



open access in the european research
area through scholarly communication

OPERAS Design Study



July 2017

introduction

OPERAS Website: <http://operas-eu.org>

Full Design Study with annexes: <http://operas-eu.org/design-study>

Contact: pierre.mounier@openedition.org

FOREWORD

The following report, OPERAS Design Study, has been composed thanks to the OPERAS-D project, in its first stage. The report joins four studies that explore the landscape of OPERAS' field of activity, establish the technical mapping of the OPERAS Consortium, survey users' needs regarding scientific communication and academic publishing, and finally look ahead to the development of the governance structure and business model of the future infrastructure within the ESFRI framework.

One word can synthesize what these studies and reports tell us about OPERAS' field of work: fragmentation. Hence, the OPERAS vision and mission, which arises naturally from this finding: integration.

THE STRUCTURE OF RESEARCH IN SOCIAL SCIENCES AND HUMANITIES

Fragmentation, as evidenced by the Landscape Study in this report, is one of the major characteristics of the communication and publication sector in humanities and social sciences: an impressive number of small size players of different types operate across the European Research Area to offer communication and publication services to researchers. The reasons behind this situation can be analysed at multiple levels, but those reasons shouldn't overshadow the most important one: the scientific community they serve is itself highly fragmented.

Many studies and authors have explored and revealed the fragmentation of the research landscape in the humanities and social sciences (see the Scientific Case), across multiple disciplines and sub-disciplines, as well as in small research units, and, of course different languages. To take a striking example, the recent INTERCO-SSH project, that studies internationalization of SSH found that, despite the growing importance of English as a communication language in social sciences and even humanities, the need for academic publications in native languages remains central in many cases. According to the same project findings, most of the time internationalization of SSH doesn't mean going from local to global, rather it goes through what is described as transregional integration which reflects the structure of scientific networks that connect researchers across national boundaries and not always globally.

Therefore, the particular structure of the publishing sector in social sciences and humanities, composed of small and many players serving local scientific communities and specializing in narrow fields of research, cannot be considered as a flaw, but in fact more as a feature, an adaptation to the reality of the scientific ecosystem in these disciplines. The flaw emerges, however, when the actors playing in the field operate as isolated beings, unaware of what others are doing, reinventing the wheel in their own corner or even competing through unfair practices. In such circumstances, fragmentation turns into disintegration and the scientific community which needs efficient partners to circulate ideas and research findings across borders could be negatively impacted. The flaw is also apparent when the whole system experiences a global change in which all players need to redefine and renew their mission, the services they deliver, and consequently their workflow, business model and tools. The global change experienced by everyone is well known: the digital turn and its consequence: Open Science.

OPEN SCIENCE IN PRACTICE

Open Science has so far mainly been debated as a principle. And as a principle, a growing part of research funders – funding agencies, ministries and the European Commission – seem inclined to adopt it. In 2016, the Dutch government took the opportunity of its European presidency to push for a new stage in the development of Open Science, from principle to reality, resulting in the The Amsterdam call for action on Open Science. But, as mentioned in the document, to put Open Science into practice requires strong coordination between the different stakeholders of the scientific community: researchers, funders and national authorities, libraries and finally...publishers. That's why the implementation of Open Science in humanities and social sciences is a specific challenge for a sector that is currently loosely connected and highly fragmented.

As it appears from the Technical Mapping study undertaken during the OPERAS-D project, the structuration of OPERAS partners in terms of technologies used, types of services offered to the community but also and may be even more importantly, the business models and workflows, is heterogeneous. Therefore, to increase integration in terms of interoperability and complementarity through cooperation across different institutions and European countries with their particular academic cultures, requires a particular effort which is, in many cases, beyond their reach on their own as resources are limited for each of them. Yet, the same study reveals that divergence is not total, particularly at a technical level, with most partners having chosen more or less the same technological bases, adopting more or less the same standards, and aiming more or less at the same practices. The whole challenge resides in this 'more or less' that reflects on the one hand the need for flexibility that fits local contexts, but on the other hand presents serious obstacles to practical integration. For decades, players in the scholarly communication field haven't felt a particular motivation to become more integrated.

Cooperation between scholarly communication players locally or nationally, often remains loose, and limited to exchange of information during scientific and professional conferences (such as Elpub, APE, ALPSP), and exchange of rights during book fairs (such as Frankfurt, London, Turin). In the print era and at the very beginning of the digital turn, networking was enough. But the recent development of Open Science changes everything, and at a fast pace, especially when it comes to putting into practice the goal of the European Commission to set up a European Open Science Cloud within a few years. For this purpose, networking is not enough anymore and integration must be achieved in reality for disciplines that would otherwise risk being left behind. Indeed, if SSH disciplines do not integrate with the Open Science Cloud, the entire scientific ecosystem will fail to reach its full potential since it will be missing the publications and other research outputs from half of the scientific disciplines.

THE 'LONG TAIL SCIENCE' MODEL

How can many small collections of materials provided by many small teams integrate in practice? OPERAS, as a Research Infrastructure project, aims to provide a sensible and practical answer to this question. The intellectual model that governs the way the OPERAS project is structured, and the main choices that have been made to plan its development, has been aptly summarized by a European Commission officer during an informal conversation about EOSC and how OPERAS could contribute to it: The main challenge OPERAS wants to tackle is the integration of 'long tail science' into the Open Science framework. This long tail model operates at two levels: the definition of services the future infrastructure will offer, and the type of structuration it will adopt for its operation.

The plan adopted by OPERAS for its services and their structuration is the result of numerous studies undertaken by OPERAS partners (see the Bibliography in this report), individually or collectively, and of seminars, workshops and conferences attended together where a continuous conversation was fed and progressively structured. For the record, the original idea of setting up an infrastructure for open scholarly communication in SSH came from a workshop organized by Victoria Tsoukala (EKT), Emmanuelle Corne (AEUP), Pierre Mounier (OpenEdition), Eelco Ferwerda (OAPEN) and Brian Hole (Ubiquity Press) during the Elpub conference in 2014. From that starting point and those five original participants, representing a variety of situations and experiences, to today with more than 30 partners from 11 countries participating in the project, the important work of designing the infrastructure has taken place.

DESIGNING THE SERVICES

The design of future OPERAS services has culminated in an online survey to test OPERAS' proposition against users' needs, in particular those of researchers, libraries and publishers, the most important stakeholders for OPERAS. The main principle that has emerged from this Design Phase is the need to define future OPERAS services at several levels and distribute them following a principle of subsidiarity.

FIRST LEVEL: SHARED SERVICES

At a first level, the partners offer communication and publication services to their community, whether it is regional, national or limited to a specific language. At this level, what OPERAS proposes is not to merge the existing services into a pan-European one, but on the contrary to support the partners to improve and upgrade the services they already offer to their own users. During the Design Phase it was identified that specific support is needed in the following three key areas:

- The definition and adoption of best practices that allows for a common level of quality and compliance with Open Science principles
- Research and development activities aimed at developing publishing tools and technologies that partners can use from a shared toolbox in their adoption of common best practices and to support the redefinition of their workflows
- Support for innovative open access business models by developing shared components such as a common market place, a journal flipping mechanism and a funding model that involves libraries in supporting open access.

As such, OPERAS services have been designed as 'Shared Services' between partners, aimed at supporting and improving their existing activity, not replacing it.

SECOND LEVEL: EOSC INTEGRATION

A second level of more integrated services then had to be defined, to prepare content to be accessed and used through the EOSC. This level of services was more difficult to design because EOSC is in its first stage of development and has not yet been put into practice. It was decided that the best way to prepare for future integration is to upgrade existing dissemination platforms in the OPERAS Consortium with rich metadata and machine-readable content allowing for efficient text and data mining from third parties. We started with a specific project within the H2020 framework programme,

focusing on open access books platforms which required specific development, as books are the most difficult objects to integrate considering their specificities. The HIRMEOS project allows for the implementation of standard identifiers such as DOI, ORCID and Fundref for books, but also other more innovative types of metadata, such as reader annotation and new usage metrics.

More importantly, HIRMEOS was used to test and deploy a common methodology that enables different partners operating platforms based on different software and technologies to implement common standards. Based on a uniform definition of implementation levels, and a governance framework that commands distribution of work among partners, the HIRMEOS method will be used in the future development phase of OPERAS to extend standards implementation beyond the project, beyond the five dissemination platforms participating in it, and of course beyond the books themselves.

Considering the specificities of SSH content and the importance of its distribution across several languages rather than one, it was clear to OPERAS partners that special attention should be paid to multilingualism to facilitate the process of integration into the EOSC. In the development of the infrastructure, it is therefore planned to undertake specific work, first on the alignment of metadata describing content with ontologies in several languages, and second to support metadata translation that improves content discoverability.

THIRD LEVEL: OPERAS PLATFORMS

Finally, the most important services to be delivered at European level, which are meant to address all stakeholders' needs across the European Research Area and across different languages:

- Research funders and libraries need a certification service to implement their open access policies for the former and to deliver good quality content to their users for the latter. This service has to be delivered globally because certification needs to be independent from local constraints and free from local interests; in all cases, certification must come from external authorities.
- Researchers need an open and efficient Discovery platform to find content relevant to their research topics. Since SSH researchers read if not write in several languages, the platform should be able to support multilingual content, which is a sufficient reason to set it up globally, and index different types of content: publications of course, but also primary data and other grey literature content. The Discovery platform will also serve as the main interface with the EOSC.
- Society and different types of socio-economic actors (media, citizen, administrations and SMEs) need more than just access to academic content. In the context of citizen science which is implied by the definition of Open Science, they need a common framework to collaborate with research teams to achieve research projects that tackle their specific concerns, namely societal challenges. Therefore, OPERAS will prepare and deploy a Research for Society platform that addresses those needs that will be open to be used across all disciplines, including both SSH and STM, in a multidisciplinary perspective.

PLANNING OPERAS DEVELOPMENT

The promise to deliver three pan-European platforms by the OPERAS project could be considered too ambitious, potentially exceeding the Consortium's resources and capability. This might be true if the aim of OPERAS was to build those platforms from scratch and to develop them at the same time.

However, the method adopted during the Design Phase was to identify existing platforms provided by OPERAS partners and to upgrade them with the new functions they will need to deliver and scale up to the European level. The maturity of the three chosen platforms is different, which will allow for smooth development phasing across the Design, Preparation and Construction stages of the infrastructure:

- The certification platform is the first to be developed. Based on the existing Directory of Open Access Books (DOAB: <http://doabooks.org>), operated by OAPEN and OpenEdition, its main development is currently supported through the HIRMEOS project to upgrade the structuration of the platform in terms of workflow organization and technical capability. The development of DOAB as OPERAS' certification platform is currently supported across the Design and early Preparation Phase of OPERAS within the HIRMEOS project.
- The Discovery platform is based on the very mature Isidore platform (<http://rechercheisidore.fr>) developed by the French Research Infrastructure Huma-Num. Launched in 2011, Isidore has proved its sustainability at least at national level, and, with more than a million visits a year, it is clearly meeting researchers' needs. From a technological perspective, Isidore meets much more than the minimal criteria to be compliant with a state-of-art digital delivery platform. Following the principles of the web of data, Isidore enriches indexed content with metadata aligned across several scientific vocabularies and provides access to data through several means, including a Sparql Endpoint. The planned development of the Discovery platform will extend Isidore to a wide array of different languages from French, English and Spanish which are currently supported, and will be delivered during the Preparation Phase.
- The Research for Society platform is the less mature of the three platforms. Its main development will take place during the Construction Phase after prototyping during the Preparation phase. Its envisaged starting point will be the Hypotheses platform which, as an academic blogging platform is completely mature. With more than 2,000 active blogs, and a structured user community in several of the most important European languages (French, English, German, Spanish, Portuguese, Italian), the Hypotheses platform, operated by OpenEdition, has already reached a pan-European scope. Nonetheless, many components are still to be developed to set up a real usable environment to support collaboration across academic boundaries.

STRUCTURE THE INFRASTRUCTURE

SET UP THE GOVERNANCE SCHEME

Having defined the services OPERAS will provide and planned the timeline along which they will be implemented, the last component that had to be designed was the structure of the infrastructure supporting the services, including its governance scheme and future business model. The plans resulted from a specific study achieved during the OPERAS-D project and included in the report. Here again, the structure adopted reflects the particular landscape of the SSH community. For such a community, a centralized infrastructure wouldn't address the complexity of SSH activity, but given its fragmentation, the risk that the infrastructure could lose its direction was taken into consideration. In this matter, the experience of other European Research Infrastructures, close to OPERAS in their scope and the users they serve, was enlightening and helped us to ensure a delicate balance between centralization and federation. The principles that have been adopted to prevent loss of guidance and lack of integration are, first, to rely on a strong hub for coordination, supported by one institution

and the Coordinator, and, second, to create a legal entity early in the course of the development, to prepare final incorporation into an ERIC.

The choice of OpenEdition as Coordinator and host of the hub, results from its position in the Consortium, its size and, of course the commitment coming from its supporting institutions and national authorities for the development of OPERAS. Placed in Aix-Marseille University, the management team will be strongly embedded in the OpenEdition team (50–60 persons), will benefit from OpenEdition’s supporting institutions (Aix-Marseille University, CNRS, EHESS, Avignon University) and will be located in the Technopole of Chateau-Gombert, an institution that incubates a large number of high technology companies and scientific laboratories. OpenEdition’s capacity to coordinate such a project was assessed through an evaluation exercise performed in June 2017 by an external audit company (the executive summary of the report is included in this Design Study).

Lessons from other infrastructures were also learnt, showing that to prepare and set up a pan-European entity is a long and difficult process. On the other hand, infrastructures which are not incorporated as a legal entity and perpetuate only through projects can lose consistency and drift in different directions following the divergent opportunities defined by the projects they run through. Therefore, OPERAS aims at preparing for ERIC incorporation through an intermediary stage, the creation of an international association that on the one hand will reflect the current organization of the project, and on the other will prepare for the organization of the future ERIC, and ensure a progressive transition towards implementation and start operating the functions of the infrastructure: project management through the hub, independent scientific monitoring, political representation of the Member States and executive participation of the partners.

COMMUNITY MANAGEMENT

To be effective and change the landscape of scholarly communication in Europe, OPERAS must be able to gather a high number of partners, and manage and coordinate them. For this reason a light commitment scheme was defined, allowing small size partners, with few resources, to participate in the project through thematic Working Groups that align with the structuration of the services OPERAS will deliver. The Working Groups will be used to prepare the future H2020 projects that will support the development of the infrastructure.

With a large community composed of many partners, a more complex structuration is needed. A Core Group was created during the Design Phase, gathering the partners willing to commit more than the others, to organize their national community and manage the Working Groups. In the future, the Core Group will transform into an Executive Assembly and gather National contact points as well as other representatives. The management office installed in the hub will support the work of the different groups and ensure effective coordination between partners at different levels.

OPERAS also coordinates with other ESFRIs such as DARIAH, CLARIN, CESSDA, and e-infrastructures such as OpenAIRE, as well as other projects that complement OPERAS’ core activity such as ENRESSH, and international partners such as Scielo.

BUSINESS MODEL

Finally, we had to define a business plan to finance the development of the infrastructure, that reflects its structuration. The funding of the hub and the project management team is ensured by the Coordinator, OpenEdition is supported by French authorities, and the development of the services will

be funded through projects. The participation of the partners in Working Groups and Core Groups is self-sustained through in-kind contribution.

When the infrastructure is in operation after the creation of the ERIC, another business plan will be adopted, relying on Member States' annual contribution to fund the hub and project funding to develop new services. OPERAS platforms will be operated by identified partners who will support the operational costs of the platforms. They will be funded through a mix of upfront funding and commercialization of premium services.

CONCLUSION

OPERAS' name epitomizes in many ways the mission that guides its development plan. First and foremost, OPERAS stands for 'open access in the ERA through scholarly communication'. This name is not only an astute way to align keywords in a single sentence; it means something more. It means that for the partners of this common effort, open access to publication shouldn't be defined outside and independently from the scientific community. Since the Second World War, scientific publication has been progressively outsourced to commercial entities by scientific institutions, which seemed a good solution in the first place to improve quality by professionalization. But control was progressively lost by the scientific community over a strategic part of its activity, the part that conditions its very existence as a community: communication. Some forward-thinking scholars and librarians in the 1980s started to become alarmed by the situation and considered the path that had been taken during the previous decades to be a tragic mistake. Different initiatives were taken from there, scarce and small at the beginning, to allow the scientific community to take back control over its own communication system. This movement took different forms, from the creation of new university presses to the commitment of research libraries to content dissemination from their institution, and all these took another dimension when Internet became the standard. Then the open access movement started and led to a new and more complete concept, Open Science. But all this evolution shouldn't overshadow its origins and its original meaning: the need for the scientific community to reclaim its own communication system.

That is why we, the OPERAS partners, consider that the best way to achieve open access movement in Europe, is to do it *through* scholarly communication, which simply means *from within* the scientific community, by close cooperation between its different stakeholders and always considering primarily its specific needs over all other considerations, in particular commercial. That is why when we tried to define our initiative, to put a name on what we were aiming at, it came to us very naturally that we were on the course of creating an infrastructure, but more importantly, a *Research* Infrastructure that should stand by the researchers and operate *inside* the scientific community to support an essential part of its activity: scholarly communication.

OPERAS is also a metaphor, of course. Opera is one of the most sophisticated and complex performance arts, because it involves so many different components, symphonic music, lyrical art, drama, and even visual art through scenery and costumes. To perform it correctly, it requires thorough understanding between all the different performers, strong coordination and close cooperation. And then, after a long preparation, when the time for the performance has arrived, the complex machinery must become invisible and serve the artwork smoothly and gently as if it was all natural. An inspiring model, certainly.

Pierre Mounier
OPERAS Coordinator