

Management of open access research infrastructures in large EU projects: The “CultureLabs” case

SPINELLO, Andrea Orazio, GIGLITTO, Danilo and LOCKLEY, Eleanor
<<http://orcid.org/0000-0001-8257-9543>>

Available from Sheffield Hallam University Research Archive (SHURA) at:
<http://shura.shu.ac.uk/29453/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

SPINELLO, Andrea Orazio, GIGLITTO, Danilo and LOCKLEY, Eleanor (2021).
Management of open access research infrastructures in large EU projects: The
“CultureLabs” case. CNR-IRCrES Working Paper, 9/2021.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

CNR-IRCrES Working Paper

**Management of open access
research infrastructures in
large EU projects:
the “CultureLabs” case**



9/2021

**Andrea Orazio Spinello
Danilo Giglitto
Eleanor Lockley**

Direttore Emanuela Reale

Direzione CNR-IRCrES
Istituto di Ricerca sulla Crescita Economica Sostenibile
Via Real Collegio 30, 10024 Moncalieri (Torino), Italy
Tel. +39 011 6824911 / Fax +39 011 6824966
segreteria@ircres.cnr.it
www.ircres.cnr.it

Sede di Roma Via dei Taurini 19, 00185 Roma, Italy
Tel. +39 06 49937809 / Fax +39 06 49937808

Sede di Milano Via Corti 12, 20121 Milano, Italy
Tel. +39 02 23699501 / Fax +39 02 23699530

Sede di Genova Corso Ferdinando Maria Perrone 24, 16152-Genova, Italy
Tel. +39 010 6598798

Comitato Redazione

Emanuela Reale, Giuseppe Giulio Calabrese, Grazia Biorci, Igor Benati, Antonella Emina, Serena Fabrizio, Lucio Morettini, Susanna Paleari, Anna Perin, Secondo Rolfo, Isabella Maria Zoppi.



redazione@ircres.cnr.it



www.ircres.cnr.it/index.php/it/produzione-scientifica/pubblicazioni

The Working Papers published by CNR-IRCrES represent the views of the respective author(s) and not of the Institute as a whole.

CNR-IRCrES Working Paper 9/2021



dicembre 2021 by CNR-IRCrES

Management of open access research infrastructures in large EU projects: the “CultureLabs” case

ANDREA ORAZIO SPINELLO^a, DANILO GIGLITTO^b, ELEANOR LOCKLEY^b

^a CNR-IRCrES, Consiglio Nazionale delle Ricerche, Istituto di Ricerca sulla Crescita Economica Sostenibile, via dei Taurini 19, Roma – Italia

^b Sheffield Hallam University, Howard St, S1 1WB, Sheffield – United Kingdom

corresponding author: andrea.spinello@ircres.cnr.it

ABSTRACT

Research funding organizations, particularly at international level, are increasingly promoting the creation and maintenance of open access research infrastructures (RI). These resources have assumed a pivotal role as support for the new open and networked science in their dimension of technical and operational frameworks that allow scientists and stakeholders to collaborate and share scientific data and results. In Social Sciences and Humanities (SSH), the creation and exploitation of open data platforms is still attempting to catch up with longer-standing practices in the “hard sciences” as the resistance to wider data sharing has not yet been completely overcome. This paper aims to describe how a large project, financed by the European Commission, managed the creation of a RI in the field of SSH, showing the steps undertaken to comply with the GDPR regulations and prepare the data for useful sharing and reuse. In this regard, the authors present the case study of the Horizon 2020 “CultureLabs” project, placing emphasis on some specific practical factors that they believe are particularly important for implementing open access principles in the establishment and maintenance of RIs in the new course of science based on sharing and openness. In particular, the authors will focus on creation of “useful” and GDPR-compliant data and the impact on research activities as a result of their (re)utilisation; the control of the data management process; and the compliance with funders’ requirements (e.g. in terms of data security). The reflection on the interplay of these aspects, operated through a case study, appears to be crucial in moving away from a merely theoretical approach to addressing the issue of open access, and it hopes to serve as a guide or a warning for those who create and administer open RIs.

KEYWORDS: open access, GDPR, data management, research infrastructures, SSH.

DOI: 10.23760/2421-7158.2021.009

HOW TO CITE THIS ARTICLE

Spinello, A. O., Giglitto, D., Lockley, E. (2021). *Management of open access research infrastructures in large EU projects: the “CultureLabs” case* (CNR-IRCrES Working Paper 9/2021). Istituto di Ricerca sulla Crescita Economica Sostenibile. Disponibile da <http://dx.doi.org/10.23760/2421-7158.2021.009>

CONTENTS

1.	INTRODUCTION	3
2.	THE NEW PATH OF SCIENCE BASED ON OPEN AND SHARED RESEARCH INFRASTRUCTURES	4
2.1.	Research infrastructures in SSH and resistances to data sharing	5
3.	CRITICAL ASPECTS RELATED TO OPEN DATA	6
4.	THE CULTURELABS CASE: PURPOSE, CREATION, MANAGEMENT	7
4.1.	The CultureLabs project	7
4.2.	The CultureLabs infrastructure: creating <i>useful</i> open access data and <i>impact</i> on research design	9
4.3.	Responsibility and control: the actors in the research data management process	10
4.4.	Practicalities: meeting the funders' requirements	11
5.	CONCLUSIONS AND DIRECTIONS FOR FUTURE STUDIES	12
6.	ACKNOWLEDGMENTS	13
7.	REFERENCES	13

1. INTRODUCTION

Research funding organizations, particularly at international level, are increasingly promoting the creation and maintenance of *open access* research infrastructures in all research fields (cf. ESFRI, 2018; European Commission, 2020b). Referring to the domain of Social Sciences and Humanities (SSH), research infrastructures can serve as long-term institutions anchored in public policies and non-competitive resources accessible to all, addressing not only the needs of scientific communities but also a diverse range of stakeholders (Renschler et al., 2013; Duşa et al., 2014). On the one hand, the shared exploitation of these resources is fostering a new mode of knowledge production characterized by the promotion of network-building processes and a broadening of research impact, nurturing the idea of ‘networked collaborative science’ (Nielsen, 2011). On the other hand, their maintenance and opening can determine a profound effect on the way in which research in the SSH is organized, offering two visible advantages – cost efficiency (since the research data is produced once and reused) and the promotion of a balanced competition. All whilst proposing new challenges, from the tension between data access and confidentiality, to overcoming fragmentation and financing issues (Ciula et al., 2011).

Despite the trend toward the establishment of open data platforms, particularly in the SSH domain, a significant proportion of knowledge remains inaccessible, and the sharing of empirical data appears to be excessively fragmented (Dutton & Meyer, 2009; Jarolimkova & Drobikova, 2019; Spinello, 2019). Scientific data is more likely to be opened and shared when career benefits are perceived as connected with these practises and above all when regulations from funding institutions encourage researchers at the cost of depriving them of research funding (Kim & Adler, 2015; Schmidt et al., 2016; Chawinga & Zinn, 2019).

Under the last European Framework Programmes and Horizon 2020, the European Commission set itself as a funding body for constructing, extending, opening and/or updating research infrastructures for all scientific disciplines, incentivising and regulating open access practises, and making them essential prerequisites for funding (European Commission, 2020b). Creating open data research infrastructures means developing a purposeful attempt to make all raw research data fully available for public access following the principles of transparency and reproducibility. This process of opening data raises several issues, ranging from privacy and security requirements to the generation of “useful” data, the meeting of funders’ requests, and the impact on research activities as a result of their exploitation.

This paper aims to describe how a large project, financed by the European Commission, managed the creation of a research infrastructure in the field of SSH, showing the steps undertaken to comply with the GDPR regulations and prepare the data for useful sharing and reuse. In this regard, the authors present the case study of the Horizon 2020 “CultureLabs” project by reflecting on the processes involved to ensure open access data requirements in the framework of a new mode of knowledge production based on data sharing. In particular, the authors will focus on the provision of “useful” (and compliant with GDPR) data to be shared and the impact on research activities as a result of their reuse, the control of the data management process, and the practicalities related to adhering to the requirements imposed by the funders.

This contribution is organized as follows. Section 2 briefly discusses the key terms characterizing the new knowledge production process – openness and sharing – and introduce the concept the research infrastructures as ideal basis for the implementation of these practises; the subparagraph 2.1 reports the peculiarities of research infrastructures specifically in the SSH domain; Section 3 will present some critical aspects related to data openness and introduces the related dimensions of interest for this contribution; Section 4 and the respective sub-paragraphs will present the CultureLabs project case, presenting the infrastructure and the selected aspects related to the open data management process; the last section is dedicated to the perspectives for management of the open access data hosted in research infrastructures.

2. THE NEW PATH OF SCIENCE BASED ON OPEN AND SHARED RESEARCH INFRASTRUCTURES

Over the past two decades, the research world has revised processes and practices related to the production and sharing of new knowledge across all disciplines (cf. Gibbons et al., 1994; Martin, 2003; Borgman, 2007).

Some authors have outlined the contours of a “new path for science”, an unprecedented paradigm based on the *sharing* of information wealth and the intensity of data (the so-called “data deluge”) with a greater potential for the resolution of scientific issues (Hey et al., 2009). The abandonment of a static view of the scientific process, characterized by rigidity of roles and disciplines, has given way to the emergence of *distributed and collaborative knowledge networks* based on the interdependence of researchers’ skills (Nielsen, 2011).

In parallel, the progressive institutionalization of research data as a common and shared good has given relevance to the concept of *openness* in terms of data accessibility oriented towards ever wider audiences of researchers and society (Boulton et al., 2012). The production of knowledge has found more and more space outside academic institutions, which have been required to collaborate more intensively with civil society through sharing (or transferring) knowledge, skills and results achieved, for maximizing the benefits obtainable from scientific research. There are many actors involved in this collaborative process, and each has their own specific needs: (i) the researchers with their ideas and results (constantly looking for new funding opportunities); (ii) universities and research institutions called upon to adapt to new global, European or national policies (e.g. the evaluation of the social impact of research); (iii) the private sector that needs to overcome the economic recession; (iv) governments that promote the evaluation of their investments in order to justify the expenditure of public resources.

In this context, *research infrastructures* have assumed a pivotal role as support for the new open and networked science in their dimension of technical and operational frameworks that allow scientists and stakeholders to meet, collaborate, and share scientific data and results. Research infrastructures refer to “facilities, resources and related services that are used by the scientific community to conduct top-level research in their respective fields”. This “institutional” and very generic definition (as noted by Ciula et al, 2013) covers: “major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives or structured scientific information; enabling ICT-based infrastructures such as Grid, computing, software and communications; any other entity of a unique nature essential to achieve excellence in research” (European Commission, 2008, p.10). In the domain of SSH, the reference is focused on data archives, repositories, and collections; national and international statistical facilities and research data service facilities; registers and survey-led studies/databases (MERIL project, 2011).

Recognizing the strategic importance of research infrastructures, the European Commission has qualified them as research resources to facilitate interdisciplinary and institutional collaboration as well as the enabling of public-private partnerships (cf. European Commission, 2020b), investing in their creation and maintenance within its Framework Programmes and Horizon 2020. Over recent years, many infrastructures at the European or national level that have arisen within time-limited research projects have developed into more stable institutions, increasingly incorporating user groups through methodological innovations and the incorporation of new research topics (an example is FP7-RISIS project, see RISIS project, 2013).

With a view to the structuring process of the European Research Area, the European Strategy Forum on Research Infrastructures (ESFRI) is responsible for identifying the most relevant research infrastructures on the basis of scientific excellence and competitiveness, facilitating multilateral initiatives that allow their development and better use (cf. ESFRI, 2018). Integration, interoperability, and – above all – management of openness to the scientific community and relevant stakeholders are among the basic attributes of the new research infrastructures of European relevance. This ESFRI initiative of mapping, supporting, and promoting research infrastructures is part of a larger effort to encourage a joint approach to data sharing across Europe that is driven by research policies and compliance with GDPR regulations.

2.1. Research infrastructures in SSH and resistances to data sharing

As reported above, research infrastructures in the domain of SSH refer to categories of academic material collections, data repositories/archives/platforms, and structured research databases. A further classification may be relevant when it comes to platforms or databases that allow differentiating between *project-driven platform/databases*, derived from specific projects limited in time and scope, and *infrastructures of European interest or of international standing* (cf. German Council of Science and Humanities, 2011).

As for the latter category, there is a growing presence of institutionalized data collections at the European level, characterized by wide openness to the research community, updating and interoperability. ESFRI identifies research infrastructures of pan-European interest distinguishing between landmarks (e.g., CESSDA – Consortium of European Social Science Data Archives; CLARIN – Common Language Resources and Technology Infrastructure; DARIAH – Digital Research Infrastructure for the Arts and Humanities) and projects (infrastructures not yet mature and developed but of scientific excellence). The ESFRI Roadmap emphasizes their strategic importance for European research development and advocates for their timely support and implementation (ESFRI, 2018).

Nevertheless, SSH are characterized by the presence of a multitude of smaller infrastructures, developed in relative independence within specific research projects (German Council of Science and Humanities, 2011; Spinello, 2019). In this domain, it is very common to create data collections which are focused and tailored to answer punctual research questions and intended for the use of primary creators (Dutton and Meyer, 2009). Usually, when these resources are no longer able to meet the expectations or when they lose their connection to the development of research methods, their exploitation is terminated. The external availability and the relative openness of the data from these kinds of infrastructures depend on specific agreements and conditions stipulated with the sources of the data itself or with the funders of research (cf. OECD, 2013; Kim & Adler, 2015).

In general, the mode of usage and the mutual exploitation in the SSH scientific community of research infrastructures is attempting to catch up with longer-standing practices in the so-called “hard sciences” as the resistance to wider data sharing has not yet been completely overcome. The idea that data openness is beneficial to science does not appear to be deeply embedded in the culture of some researchers; on the contrary, the desire of scientists to maintain a competitive advantage over their peers persists. Borgman (2012) discusses a true conundrum which is inherent in the issue of balancing the benefits and drawbacks of opening scientific activity: researchers fear losing competitive power by sharing their findings as they fight for funding, career, and publications. Jarolimkova & Drobikova (2019) found that social scientists are less willing to share their data and are more concerned about potential ethical and legal problems. In the Italian context, Spinello (2019) noted that the community of social scientists is still grappling with the early stages of the transition process towards a way of producing knowledge based on open and shared research infrastructures, characterized on one side by a relative lack of opening practises with reference to the project-driven ones and on the other by a reduced exploitation of the open data from the European-level ones.

Although an infrastructure can be established as part of a specific research project, with relatively few users, it may potentially be expanded into a resource open to a wider community. In this regard, what is hoped by many observers is that the data contained in project-driven infrastructures for SSH can be increasingly open and shared, allowing the social sciences to abandon the model based on isolated databases and to embrace one focused on infrastructures that contribute to scientific provision, offering the same starting conditions to every member of a scientific community in terms of access to data and information, without restrictions (cf. Duşa et al., 2014).

3. CRITICAL ASPECTS RELATED TO OPEN DATA

Given the persistence of resistances, the processes of data opening still needs a promotion through mechanisms such as the obligation imposed by national or international funding agencies (cf. Schmidt et al., 2016; Chawinga & Zinn, 2019). Nevertheless, these obligations are transforming into an increased awareness within the research communities about the importance of the practices of opening and sharing data. The publication of the Royal Society's report "Science as an open enterprise" (Boulton et al., 2012) signalled a significant paradigm shift in the dynamics of scientific knowledge production, highlighting a general trend toward a greater openness among scientists, in terms of incentives for collaboration and their relationship with the wider society, the growing recognition of the value of open scientific data collection and their visibility, and the use of common standards for sharing information in order to make it more widely usable. One of the report's primary points can be summarized with "open access to open data", and research infrastructures are becoming the appropriate means for dealing with the requirements of this new course of science.

According to Paschetto et al. (2015), the debate over open data and how to access it is characterized by eight recurring dimensions of interest: users; legal aspects; dissemination methods; access standards; management and conservation; sharing scale; material references; and related benefits. Sharing research data for the use of others (namely, individuals who did not participate in their production) implies the identification of the target audience. The extension of the range of sharing varies depending on the "reuse" perspective of data assumed by data producers. The conditions of "reuse" are foreseen in the strategy of administration, preservation, and care, which are stated in the preservation policy documents. From a legal perspective, open data producers can allow external access while protecting their intellectual property and limiting operations on the data in a variety of ways. The most commonly adopted solution is the use of an Open Data Commons license, which demands that data is reused provided that the source, the privacy configuration, and the ethical constraints are cited. These are specified in appropriate GDPR-compliant documentation that describe the collection methods, sources, and purposes for which they were obtained. Technical data access is intrinsically linked to data dissemination modalities, which can range from online publications on personal or institutional websites, to arrangements in archives or repositories, to the creation of more complex relational structures in the form of public databases. Direct contact with data is generally obtained through digitally-mediated solutions such as Digital Object Identifiers (DOI), which are based on the World Wide Web standard and Linked Open Data and allow for the generation of "executable sheets" to extract data and perform analysis. Although most open databases have access requirements that are more or less identical, there is no coordinating mechanism for international data exchange currently in place, and research data sharing procedures vary by format, management style, and disciplinary field.

In this contribution, the authors focus on some specific practical factors for implementing open access principles in the establishment and maintenance of research infrastructures in the new course of science based on sharing and openness. They can be summarised as follows: (i) creation of "useful" and GDPR-compliant data and the impact on research activities as a result of their (re)utilisation; (ii) the control of the data management process; (iii) the compliance with the funders' requirements (e.g. in terms of data security). Based on the Horizon 2020 CultureLabs project's experience (two out the three authors were involved in the project), in the following Section the authors will describe how a large European Commission-funded project implemented these critical points. The reflection on their interplay, operated through a case study, appears to be crucial in moving away from a merely theoretical approach to addressing the issue of open access, and it hopes to serve as a guide or a warning for those who create and administer open research infrastructures.

4. THE CULTURELABS CASE: PURPOSE, CREATION, MANAGEMENT

This Section explores the practices and the steps undertaken within the CultureLabs project to pursue a research data open access policy in the attempt of establishing and maintaining a research infrastructure in an ethical manner. The first subsection (4.1) describes the CultureLabs project, including its main objectives, the type of data generated as part of its research activities, and the main intended beneficiaries of this data in its open access form. The reflections stemming from a retrospective look at the data management-related actions and practices within the project have been clustered under three main categories that see a dedicated subsection each: the pursuit of *usefulness* in opening the access to the research data and the subsequent *impact* that pursuing an open access policy may have on the research design (4.2), the distribution of *responsibilities* among the actors involved in the data management process (4.3), and the *practicalities* related to adhering to the *requirements* imposed by the funders (4.4).

4.1. The CultureLabs project

CultureLabs (2018–2021) was a 3-year European project funded under the Horizon 2020 Societal Challenges pillar. The project aimed to develop novel methodologies and ICT tools to facilitate the creation, management, and implementation of social innovation-compatible and cultural heritage-oriented participatory projects. This aim was mainly pursued through a dual approach: on the one hand, the design and development of an online research infrastructure hosting an extensive pool of resources to generate new, or enrich existing, participatory projects; on the other hand, the deployment of four pilot actions that serve as best practices for the involvement of disenfranchised, marginalised, or disadvantaged groups in the fulfilment of cultural heritage-related objectives (Kaldeli et al., 2019).

The CultureLabs Consortium was comprised of nine partners from across Europe, including actors with different expertise such as cultural professionals from museums with experience in community-oriented approaches (People’s History Museum in the UK and Museovirasto/Finnish Heritage Agency in Finland), field staff from non-governmental organisations working with migrants (Cooperative Social COOSS Marche Onlus in Italy), social innovators (Fondazione Sistema Toscana in Italy and Platoniq CoLaboratory in Spain), researchers from the fields of computing and social science (Sheffield Hallam University in the UK and the European Forum for Migration Study of the University of Bamberg in Germany), and technical experts (the Institute of Communication and Computer Systems of the National Technical University of Athens and Singular Logic in Greece).

The several macro areas CultureLabs operated within were translated into a variety of research data collection processes. Throughout the project, the CultureLabs Consortium carried out five areas of research as summarised by Table 1.

Table 1. Summary of research areas and related methodology and research participants of the Horizon 2020 CultureLabs project

RESEARCH TOPIC	METHODOLOGY	RESEARCH PARTICIPANTS
Institutional Stakeholders’ needs and services analysis report	90 surveys 17 semi-structured interviews	Representatives from the GLAM sector (Galleries, Libraries, Archives, and Museums), non-governmental organisations (NGOs), and educational institutions in Europe

Communities needs and living heritage analysis report	101 surveys 27 semi-structured interviews	Representatives from migrant and refugee communities, cultural sector, and NGOs active in cultural work with migrants and refugees
Platform's functional requirements	66 surveys 24 semi-structured interviews	Cultural professionals and community facilitators working with/for migrants and refugees
Platform usability evaluation	30 surveys 5 cooperative evaluations 9 focus groups' participants 8 semi-structured interviews	Cultural professionals, community facilitators working with/for migrants and refugees, human-computer interaction and interaction design experts, pilots' collaborators and participants
Pilots monitoring and evaluation	48 surveys 37 semi-structured interviews 12 focus groups' participants	Pilots' participants (including people with a migratory background) and collaborators (including cultural professionals and professionals dealing social inclusion issues)

Each of these research endeavours was carried out independently, with its own research design and set of research questions, methodologies, and research participants. All research was carried out across Europe with a geographical scope focused on – but not limited to – the partnered countries (namely, Italy, Finland, the UK, Spain, Greece, and Germany). The bulk of the research participants was composed of institutional stakeholders and members of migrant communities. The stakeholders were professionals from a variety of fields related to the wider topic of social innovation via cultural heritage. Stakeholders participating in one or more of the research outputs included, among others, cultural professionals, community facilitators working with and for migrant and refugee communities, and educators. The members of migrant communities involved were people who moved away from their place of usual residence across an international border, temporarily or permanently, to countries in which the research took place. Alongside established first- and second-generation migrants, some of the research endeavours included refugees and asylum seekers.

To systematically pursue openness and transparency, in the early stages of the project the CultureLabs Consortium established a comprehensive Data Management Plan (DMP). The creation of a DMP was required by one of the nine work packages of the project to investigate the existing legal and ethical frameworks at the national and EU level, within which the variety of research data produced as part of the project's research activities (including, among others, surveys', interviews', and focus groups' data as reported in Table 1) was to be managed. With this task in mind, the project focused on creating *open access research data* in light of enhancing the ability of identified stakeholders (namely, cultural heritage institutions, NGOs and other organisations working with/for migrants and refugees, and public administrations) to integrate participatory and collaborative practices in their work, as well as to provide researchers, practitioners and, possibly, immigrant communities around Europe, with information about participation in cultural heritage (CH)-related activities. Together with the wider community of academics and researchers in the field of cultural heritage and social innovation, these listed figures represent the main target audience of the open access data made available by the project.

The choice of these targets is consistent with the idea of contributing to a wider deployment of participatory projects, which represented one of the main ambitions of the CultureLabs project. The lack of easy access to truthful information on best practices is considered one of the main barriers experienced by cultural professionals and community facilitators when trying to deploy digital technologies to facilitate the participation of communities at risk of exclusion in CH-related activities (Giglietto et al., 2019). CultureLabs contributed to overcoming this systematic barrier not only through the design and development of the CultureLabs platform but also by making the research data gathered available to these interested parties.

Such an endeavour led to many decisions from which some lessons learnt as well as considerations revolving around the idea of creating useful open access data, the actors involved in the data management process, the meeting of funding requirements, and the impact on research design have been drawn. These considerations will be discussed in the following four subsections.

4.2. The CultureLabs infrastructure: creating *useful* open access data and *impact* on research design

Although European-funded projects can take advantage of a variety of helpful guidelines and procedures (European Commission, 2014; European Research Council, 2017), managing a research infrastructure of this scale whilst ensuring accessibility and transparency is not an easy task. This case study asserts that the key to creating open access data is making it useful to actors outside the project. To ensure this, careful planning and consideration from the beginning of the project were essential, especially when gathering sensitive data and simultaneously adhering to individual partners' organisational research ethics protocols. For CultureLabs, an additional consideration was that the project's research data was being gathered by several partners with differing levels of experience in relation to conducting and implementing research practice. This meant the need for a shared understanding about applying ethical research practices, towards making the research data useful to external agencies, as well as consideration for which these audiences might be, for example, research communities as well as the professionals figure described in Section 4.1.

Although the emphasis of each of the research activities listed in Table 1 was on gathering full details on the professional circumstances as well as the social and cultural contexts of the research participants, this emphasis was sidelined when it was time to openly share this research data. In fact, the decision was taken to completely anonymise all open access data to protect the individuals involved who had shared sensitive data about themselves and their contexts. The refugee or pending status of some individuals further influenced this approach as part of the Consortium's research duty of care. For the open access data, the process of anonymisation meant that the finer details were partially lost, creating a paradox by which the data managed internally is richer and potentially more useful than the data made openly available to interested parties. On the one hand, omitting data to protect participants' individual identities is essential for ethical data collection and wider dissemination; on the other hand, removing contexts can reduce the richness and meaningfulness of the data.

CultureLabs chose to gather detailed research data that was needed for the deliverables – especially as personal information was helpful for future contact and subsequent methodologies. As the project opted for not anonymising the data stored privately by keeping sensitive content, the research design was not compromised and rich and meaningful data could be gathered. It was only upon the second planned iteration of the DMP, almost three years into the project, that the need for open access data was acted upon by removing all personal information (including names, countries of origins, country of residence, and affiliation) before uploading the datasets publicly. In hindsight, finding a balance for anonymising data by replacing personal details with contextual pseudonyms so that individual identities were not revealed but context was preserved through metanarratives would have been a preferred approach, but time constraints impacted the approach taken to be compliant.

It is important to reflect upon the consequences of this choice in terms of the usefulness of open access data. The cultural context for research data can be what makes it meaningful to others. CultureLabs' Consortium administered research across several European countries and the context was unique in each case. Therefore, the researchers also needed to think about how the cultural context was represented (ethically) within the data set and the management of this can be difficult when individual case studies may reveal such different results. It would have been ideal for CultureLabs to have designed the research around the open access nature of the data management so that certain questions such as personal individual and demographic data were never included. However, this would have contradicted social science data collection protocols whereby demographic details and personal information are gathered in order to enrich, cross-compare the data, and conduct follow-up activities with participants. This final arrangement can be considered partially fruitful in terms of usefulness of data. The research data generated by the project and stored internally can serve as a basis to explore, understand, and disseminate about complex issues such as practices of institutional stakeholders involved in cultural approaches and/or the promotion of social integration of disadvantaged groups (Giglito et al., 2019), the needs and expectations of migrant and refugee communities in relation to sharing their intangible cultural heritage, and the role of digital technologies in bridging cultural heritage communities (Giglito et al., 2021). However, the data ultimately uploaded publicly has undergone an additional process of anonymisation to be GDPR compliant that caused some of its heuristic value to be sacrificed in the name of open accessibility.

4.3. Responsibility and control: the actors in the research data management process

A key challenge in devising a research infrastructure is establishing responsibility for the research data management processes. This includes where the data is stored during and after the research has been conducted, the length of time it is available as open access, as well as upon whom the responsibilities of monitoring it lie beyond the lifespan of the project. For CultureLabs, some of these issues were negotiated early in the project and based upon the project's description of work but the finer details to action these protocols were established in a more reactionary manner. For instance, it was agreed that a cultural institution was responsible for the DMP based upon the organisation's large size and capacity to execute it and the assumed expertise related to intellectual property rights based on their extensive archives and collections. The DMP developed by CultureLabs was designed with these considerations in mind and Sheffield Hallam University (SHU) – one of the project partners – took responsibility for the storage and management of the research data, given the strength of their existing research concordat and their institution's technical infrastructure. It is crucial that large research projects dedicate, from the outset, a proper amount of time for identifying the suitable expertise and available resources to lead the data management process, in light of ensuring smooth processes concerning complicated procedures such as storage and long-term preservation as well as the legal and ethical aspects associated with it.

For CultureLabs, it was also helpful to ensure that there was an element of knowledge transfer in relation to key aspects of research governance across the project. Whilst academics may adhere to responsible research protocols (to help ensure robustness and integrity), partners with less practical research experience may be unaware of professional standards in relation to research practice (UKRIO, 2009; ALLEA, 2017). Gathering research data was not part of the organisational mission of some partners, who therefore did not have any formal regulations or policies concerning data management, research ethics, or research integrity that they could draw from within their organisations. A key consideration for the project as a whole was negotiating between the existing data governance models, and therefore partners with stricter and more established policies and codes of practice took more ownership towards guiding those with less extensive experience. Whilst some time was factored to consider these aspects, the reality is that

organising and coordinating several different partners towards a concrete working agreement may take more time than planned.

4.4. Practicalities: meeting the funders' requirements

To comply with the EU data protection rules (European Commission, 2020a), the data produced within the CultureLabs project data needed to be stored securely. Simultaneously, it was also essential that multiple partners based across Europe were able to access a central system whereby they could upload, store, and access data. There also needed to be a separate repository allowing the data to be openly accessible by interested parties.

To fulfil the first objective, CultureLabs used an existing secure server supplied by SHU which allowed access to the project partners through a password-protected system that ensured the project data was stored and shared securely. Whilst setting up such a system may appear straightforward, it takes additional time for coordination and continuous use. For instance, requests for access to the servers were dependent upon a central IT service within the host institution. The hosts were therefore responsible for mediating all communications between the partners and the IT service.

To openly share data, CultureLabs made use of Zenodo, which is a repository that supports most file formats, while being Open AIRE-compliant (Sicilia et al., 2017). Given that it placed no restrictions on licenses, it allowed the Consortium to consider suitable Creative Commons licenses. For this project, data associated with academic publications was licensed under Creative Commons Attribution-NonCommercial 4.0 International, whilst all metadata can be exported via OAI-PMH and harvested. Submissions to Zenodo are given a DigitalObject Identifier (DOI) and items are retained for the lifetime of the repository. Data files and metadata are backed up every 24 hours and copies of the data are replicated into an online system. Whilst registration is required to upload and submit data, it is not needed for downloading data that is under open access status from the repository (CERN & OpenAIRE, n.d.).

Unsurprisingly, European projects are required to be compliant with the open access regulation (European Commission, 2014) but unless there is existing key expertise within a project's consortium, additional considerations for the time and administration that is involved for training and complying with open access requirements needs to be given. A clear understanding of the requirements for compliance for open access is essential from the start of any EU project, therefore, prospective consortiums should carefully consider a partner who can lead to ensure that the project will be compliant. Additionally, part of the data management expertise concerns not only the consideration for who will be responsible for the management of the data, but also the decision about who will be responsible for monitoring the infrastructure and ensuring compliance with GDPR guidelines as well as any additional national regulation in place.

Being GDPR compliant also means reflecting upon which digital survey tool to use for data collection. For instance, those based with servers outside of the EU are less likely to be GDPR compliant. For CultureLabs, that meant that online surveys could not be distributed via a commonly used service like Google Forms, nor could data be temporarily stored for example on Google Drive or in Dropbox, which were all platforms the partners were familiar with. For survey distribution, CultureLabs chose to make use of an open source GDPR compliant software called KoBoToolbox. Although fit for purpose, the use of this tool required the partners to develop the necessary skills to learn how to use it effectively. Besides, the project needed the involvement of an IT expert that could create and manage a local version of the software to be adapted to the needs of each research design. This demonstrates how the choice of appropriate survey tools for data collection should be therefore planned in advance, consistently with the available expertise within a consortium. Whilst adhering to this regulation should be standard practice, regulating individuals' use of tools may be challenging, especially across international boundaries. Partners can recommend and guide others but they ultimately cannot easily control which systems individuals use. This is particularly poignant considering the compelling fines established for

infringing GDPR (Wolford, n.d.).

Finally, the multi-linguistic aspects of the data should also be considered. Ideally, research gathered in one language needs to be translated for wider dissemination for international open access. Although an inevitable part of most collaborative international projects, translating data while adhering to the GDPR framework is not an easy task, and the time or the money to do it are not usually factored in. International projects have to rely on expensive translation services from companies that are aware of or sensitive towards GDPR issues or, if possible, take advantage of internal resources such as bilingual partners. The latter option was the one adopted by the CultureLabs' Consortium, as part of the collected data was in a variety of languages such as Finnish, German, and Italian, and had to be translated into English in order to enhance the reusability of the resulting shared datasets.

5. CONCLUSIONS AND DIRECTIONS FOR FUTURE STUDIES

In the era of collaborative and open science, research infrastructures play a critical role, promoting dynamic research processes and in conveying the expanding “data deluge” (Hey et al., 2009) available to scientists and other potential users. On the one hand, these resources aim to support fluid ways of organizing scientific work by furthering a collaborative model based on common research interests by sharing scientific and technical knowledge and methodological practices (Nielsen, 2011); on the other hand, they contribute to the institutionalization of scientific results as a public good (cf. Boulton et al., 2011) by redefining the relationship between the field of research and society, as well as by providing a common space for discussion.

The case of the EU Horizon 2020-funded project CultureLabs' showed what the process of building and managing a research infrastructure that adheres to the principles of open access entails. Using the CultureLabs project as a case study, this paper highlighted a number of important considerations to be addressed which are generated as a result of having devised a research infrastructure for the compliance with open access data. Particularly, the authors concentrated on some selected points in this regard: i) creation of “useful” and GDPR-compliant data and the impact on research activities as a result of their (re)utilisation; ii) the control of the data management process; iii) the compliance with the funders requirements (e.g. in terms of data security).

As this paper draws directly from CultureLabs' experience, the reflections reported here may constitute a preliminary framework for in-depth studies about the management of research infrastructures. Following, the authors outline five possible directions for studies concerning the management of research infrastructures. The various challenges that arise when considering open access rules in research and how to best handle them are left for discussion as open questions.

Dissemination and awareness. In the case of the CultureLabs project, one of the reasons for creating open access data was to ensure that cultural heritage institutions, NGOs and other organisations working with/for migrants and refugees, and public administrations as well as researchers, practitioners and, possibly, immigrant communities around Europe could more easily access data information about the project. In the actual experience of the project, it was found that ‘access’ is more complicated than simply accessing data files. Individuals need to have knowledge of the data repositories, as well as the technological skills and system capabilities to access the files and make use of them (even if data has been organised and made accessible). Just because data is made available does not mean that it will be accessed, consulted, or used. How effective is research infrastructure dissemination in clarifying contents and possibilities? How much of it is aimed at certain target audiences? And, to what extent is the intended audience aware of the presence of these resources?

Incentives for a definitive breach for sharing. Some researchers are still concerned about someone “stealing” their data, rather than having a vision open to the concept of “sharing” from which it is possible to obtain mutual benefits in terms of knowledge growth. Particularly in the social sciences, the motivation to share does not appear to have spread, and the exploitation of

shared resources is still in its early stages. From this point of view, we question if there is a need to include open science practices in research evaluation for career advancement. Furthermore, possible directions for future research can turn towards determining whether seniors favor or discourage open science practices among junior colleagues.

Open access ‘best practises’. Evaluation of research infrastructures is essential for generating “best practice” for data research open access as well as identifying possible limits in current approaches to managing open access data. But, since data differs so greatly from project to project, what really is best practice? While practices in regard to open access data may be improved by considering specific research infrastructures, they may be improved even further by monitoring who actually accesses the data and for what purpose so as to understand how data is reused. To do this, it would be costly for the projects both in terms of time and access to the monitoring systems but, in the long term, detailed insight of how the data is used could help to further shape practices.

Expertise for building and managing open research infrastructures. Expertise is needed not only for the production of the contents of a research infrastructure (scientific expertise) but also for understanding and applying policy such as GDPR (and what limitation it imposes) and other regulations (administrative and technical expertise). The working groups that are involved in the construction of an open research infrastructure must be able to count on real technical and legal support. To what extent are these skills required from the same researchers who contributed to the creation of the data? If the researcher is partly required to have expertise in computer science, how much are they willing to invest in legal training?

Interoperation and searching tools. The construction of common data structures for interoperability and reusable data flows are essential for implementing a system of research infrastructures which is accessible and user-friendly. Common agreed rules on metadata can also facilitate the processes of searches. Sometimes, good metadata services, both for stand-alone and integrated infrastructures can be more important than the actual data since they can summarize the real contents for the users. Furthermore, given the “data deluge”, if open access data becomes standard, finding and searching for specific data sets may become more difficult as the data amounts and repositories grow. How much “connectable” infrastructures are really interoperable and to what extent are metadata services being appropriately developed and using a common set of standards?

We believe that these open questions can serve as key starting points for critical discussions that are needed to enrich the practices around open access and data management, as well as to provide the ground for future analysis and empirical work.

6. ACKNOWLEDGMENTS

This work is partly supported by the European Union’s Horizon2020 project CultureLabs: Recipes for Social Innovation, under Grant Agreement 770158. The European Union is not responsible for any use that may be made of the information contained therein.

7. REFERENCES

- All European Academies (ALLEA) (2017). *The European Code of Conduct for Research Integrity: Revised edition*. Available at: <https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf>
- Boulton, G., Campbell, P., Collins, B., Elias, P., Hall, W., Laurie, G., O’Neill, O., Rawlins, M., Thornton, J., Vallance, P. & Walport, W. (2012). *Science as an open enterprise*. London: The Royal Society.

- Borgman, C.L. (2007). *Scholarship in the Digital Age. Information, Infrastructure, and the Internet*, Cambridge: MIT Press.
- Borgman, C.L. (2012). The Conundrum of Sharing Research Data. *Journal Of The American Society For Information Science And Technology*, 63(6), pp. 1059-1078.
- CERN, & OpenAIRE. (n.d.). *Frequently Asked Questions*. Zenodo.Org. Retrieved 10 August 2021, from <https://help.zenodo.org/>
- Chawinga, W.D., & Zinn, S. (2019). *Global perspectives of research data sharing: A systematic literature review*. *Library & Information Science Research*, 41(2), pp. 109-122.
- Ciula, A. Nyhan, J., & Moulin, C. (2011). *Research Infrastructures in the Digital Humanities*, ESF Science Policy Briefing 42, European Science Foundation (ESF).
- Duša, A., Nelle, D., Stock, G. & Wagner, G.G., (2014). *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*, Berlin: SCIVERO Verlag.
- Dutton, W.H., & Meyer, E.T. (2009). *Experience with New Tools and Infrastructures of Research: An exploratory study of distance from, and attitudes toward, e-Research*. In "Prometheus", 27(3), pp. 223-238.
- ESFRI (2018). *Strategy Report On Research Infrastructures – Roadmap 2018*, Dipartimento di Fisica - Università degli Studi di Milano.
- European Commission (2008). *Proposal for a Council Regulation on the Community legal framework for a European Research Infrastructure (ERIC)*, available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52008PC0467&from=EN>
- European Commission (2014). *H2020 Online Manual*. European Union. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/index_en.htm
- European Commission (2020a). *EU data protection rules*. EC.Europa.Eu. https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules_en
- European Commission (2020b) *Supporting the Transformative Impact of Research Infrastructures on European Research*. Luxembourg: Publications Office of the European Union.
- European Research Council. (2017). *Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020*. European Commission. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf
- German Council of Science and Humanities (2011). *Recommendations on Research Infrastructures in Humanities and Social Sciences*, German Council of Science and Humanities, https://www.wissenschaftsrat.de/download/archiv/10465-11_engl.pdf
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London: SAGE.
- Giglito, D., Ciolfi, L., Claisse, C., & Lockley, E. (2019). Bridging cultural heritage and communities through digital technologies: Understanding perspectives and challenges. In *C&T '19: Proceedings of the 9th International Conference on Communities & Technologies - Transforming Communities*, pp. 81-91.
- Giglito, D., Ciolfi, L., & Bosswick, W. (2021). Building a bridge: Opportunities and challenges for intangible cultural heritage at the intersection of institutions, civic society, and migrant communities. *International Journal of Heritage Studies*, pp. 1-18. <https://doi.org/10.1080/13527258.2021.1922934>
- Hey, T., Tansley, S., & Tolle, K. (2009). Jim Gray on e-science: A transformed scientific method. In Hey, T., Tansley, S., & Tolle, K. (Eds.). *The fourth paradigm. Data-intensive scientific discovery*. Redmond, Washington: Microsoft Research, pp. xvii-xxxi.

- Jarolimkova, A., & Drobikova, B. (2019). Data sharing in social sciences: Case study on Charles University. In Kurbanoglu S., Špiranec S., Ůnal Y., Boustany J., et al. (Eds.), *Information literacy in everyday life*. New York, NY: Springer International Publishing, pp. 556-565.
- Kaldeli, E., Tsakou, G., Giglitto, D., Cesaroni, F., Tzouvaras, V., & Stamou, G. (2019). CultureLabs: Cultural heritage and digital technology at the service of social innovation. In Antoniou, A. & Wallace M. (Eds.), *Proceedings of the Workshop on Cultural Informatics co-located with the 14th International Workshop On Semantic And Social Media Adaptation And Personalization (SMAP 2019)*. <http://ceur-ws.org/Vol-2412/paper9.pdf>
- Kim, Y., & Adler, M. (2015). *Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories*. *International Journal of Information Management*, 35(4), pp. 408-418.
- Martin, B. (2003). The changing social contract for science and the evolution of the university. In Geuna, A., Salter, A.J., Steinmuller, W.E. (eds.), *Science and innovation. Rethinking the rationales for funding and governance*. Cheltenham: Edward Elgar, pp. 7-29.
- MERIL project (2011). *MERIL, Research Infrastructures of European relevance. A comprehensive inventory*, European Science Foundation, ISBN: 978-2-918428-45-9.
- Nielsen, M. (2011), *Reinventing discovery: The new era of networked science*. New Jersey: Princeton University Press.
- OECD (2013). *New Data for Understanding the Human Condition. International Perspectives*. OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences, OECD.
- Pasquetto, I.V., Sands, A.E., & Borgman C.L. (2015). Exploring openness in data and science: What is "open," to whom, when, and why? In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. Article No. 141.
- Renschler, B., Kleiner, I., Wernli, B., Farago, P., & Joye, D. (Eds.) (2013). *Understanding Research Infrastructures in the Social Sciences*, Zurich: Seismo Press.
- RISIS project (2013), *DOW RISIS (grant agreement no. 313082)*, available at: <https://www.risis2.eu/official-documents/>
- Sicilia, M.-A., García-Barriocanal, E., & Sánchez-Alonso, S. (2017). Community curation in open dataset repositories: Insights from Zenodo. *Procedia Computer Science*, 106, pp. 54–60.
- Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). *Open data in global environmental research: The Belmont Forum's open data survey*. *PLoS ONE*, 11 (1), Article e0146695.
- Spinello, A.O. (2019). Le infrastrutture di ricerca nelle scienze sociali: uno studio di caso su utilizzo ed effetti nella produzione scientifica. *Welfare e Ergonomia*, 1/2019, pp. 67-78.
- UKRIO (2009). *Code of practice for Research: Promoting Good Practice and Preventing Misconduct*. Available at <https://ukrio.org/wp-content/uploads/UKRIO-Code-of-Practice-for-Research.pdf>
- Wolford, B. (n.d.). *What are the GDPR Fines?* GDPR.Eu. Retrieved 11 August 2021, from <https://gdpr.eu/fines/>

CNR-IRCrES Working Papers

2021

- N. 8/2021 [Francia-Italia: parole in campo. Intorno alla narrazione del Campionato del Mondo di calcio donne](#). Antonella Emina.
- N. 7/2021 [Covid-19 e rischio di insolvenza: il punto di vista del mercato azionario](#). Franco Varetto.
- N. 6/2021 [Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform](#). Greta Falavigna, Roberto Ippoliti.
- N. 5/2021 [Different waves and different policy interventions in 2020 Covid-19 in Italy: did they bring different results?](#). Mario Nosvelli.
- N. 4/2021 [On Search of a General Model of Technology Innovation](#). Angelo Bonomi.
- N. 3/2021 [Design and implementation of a web survey on the effects of evaluation on academic research](#). Andrea Orazio Spinello, Emanuela Reale, Antonio Zinilli.
- N. 2/2021 [An online survey on the effects of agile working in Italian Public Research Organisations](#). Serena Fabrizio, Valentina Lamonica, Andrea Orazio Spinello.
- N. 1/2021 [Technology Transfer Activities in Universities and Public Research Organizations: A Literature Overview](#). Ugo Finardi, Rolfo Secondo, Isabella Bianco.

2020

- N. 12/2020 [Unexpected loss multiperiodale e pricing del rischio di credito](#). Franco Varetto.
- N. 11/2020 [La ricerca in Nanotecnologie e Nanoscienze in Italia: spesa del settore pubblico e aree tematiche prevalenti](#). Ugo Finardi, Andrea Orazio Spinello.
- N. 10/2020 [Persistent fast growth and profitability](#). Lucio Morettini, Bianca Potì, Roberto Gabriele.
- N. 9/2020 [Binomio Burnout e Mindfulness nelle organizzazioni. Alcuni studi e scenari di applicazione](#). Oriana Ippoliti, Riccardo Briotti, Bianca Crocamo, Antonio Minopoli.
- N. 8/2020 [Innovation and communication of companies on Twitter before and during COVID-19 crisis](#). José N. Franco-Riquelme, Antonio Zinilli, Joaquín B. Ordieres-Meré and Emanuela Reale.
- N. 7/2020 [The proposal of a new hybrid methodology for the impact assessment of energy efficiency interventions. An exploratory study](#). Monica Cariola, Greta Falavigna.
- N. 6/2020 [The technology innovative system of the Silicon Valley](#). Angelo Bonomi.
- N. 5/2020 [Storia dell'industria delle macchine utensili in Piemonte dalle origini alla seconda guerra mondiale](#). Secondo Rolfo.
- N. 4/2020 [Blockchain e Internet of Things per la logistica Un caso di collaborazione tra ricerca e impresa](#). Edoardo Lorenzetti, Lucio Morettini, Franco Mazzenga, Alessandro Vizzarri, Romeo Giuliano, Paolo Peruzzi, Cristiano Di Giovanni.
- N. 3/2020 [L'impatto economico e fiscale di un evento culturale: misure e scala territoriale](#). Giovanna Segre, Andrea Morelli.
- N. 2/2020 [Mapping the tangible and intangible elements of the historical buildings and spaces](#). Edoardo Lorenzetti, Nicola Maiellaro.
- N. 1/2020 [Il lavoro agile negli enti pubblici di ricerca](#). Emanuela Reale, Serena Fabrizio, Andrea Orazio Spinello.

2019

- N. 6/2019 [Women's candidatures in local elections: does the context matter? Empirical evidence from Italian municipalities](#). Igor Benati, Greta Falavigna, Lisa Sella.
- N. 5/2019 [Research activities in Nanotechnologies and Nanosciences: an analysis of Piedmont's nanotech research system](#). Ugo Finardi.
- N. 4/2019 [Xylella fastidiosa: patogenesi, danni economici e lotta al disseccamento rapido dell'olivo](#). Maurizio Conti.
- N. 3/2019 [Flussi di traffico attraverso il tunnel automobilistico del Frejus: un semplice esercizio di forecasting e alcune considerazioni a margine](#). Ugo Finardi.
- N. 2/2019 [The Start-up Venture Capital Innovation System Comparison with industrially financed R&D projects system](#). Angelo Bonomi.
- N. 1/2019 [Complessità delle organizzazioni, complessità della formazione. Report di studio qualitativo ed analisi ermeneutica del Modello TRASE – IRCRES/CNR-IMO](#). Anna Chiara Scardicchio.

[Numeri precedenti/Previous issues](#)

Research funding organizations, particularly at international level, are increasingly promoting the creation and maintenance of open access research infrastructures (RI). These resources have assumed a pivotal role as support for the new open and networked science in their dimension of technical and operational frameworks that allow scientists and stakeholders to collaborate and share scientific data and results. In Social Sciences and Humanities (SSH), the creation and exploitation of open data platforms is still attempting to catch up with longer-standing practices in the “hard sciences” as the resistance to wider data sharing has not yet been completely overcome. This paper aims to describe how a large project, financed by the European Commission, managed the creation of a RI in the field of SSH, showing the steps undertaken to comply with the GDPR regulations and prepare the data for useful sharing and reuse. In this regard, the authors present the case study of the Horizon 2020 “CultureLabs” project, placing emphasis on some specific practical factors that they believe are particularly important for implementing open access principles in the establishment and maintenance of RIs in the new course of science based on sharing and openness. In particular, the authors will focus on creation of “useful” and GDPR-compliant data and the impact on research activities as a result of their (re)utilisation; the control of the data management process; and the compliance with funders’ requirements (e.g. in terms of data security). The reflection on the interplay of these aspects, operated through a case study, appears to be crucial in moving away from a merely theoretical approach to addressing the issue of open access, and it hopes to serve as a guide or a warning for those who create and administer open RIs.