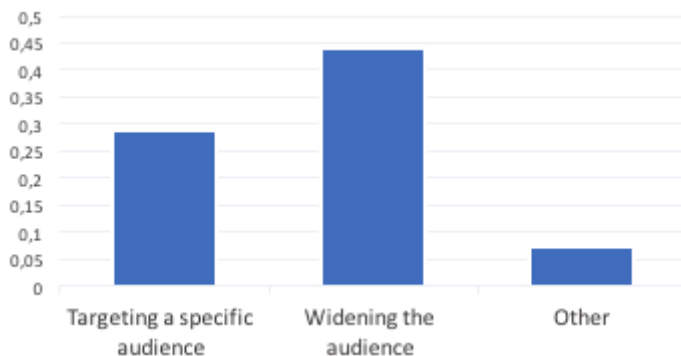


3

Multilingualistic publishing and dissemination.

When investigating the motivations of the researchers for publishing in several languages we find that they almost equally do it to widen the audience (this answer probably indicates the choice of english) as well as to target a specific audience. The "other" cases refer mostly to constrained - directly or not - multilingualism.

Here are the results for the question "To which purposes did you publish in several languages?":



Reasons for multilingualistic publishing by researchers.

Another question about the several languages used by the researchers, the publishers or the libraries complete the previous observation. After the main european languages (English,

French, German) come the other European languages to confirm that, in the SSH field, persists a rather high importance of the national language.

Regarding metadata in several languages, the results show that a majority of publishers (60%) are providing them so as to allow for multi-language search. However, the usage of multilingual metadata seems to be rather limited in the responses of the libraries (around 30%).

Seamless services

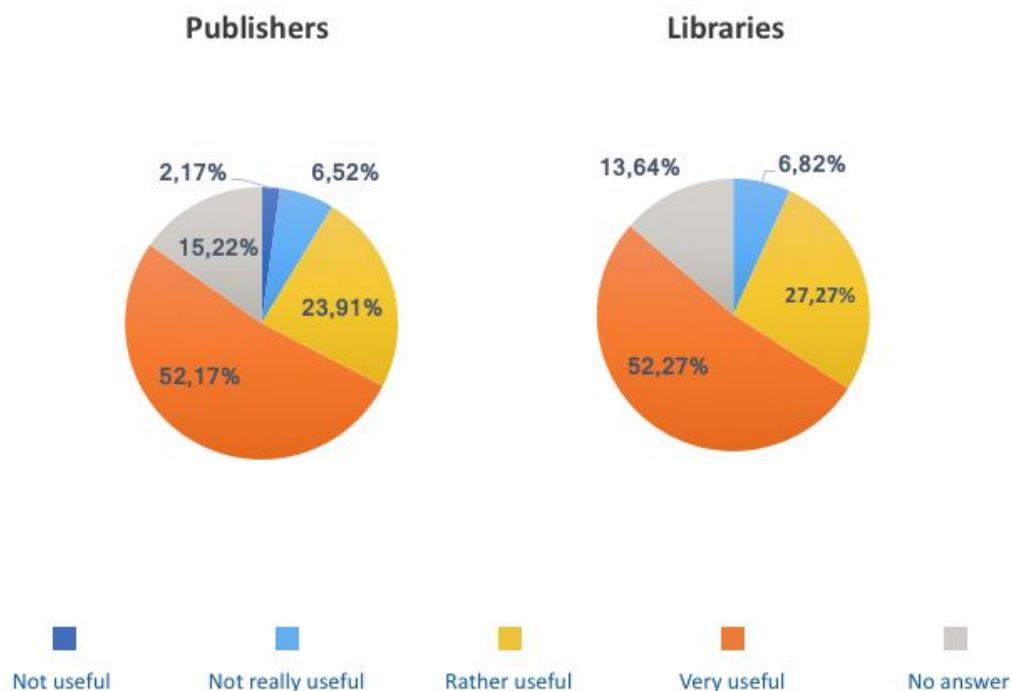
The integrated platforms are here intended as a set of integrated services corresponding to one specific area of academic interest in the digital communication field. Practically, the OPERAS infrastructure will set up three platforms providing enhanced and complete services:

- the DOAB as a certification platform;
- a discovery platform to index all research materials in the SSH;
- a platform to foster collaboration between researchers and socio-economic actors.

The questions explained the content of these new platforms and were asked to different stakeholders whenever meaningful. The results are showing a very high interest in general. Negative answers ("not useful" and "not really useful") triggered a specific open question so as to better know the needs of the participants.

Here are the results to the question about the certification service for publishers and libraries: *"The OPERAS project is planning to launch a platform based on the existing DOAB platform (<http://www.doabooks.org>). The new platform will provide a complete certification service for open access monograph publishing platforms: a classification system of peer-reviewing procedures, a list of open licences, and a tool to manage peer-review descriptions.*

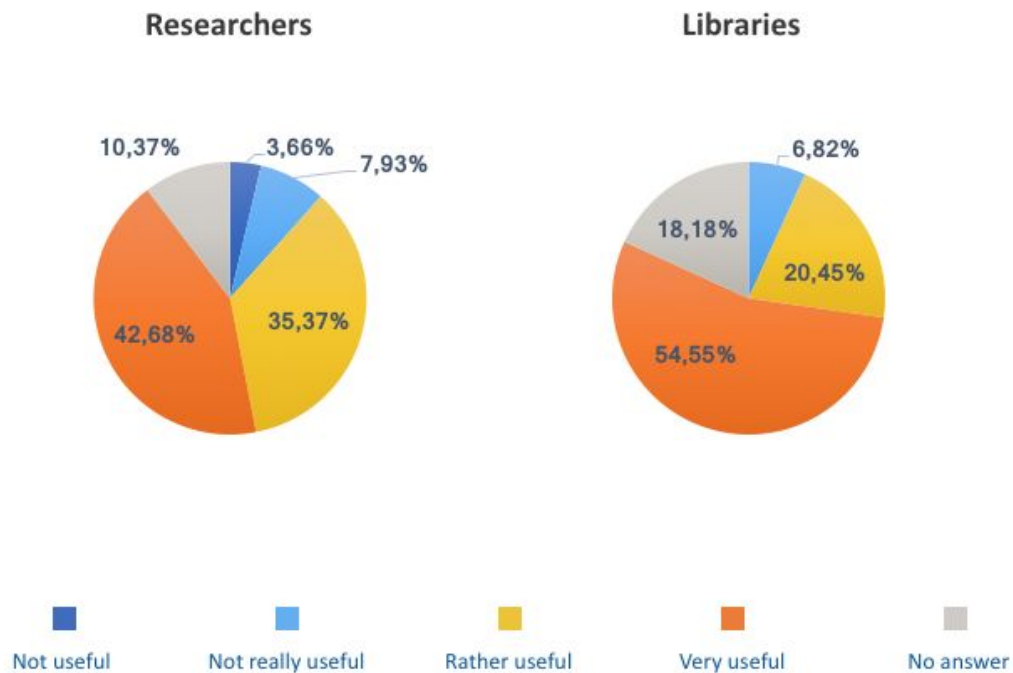
Based on your current activities and needs, how would you evaluate this prospect?"



Interest for the certification service

One of the comments made by a publisher recommended to widen the range of the service, stating that *“yes, there are needs to be continued work in the area of legitimizing open access material that has been properly peer-reviewed and I support such initiatives but special databases that list open access material seems to me be an unusual waste of limited resources”*. This comment gives indication to develop DOAB collaboration with more comprehensive and generic databases and to serve as a hub rather than an end point for open access content.

Below are represented the results to the question regarding the discovery service for researchers and libraries: *“The OPERAS project is also planning to implement a discovery platform dedicated to SSH OA. This platform could search not only through books and journals but also through blog posts and other social media. It will also index sources and data. The platform will be based on the existing Isidore platform (<https://www.rechercheisidore.fr>), which is using various reference sets to enhance resources description and discoverability. How would you evaluate this prospect?”*.



Interest for discovery service

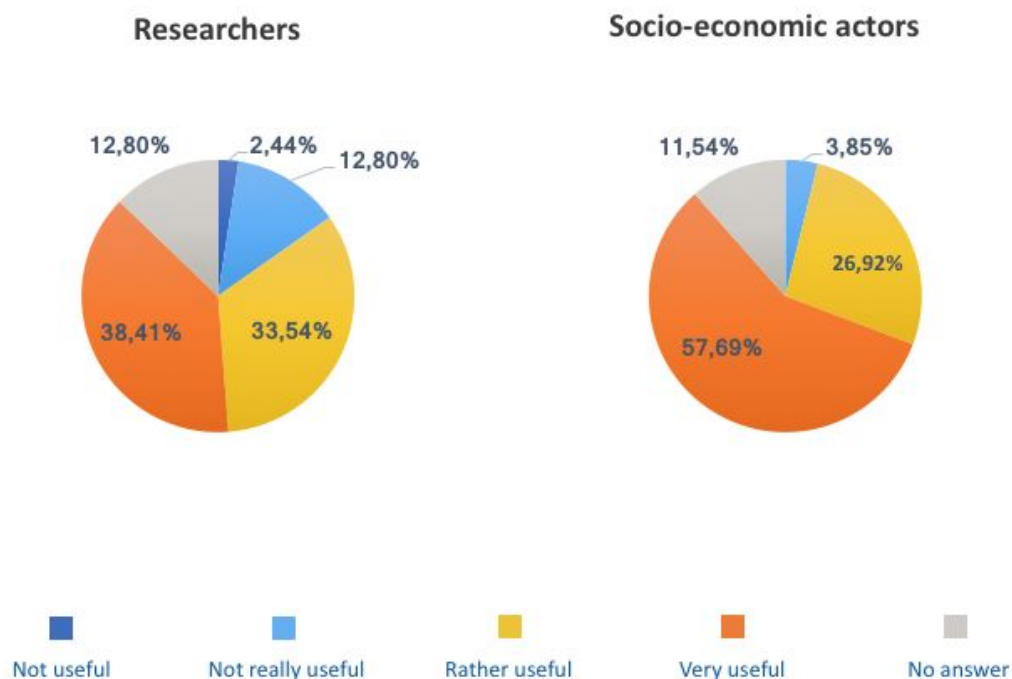
The comments contained both negative and positive answers, as shown by these two examples:

- *“I would go for a clean discovery platform with peer-reviewed only contents. And I would avoid giving the user endless results of disparate sources.”*
- *“Actually, I think this is very, very useful. But there is no feedback button for a positive answer, so I hit “not really useful” to add that I would like to integrate SSH OA into our local discovery system.”*

Once again, the interest for such a centralized service is obvious but users want it to be connected and not isolated. The discovery service must be a hub, not a dead-end. It must be able to connect and be used in local contexts to address the needs of specific communities.

The question about the collaborative service was the following for researchers and socio-economic actors: *“OPERAS intends to facilitate collaboration between academics and journalists, SMEs, administrations and citizen groups. There are plans to launch a future platform where researchers and socioeconomic actors could work together during the lifetime of research projects related to societal challenges and collectively produce and share materials and data that could be exploited and reused on a wide basis. Based on your current activities and needs, how would you evaluate this prospect?”*

In this third case also, there is a large majority of “rather useful” and “very useful” responses, like we can see in the figures below.



Interest for the collaborative service

Like in the previous question, the responses of the researchers are slightly less positive than the ones of the other stakeholder. Some comments partly explain why: “...in terms of data, it is not released until the end of a research project (if it is released at all). Would you seriously trust (all) journalists with raw data?...”. On the other hand, the socio-economic actors gave some indications on the type of guidance they would need for this service: “There should be probably an educational program going along with launching such platform. There is a need for academics to develop media literacy skills to use this platform with ease “. These two answers, together, are highly interesting: they reflect the current lack of trust and poor consideration researchers and socio-economic actors, particularly the media, have towards each other.... It gives use indication on the need to support the technical development of the platform by a strong mediation work and dedicate resources to that dimension of the platform.

Surveys details

The most important results come from the researchers’, the publishers’ and the libraries’ surveys. The responses from socio-economic actors are not representative enough and the responses from the funders are too few.

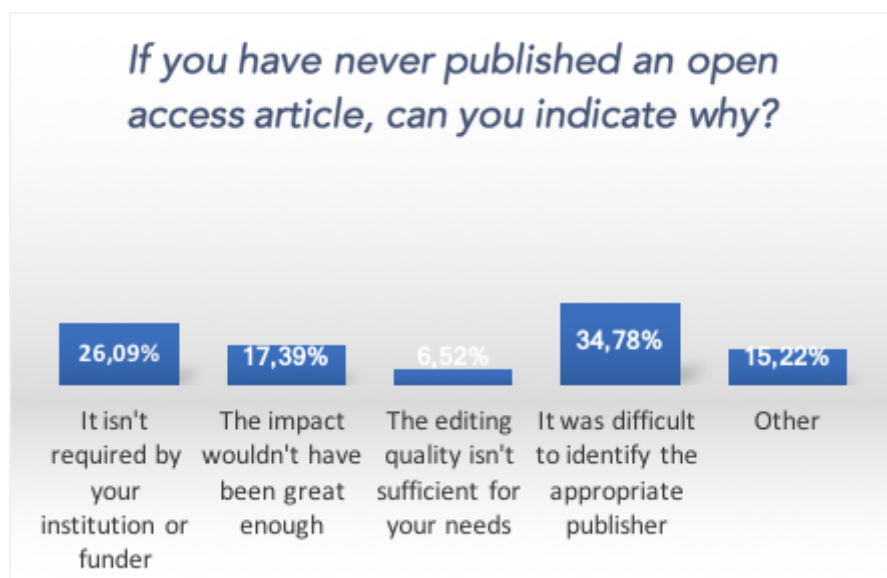
Survey for the researchers

The panel of participants in this survey offers a good level of representativeness of the academic community, because we find among them researchers, professors, PhDs, students, etc. As for their area of expertise, there is a slight majority of SSH researchers alongside with a certain number of STEM researchers.

The researchers were asked questions about their use of OA publications both as authors and as readers.

Regarding their published works, 74% of the participants declared they already published in OA, confirming a rather positive trend for OA publishing.

About those who didn't publish an OA article⁵, the chart below shows for which reasons they did not:



Reason for not publishing in OA (researchers)

The “other” category allowed for comments and, in this case, we find especially the financial issue related to APCs and BPCs. Nevertheless, the two main reasons relate, first, to the fragmentation of the publishing landscape, second, to the absence of open access policies. These two issues could precisely be addressed by the OPERAS infrastructure as an integrated service able to conduct OA advocacy and dedicate effort to communication, developing, for example a comprehensive and centralized list of open access publishers with indication on their conditions and quality. It could be a project that DOAB and DOAJ could achieve together.

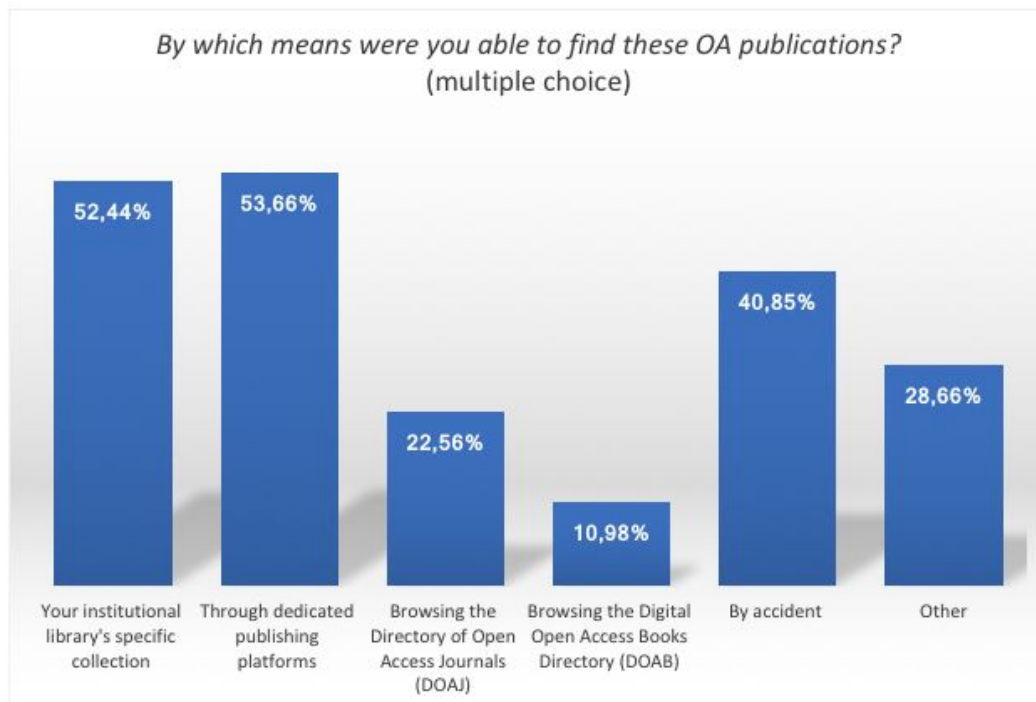
⁵ The same questions were asked about the articles and books OA publishing but the number of answers was logically more important and therefore more relevant for the articles.

For the question “If you published articles, could you indicate your level of satisfaction regarding the following aspects of OA publishing?”, the answers show a high level of satisfaction.



Level of satisfaction for OA publishing (researchers)

As readers of OA publications, 62,80% of the participants declare they can easily find OA publications. The chart below shows the answers to the question “By which means were you able to find these OA publications?”.



Finding OA publications (researchers)

The results show the importance of the local and/or personal network of the researchers, and especially the major role of their institutional documentation service for accessing to OA publications. On the other hand, in many cases the researchers access to OA publications without using dedicated tools or services: 40,85% find them by accident and 28,66% find them in another way, i.e. mainly Google Scholar, academia.edu and ResearchGate (according to the comments). In fact, the dedicated tools helping to specifically search for OA publications (DOAJ and DOAB) are the less used searching options.

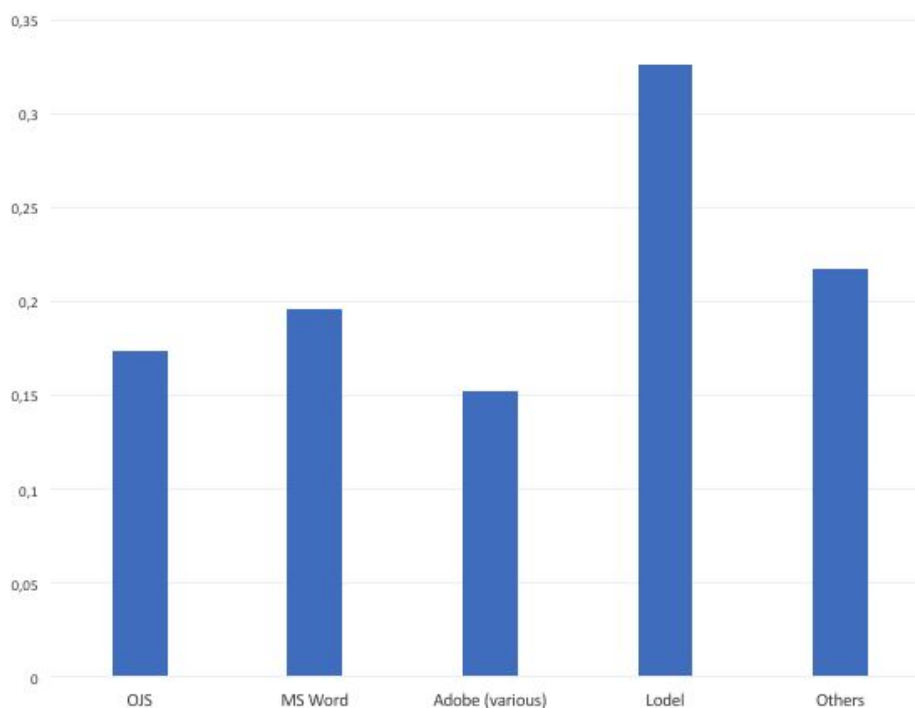
The open questions about the OA publishing in general confirm some well known concerns regarding, to summarize it, the APCs issue of the Gold OA and the impact issue of the Green OA. Some participants also mention their need for information about legal aspects and licenses: these comments legitimate the constitution of the DOAB as an integrated certification service.

Researchers also asked for what one of them called “*minimal guaranteed quality*” service, that is, mainly, providing a reliable IT system and a reasonable editing process duration. More advanced suggestions were pointing out the possibility of living-publishing or a redefinition of the search functionality based on the “*unit of knowledge*” and not only the article or the book. Those answers give us indications that search and discovery platform should develop advanced functionalities based on semantic annotation.

Survey for the publishers

The publishers who participated to the survey have collections of 170 books and 9 journals on average. Nevertheless, most of them are small or very small publishers, which is a good indication of the kind of stakeholders the OPERAS infrastructure will have to deal with. To be noted also that about a third of them have publications also in the STEM fields.

Apart from the transverse questions mentioned above, the specific questions for the publishers investigated the tools and workflows they were using. Regarding their publishing software, the publishers declared to be most of the time satisfied with it. The answers show however the tools used are rather disparate and perhaps a more precise evaluation would be useful. The chart below gives an overview of the software used.

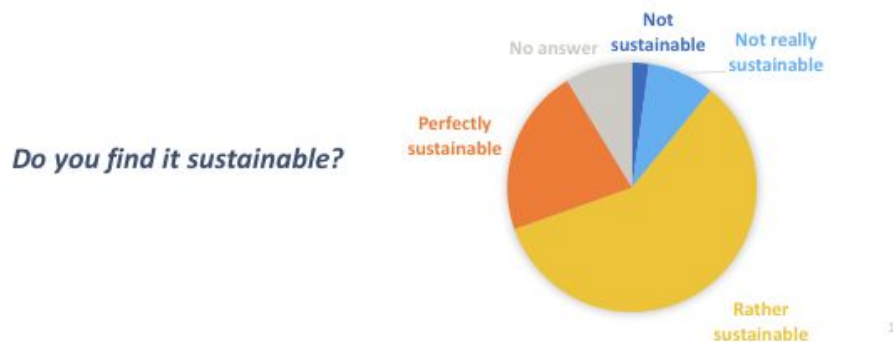
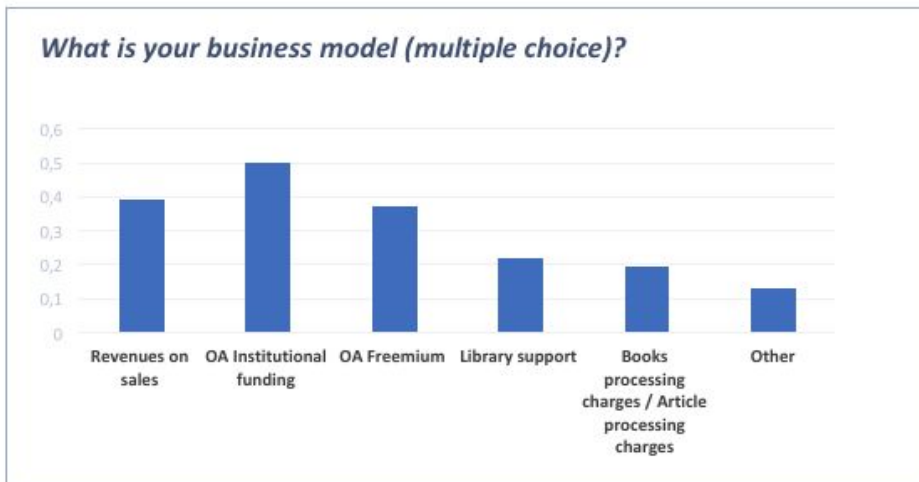


1

Publishing tools used by the publishers

Regarding the workflows, the input format is mainly DOC/PDF with some variations, and in the majority of cases the output format is PDF (85%), then come HTML (50%) and Epub (30%). Some of the publishers are working with a single source publishing process, which is almost always in XML-TEI. Among the publishers who are not yet using single source publishing, about a half would be interested to use it. This is a useful indication for the roadmap of the Tools R&D working group.

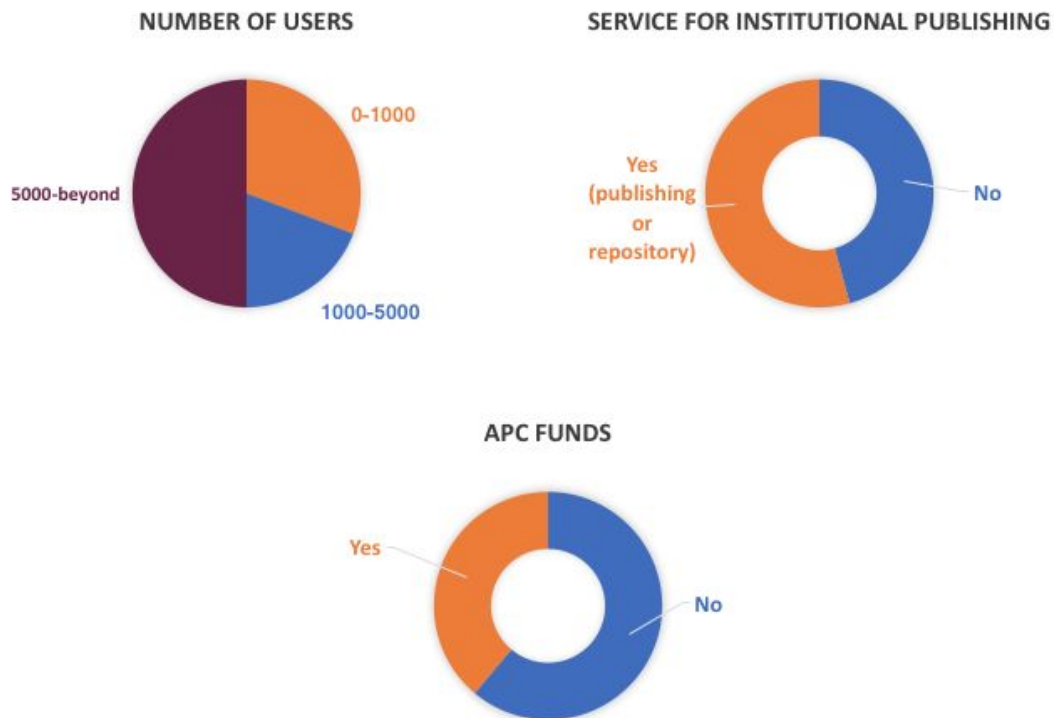
The question about the publishers' business model revealed, even with a slight majority of OA institutional funding, a rather high diversity of funding typology, especially taking into account the fact the participants could give multiple answers. However, the results also show the APC/BPC model is not the most important one. The publishers, for the most part, declared their business model was sustainable. Nevertheless, it would be interesting to make further investigation to put in relation the results of the first and of the second question.



Publishers' business models typology and evaluation

Survey for the libraries

The first questions in the survey for libraries gives us a set of information on their characteristics. Among the participants, a half have more than 5000 registered users; the other half is distributed almost equally between the libraries having from 1000 to 5000 users and those having less than 1000 users. These rather high numbers maybe explain the answer regarding the publishing service they provide and the APC funds they are keeping, as shown in the charts below.



Libraries organizational characteristics

The libraries who answered often have indeed dedicated human resources for open access publications and also a certain amount of specific funds. When there is no specific actions for open access (which is rare), it is mainly due to the difficulty to identify the right actors or to doubts on the peer-reviewing quality of those types of publications.

The libraries seem to have rather standardized techniques for finding OA contents (DOAJ/DOAB and OA platforms). And in fact, they express a high interest for an integration of DOAJ and DOAB. On the other hand, the OA business models and commercial offers (e.g. freemium model) don't seem well known. This indication validates one of the aims of the business models working group: to set up and develop a central platform for libraries where they could find all open access commercial offers in a transparent market place with low transaction costs.

The librarians gave a certain number of useful suggestions in the open questions. About the best practices, a participant suggested there should be generally *“more collaboration of publishers and libraries on metadata standards and metadata rationales”*. More precisely, another participant alerts on a topic of particular interest for the OPERAS infrastructure: *“Many Central and South European SSH journals still don't provide article-level metadata”*, validating the aim of OPERAS which is to integrate all players across the ERA and focusing on specific regions, such as southern and central Europe to bring them to a quality level in accordance with the international state of art

There are also more advanced suggestions regarding the identifiers (*“Robust linking of existing identifiers (such as ISNI, VIAF, GND) with ORCID”*) or the possibility to provide *“open peer review services”*.

If these aspect are already or will be addressed within the OPERAS consortium, some confirm the need of an infrastructure with integrated services like OPERAS: *“We need more content providers delivering rich content (rich metadata, information on peer review, licensing, terms of reuse etc.)”*; *“Make links to research data, research software and funding”*.

Survey for the socio-economic actors

The survey for the socio-economic actors has not received enough responses to be really significant. Nevertheless, it gives some information on the OA publishing reception and usage outside the academic community and possible hints for future surveys.

The participants are mainly from the information technology and administrative support area. They are all aware they can access freely to scientific content and a majority have already read an OA publication in the SSh field - with great satisfaction, it seems. However, they used generic search tools (e.g. Google Scholar) to find these OA publications and do not use dedicated tools or database. Finally, their reading present an interesting variety of motivations: they have read OA publications for their work, for their studies and out of personal interest in the same proportions.

Conclusion

The usage survey achieved during the first phase of OPERAS-D project already allows OPERAS consortium to validate a large part of the assumptions that were made during the preparation of the infrastructure project regarding the utility of the future services that would have to be deployed:

- The need for rich and multilingual metadata, SSP tools and an open access business model market place is recognised by most of the answers coming from researchers, libraries, publishers.
- The utility of the services developed by HIRMEOS (PIDs, rich indexes, annotation, alternative metrics) is also acknowledged.
- The importance of open access in researchers publishing strategies is obvious today, but a lack of information and transparency is also to be noted. That reveals the need for dedicated and integrated actions on this question and OPERAS infrastructure has to be highly instrumental in this regard.
- The need for the 3 platforms OPERAS wants to deploy appears clearly from the survey with very useful precisions coming from open answers.

Some limitations however exist: representativeness is not well balanced between the different ERA countries. More answers are needed, particularly from libraries. The

questionnaire and the dissemination strategy have to be completely reworked for socio-economic actors and funders targets. The survey will continue, be refined, and complemented by specific action during the year to come, and probably the followings. It must be understood as the first step of a continuous process enabling OPERAS infrastructure to collect feedback from its users community about the relevance of the services it offers.