

Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782

Start Date of Project: 01/01/2019

Duration: 40 months

Deliverable 9.1 Challenges that user communities face when attempting to contribute to SSHOC

Dissemination Level	PU
Due Date of Deliverable	30/04/19, M04
Actual Submission Date	18/11/19
Work Package	WP 9 - Data Communities
Task	Task 9.1 Identifying shared and unique challenges for SSH data communities
Type	Report
Approval Status	
Version	v1.0
Number of Pages	p.1 – p.24

Abstract:

This report identifies obstacles that different kinds of user- and data-communities encounter in relation to engaging with, using and contributing to the evolving EOSC/SSHOC infrastructure.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/ her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



History

Version	Date	Reason	Revised by
0.1	29/09/2019	1st draft by Cees Van Der Eijk (UNOTT)	
0.2	11/10/2019	2nd draft by Cees Van Der Eijk (UNOTT)	
0.3	05/11/2019	Peer review	Laura Morales (SciencesPO)
0.3	08/11/2019	Address peer review comments	Cees Cees Van Der Eijk (UNOTT)
1.0	13/11/2019	Final editorial review	Martina Drascic (CESSDA ERIC)
1.1.	15/11/2019	Address editorial comments	Cees Cees Van Der Eijk (UNOTT)

Author List

Organisation	Name	Contact Information
UNOTT	Cees Van Der Eijk	Cees.Van_Der_Eijk@nottingham.ac.uk

Executive Summary

This report analyses obstacles that user- and data-communities and their members are likely to experience when attempting to engage with, make use of the products and services provided by, and contribute to the evolving EOSC/SSHOC infrastructure. This remit is described in section 1.

The report then analyses (in section 2) the concept of user-communities, in order to subsequently distinguish different kinds of user-communities and different kinds of members. The reason for these distinctions is that the kinds of obstacles to be experienced cannot be expected to be uniform, but to vary because of the characteristics of communities and members. With respect to user-communities an important difference is whether they are included in the SSHOC Consortium (many SSH user communities are not, or not yet, yet SSHOC aspires to be also relevant for them). Additional differences relate to their organisational strength. With respect to members of user communities, the most relevant distinction is between 'leading' members (who occupy a role or position in the organisations of user communities) and 'regular' members.

The report then distinguishes (in section 3) different ways in which user-communities and their members can interact with the evolving EOSC/SSHOC infrastructure, and describes three of such ways: 'engaging with' (mainly in the form of attentiveness), which is seen as a first step towards 'using' (which consists of drawing on the infrastructure to acquire data, tools, and other products and services for use in research). The third way of interaction consists of 'contributing to' (which involves making one's own research, its data, tools, outcomes and publications available to others via the EOSC/SSHOC infrastructure).

In section 4 a bounded rationality perspective and a socialisation perspective are used to analyse factors that create obstacles for user-communities and their members to productively interact with the EOSC/SSHOC infrastructure. This results in the identification of the following four generic obstacles: (1) unfamiliarity; (2) expense (in terms of money, time, attention and training); (3) low priority (as a consequence of opportunity costs); and (4) inability (in the form of insufficient relevant skills and competence).

Section 5 discusses how each of these obstacles impinges on different kinds of user-communities and different kinds of their members, a discussion that is summarised in the form of low, medium or high relevance of each of the four identified obstacles for the various kinds of user-communities and members that were distinguished earlier, in section 2.

Abbreviations and Acronyms

EOSC	European Open Science Cloud
SSHOC	Social Sciences and Humanities Open Cloud
SSH	Social Sciences and Humanities
EPOP	'Election, Public Opinion and Parties' organised section of the British Political Science Association
IMISCOE	'International Migration, Integration and Social Cohesion' Research Network
ESS	'European Social Survey' ERIC
NES	National Election Study
(I)NGO	(international) Non-Governmental Organisation
MEDem	'Monitoring Electoral Democracy', Research network on Electoral research (in development)
Ethmigsurveydata	The International Ethnic and Immigrant Minorities' Survey Data Network; COST Action 16111

Table of Contents

1.	<i>Introduction and aims of this deliverable</i>	<i>6</i>
2.	<i>Varieties of (potential) user communities and individual end-users</i>	<i>7</i>
3.	<i>Varieties of uses, engagement and contributions</i>	<i>11</i>
4.	<i>Varieties of obstacles to engagement with, use of and contributing to SSHOC, and their background</i>	<i>13</i>
	4.1 The perspective of bounded rationality	14
	4.2 The socialisation perspective.....	17
	4.3 Obstacles identified by the bounded rationality and socialisation perspectives.....	19
5.	<i>Differential importance of obstacles for different kinds of users and user-communities</i>	<i>20</i>

1. Introduction and aims of this deliverable

This report focuses on challenges and obstacles that user communities in the social sciences and humanities may experience when engaging with the EOSC/SSHOC infrastructure, once it has been delivered and is available. The purpose of this report is to identify factors and conditions that may obstruct optimal engagement from relevant user communities, in order to support potential users and undertake anticipatory activities that may ameliorate any obstacles that may otherwise obstruct them from productive engagement.

In view of the very recent start of the consortium, this report can, at the time of writing be based only to a very limited extent on actual experiences of user communities with the EOSC/SSHOC infrastructure, or on surveys of the actual usage of the SSHOC infrastructure. This report is therefore mainly based on first-principle reasoning, which is, where relevant, supported by research from a variety of social disciplines.

To fulfil the remit of this deliverable, three elements must be discussed, each of which will be addressed in the following sections:

First, we must discuss different kinds of user communities and members of these communities. It is important to recognise how different these communities and their members can be because not all of these will experience the same challenges (or the same incentives), depending on their own characteristics.

Thus, in **Section 2** of this document we discuss what user communities are, in what forms they present themselves in the social sciences and humanities, and how they relate to their members (the end-users of the products and services EOSC/SSHOC aims to provide).

Second, we will discuss in **Section 3** of this document the various ways in which user communities and end-users may engage with the infrastructure of SSHOC and the various products and services that it will contain. These different kinds of usage must be distinguished, because they cannot be expected to be all subject to the same challenges.

Third, we will discuss different kinds of challenges and obstacles that may stand in the way to actual usage of and contributions to products and services that may, in principle, benefit (potential) users; this will be done in **Section 4**.

In **Section 5** we then combine the 'building blocks' developed in the previous sections by linking different kinds of challenges to particular kinds of user communities (and end-users) and to particular kinds of usage that they can make of what SSHOC will have on offer.

2. Varieties of (potential) user communities and individual end-users

When discussing challenges that may obstruct user communities and their members of making use of and contributing to SSHOC it is necessary to distinguish between different kinds of user communities and different kinds of users. These distinctions are important because not all users or user communities will necessarily face the same challenges. In this section what user communities are, how they may differ in relevant aspects, and the ways in which they may be related to their individual members.

SSHOC focusses its aims and aspirations on *communities* which are described in the original consortium proposal as, respectively, the *research community*, the *SSH scientific community*, *data research community*, *user community*, or *data community*. This focus on communities is also evident from the frequent reference to the need for *community driven* developments, for *community building* and *community interaction*, for *community features* in the aspired infrastructure, and so on. In spite of these frequent references to community in the singular, it is more useful to think of communities in the plural.¹

They can be distinguished in a variety of ways which are often cross-cutting. A popular way to distinguish between different communities is on disciplinary grounds, leading to communities of, e.g., sociologists, educationalists, economists, and so on. In view of the very broad character of these disciplinary communities it has become quite common to make further distinctions in terms of sub-disciplines (e.g., within sociology there are clearly recognisable sub-disciplines such as sociology of religion, sociology of the family, political sociology, etcetera).

These disciplinary and sub-disciplinary communities can generally be identified by their institutional forms, such as the existence of professional organisations, and organised channels for internal scholarly communication and research exchange in the form of, e.g., conferences and specialised professional journals. Yet, these disciplinary and sub-disciplinary communities are generally far from homogeneous, and they contain many smaller groups that are more united in terms of paradigms, conceptual frameworks, research topics, empirical approaches and so on. These smaller and more homogeneous communities are occasionally even large enough to host their own more specialised conferences and publish their own journals. One may think in this respect of such groups as the German *Forschungsgruppe Wahlen*, the EPOP organised section (Elections, Public Opinion and Parties) of the British Political Studies Association, or the migration studies community organised around the IMISCOE network (which emerged from an FP6 Network of Excellence). It is particularly at the level of these kinds of relatively homogeneous research communities that we find that large numbers of

¹ At a higher level of abstraction, one can speak more easily in the singular, as when referring, for example, to 'the scientific community', but that is of little practical use for the concerns discussed in this document.

their members find particular datasets or data collections of central importance in their scientific work, and it are these groups that are referred to in the SSHOC proposal as *user communities* or *data communities*.

Referring to these relatively homogeneous groups of researchers as *communities* suggests that they are – to a relevant degree – cohesive in outlook, purpose, behaviour and preferences. As already hinted at above, organisational structures (such as professional societies, conferences and professional journals) can be regarded as indicators of the existence of such communities, while, at the same time, they help constitute them. Therefore, the partners in the SSHOC consortium, in their presence as organised and structural providers of scientific information or services generate as well as represent their own user- or data communities. Many of these organisations require users to register in order to be eligible to make use of their products or services; this makes it possible to inform them of new or updated materials, and so to help develop a group of users into a user community. Such registration as requisite for downloading and using data is enacted by many large and infrastructural data and service providers. As a case in point, the ESS (European Social Survey) recorded in April 2019 that they had 139,864 registered users.² Cohesiveness of a community can manifest itself in interactions between its individual members, for example via collaborative research projects, citations to each other's work, co-authorships, etc.

How homogeneous user-communities actually are that rely on specific (collections of) datasets depends to a large degree on the width of the scope of uses for which the data in question can be used productively. Amongst the avid users of the ESS, for example, we find many political sociologists, who find the contents of the ESS data highly relevant for their substantive interests. In this respect they are joined by others who are not political sociologists at all, but who may be sociologists of different ilk (who also find data of central relevance in the ESS), or who do not consider themselves sociologists at all, but rather political scientists, or communication researchers, or social psychologists and so on. In short: these data- or user-communities represented by SSHOC consortium partners may, in spite of their shared interest in particular data or services, nevertheless be quite heterogeneous in terms of disciplinary background and other characteristics. Moreover, these user- or data-communities are not mutually exclusive. Thus, some of the researchers who belong to, e.g., the ESS user community will also belong to, e.g., the SHARE user community, and yet other ones. The reasons for this non-exclusivity is straightforward: individual researchers will use data that are *relevant* for the research questions they pursue, irrespective of their specific origins. They cannot afford to rely exclusively on data from

² See https://www.europeansocialsurvey.org/docs/data_users/ESS_data_user_stats_apr_2019.pdf. This number understates in some respects the actual ESS user community and overstates it in other ways. It understates it because researchers may take a close interest in analyses of ESS data that are published in the scholarly literature, and for which no registration is required, or they may be part of groups which use ESS data which are obtained by only a single (registered) member. At the same time, it overstates the user community somewhat as not all registered users have actively used ESS data. Of the registered users 101,696 (or 73%) have downloaded (and therefore in likelihood actively used) ESS data. Moreover, these numbers are cumulative totals for the entire period since the launch of the ESS in 2002, so that at any specific moment a considerable proportion of them will not be active anymore, or not yet be active.

a particular source if other relevant data also exist, as doing so will diminish their chances of getting their research published in high-ranking outlets. As we will see below, the relevance of this discussion about user- and data-communities is that the extent of their non-exclusivity and heterogeneity of members helps define and constrain the relationships between these organised communities on the one hand, and their individual members on the other hand.

The partners of the SSHOC consortium represent user- and data communities in the social sciences and humanities, but obviously not exhaustively so. Many large and vibrant research communities are not (yet) included, such as those in communication science, in modern slavery research, in history, in electoral research, in education, and so on. Many of the user communities in these, and other areas are highly developed, and associated with important and widely used infra-structural data collections. The field of electoral research, for example, is linked to NESs (National Election Studies) in virtually all established democratic systems, which provide not only data for research communities in their own countries, but also for comparative researchers. In addition to nationally based studies, this field has also since long generated primary and infrastructural comparative data, for example through the EES (European Election Studies) and the CSES (Comparative Study of Electoral Systems). Similar observations can be made for other user-communities referred to above, as well as for many others which remain unmentioned.

The aspiration of SSHOC is to be relevant not only for its consortium partners and their associated user-communities, but also for other user- and data-communities in the social sciences and humanities. Indeed, the SSHOC proposal recognises explicitly that the SSH domain is wider than what is covered by the consortium partners; that the products and services of the EOSC/SSHOC infrastructure must be relevant for users beyond the consortium; and that the consortium aims to link up to user communities in the social sciences and humanities that are not yet included in its activities. The potential of such relevance for other user communities derives at least to some extent from that many of the deliverables to be produced within SSHOC can serve as prototypes or examples that can then be followed in other user- or data-communities.

The challenges and obstacles for all these different user communities and their members to engage with and contribute to SSHOC will vary as a function of their characteristics. Without necessarily being exhaustive we can, in this respect, distinguish between user communities along the following lines:

- User communities that are **included in the SSHOC consortium** and those that are **not**. Communities that are part of SSHOC will have more opportunities and experience fewer challenges and obstacles to using and contributing than others for obvious reasons: greater awareness, more and more timely information, having a direct stake in the consortium, having already defined the most obvious ways in which they can contribute, and so on.
- Particularly for user communities that are not (yet) part of or closely linked to the consortium: the **organisational strength** of the community. User communities that are more developed in organisational terms offer greater possibilities to contribute to SSHOC for a variety of reasons. On the one hand, they can more easily be identified and approached by SSHOC and be provided with relevant information. Similarly, a developed organisational structure gives a community more opportunities to become aware of and approach SSHOC, and subsequently undertake activities that either involve direct engagement, of that will

facilitate this. Additionally, as will be discussed shortly, organisationally better developed user communities have more opportunities to inform, motivate and activate its own members (its individual researchers) to engage, make use of and contribute to SSHOC.

Organisational strength can manifest itself in several ways. One of these consists of the presence of a formal board or executive, and hence an established institutional actor for structured contact with SSHOC. A second aspect of organisational development concerns the degree to which membership is established (see the earlier discussion illustrated on the case of the ESS); the more that this is the case, the greater the possibilities to reach the community of individual researchers with relevant information about products, tools and services available in EOSC/SSHOC infrastructure under development. Linked to this as an indicator of organisational development is the extent to which the membership is informed via push technology (which does require a membership list of some kind to push information to, irrespective of whether that list is exhaustive). The alternative is that exposure to relevant information is dependent on active information-seeking by individual members, which is a much less certain route for information dissemination, for reasons to be discussed in section 4. Yet another aspect of organisational development consists of the extensiveness and density of the network in which a user community is embedded itself, and its centrality in that network. Such networks exist of (organised) user communities, research facilitators (e.g., research councils and other funding bodies), disseminators (commercial and non-commercial publishers), curators and archival institutions, etc.

All in all, the more strongly user communities are developed, the more that some of the necessary conditions for engagement with and contributing to SSHOC are fulfilled. The user communities that are included in the SSHOC consortium are –unsurprisingly– organisationally highly developed, but other user communities exist outside the consortium that are of similar calibre, or that are actively organising themselves to similar standards. Such communities in particular offer the greatest opportunities for SSHOC to attain its aims for a larger part of the entire domain of the social sciences and humanities than what is covered by the consortium partners themselves.

The communities to which SSHOC aims to cater its infrastructure consist of individual people, which are both those who populate the communities' organisations, as their other members –the researchers, scientists, analysts who may be working for academic institutions, for business organisations, for governmental or other public bodies, for civil society organisations, for media, etc. Again, challenges and obstacles for engaging with and contributing to the EOSC/SSHOC infrastructure will be different for different kinds of these individuals. An important distinction at this individual level is thus the following:

- **Leading members** of user communities are those individuals who occupy a role or position in the organisations of user communities. It is irrelevant on what basis they obtained such positions –one can think of individual entrepreneurship, election, co-optation, employment, etc.– as long as their work involves giving direction to the activities of their user communities. Such leading members can be further distinguished on the basis of whether or not their user community is part of the SSHOC consortium.

Leading members of communities that are represented in the consortium will obviously experience many opportunities and comparatively few challenges or obstacles for engaging with and contributing to SSHOC. Doing so is part of the role and position that these leading members occupy.

This does not in and of itself guarantee sufficient awareness to engage productively with the evolving EOSC/SSHOC infrastructure, but the leading members of communities included in the SSHOC consortium have an obvious advantage in these respects. This is very different for leading members of user communities that are not involved in SSHOC. They may experience a variety of challenges and obstacles when it comes to contributing to the EOSC/SSHOC infrastructure of data, products, tools and services. These opportunities and obstacles will be discussed in more detail in sections 4 and 5 of this document.

- **Regular members** of user communities do not occupy leading or organisational positions and are the individual researchers who collectively, and across user communities constitute the research capacity of the social sciences and humanities. They can be found in academia, but also in business, public sector organisations, (I)NGOs, cause groups, etc. Regular members may experience opportunities to contribute to SSHOC but will almost unavoidably also experience challenges and obstacles to do so. The nature and importance of these will, again, be different dependent on whether their user community is or is not directly involved in SSHOC. A further discussion of the kinds of obstacles involved will be included in sections 4 and 5 of this document. There we will refer to additional distinctions between regular members that may be relevant, such as the number of user communities in which they are actively involved; the stage of their research career; the security of their position. All of these impinge on the opportunities for and obstacles to contributing to the infrastructure the consortium develops.

To conclude this section, we reiterate its main objective. The long-term success of SSHOC is to some extent dependent on the contributions that user communities in the social sciences and humanities make to the evolving EOSC/SSHOC infrastructure. Some of these user communities are part of the SSHOC consortium, but many other ones are not, and the consortium aims to be also relevant for them and enjoy contributions from them. As opportunities for and obstacles to such contributions can be expected to vary for different user communities, we elaborated in this section the concept of user communities and distinguished different kinds. We extended such differentiation to members of user communities to avoid unwarranted reification and anthropomorphication of the communities under discussion.

The next sections will discuss, respectively, kinds of engagement, uses and contributions that could be made by user communities and their members, and subsequently, different kinds of opportunities and obstacles to such activities. Those sections provide, together with the current one, the building blocks for the final section in which different kinds of challenges and obstacles are linked to different kinds of activities for different kinds of user communities and their members.

3. Varieties of uses, engagement and contributions

Opportunities for and obstacles to contributing to the EOSC/SSHOC infrastructure do not only vary for different kinds of user communities and end-users, but also for different ways in which they can do so. This section therefore discusses different forms that such engagement and contributions can take.

In very broad terms we may distinguish between engagement with, active use of, and making contributions to the developing EOSC/SSHOC infrastructure. In a non-deterministic way these three forms of interaction are likely to be cumulative, implying that, in the order in which they are mentioned, each constitutes a facilitating condition for the next (thus engagement facilitating active usage, and that in turn facilitating making contributions).

1. **Engaging with** SSHOC (without further use of contribution) can best be conceived in terms of being interested in SSHOC and what it has to offer. This can take a variety of forms, including willingness to absorb and process information about SSHOC that one is exposed to; actively pursuing information and evaluating it in relation to one's own research; discussing the relevance of the EOSC/SSHOC infrastructure with colleagues and at meetings; and so on. Indeed, such engagement can be regarded as a precondition for more involved ways to relate to it. These forms of engagement can be undertaken by user communities (via their leading members –see the discussion in the previous section) as well as by regular members of user communities. For user communities such engagement is a prerequisite for undertaking activities such as disseminating and/or recommending information about SSHOC to their members; and for deciding whether and in what form to link user community activities to the infrastructure. For individual end-users (regular members of user communities) this engagement can be expected to be particularly attuned to their individual research needs and activities.
2. **Using SSHOC** involves actively drawing on the infrastructure to acquire data, tools, other products and services to be used in research activities. For obvious reasons, using SSHOC requires a minimum level of engaging with it, at least to the degree of being aware of the existence of the products or services one could use, and of how to access these. For user communities, usage may also involve drawing on standards and protocols developed within the consortium; or drawing on services that are available in the infrastructure for curating, archiving and disseminating data and other products generated by the user community; and so on. For individual end-users (regular members of user communities) usage will consist mainly of obtaining data, tools and services to be employed in their own research.
3. **Contributing to SSHOC** consists of making one's own research, its data, tools, outcomes etc. available to other researchers and other research communities by way of the EOSC/SSHOC infrastructure. Here, again, and for obvious reasons, contributing can only be realised on the basis of sufficient awareness of possibilities to contribute, and the technical competence to do so. In other words, it requires a minimum level of pre-existing engagement with SSHOC. Contributing by user communities can be expected to exist often in the form of making available, through the facilities offered by the consortium and its partners, primary data, metadata and other information of relevance for researchers in the social sciences and humanities that take an interest in these products. Moreover, such contributions may also involve data created by user community organisations by way of re-using or re-purposing primary data (one can think of, for example, harmonised data created from a variety of originally separate datasets that are not directly comparable). Yet other forms in which user communities may contribute is in the form of providing analytical tools and procedures, and so on. Such contributions can be made available via curating and archiving facilities provided by some of the SSHOC consortium partners (e.g., GESIS), or via the

Marketplace to be developed by the consortium.

Contributing to SSHOC by individual researchers (regular members of user communities) can be expected to be less frequent and on a smaller scale than by (the organisations of) user communities. Yet, such contributions are, in principle, not only quite feasible, but also extremely important in data life-cycles (see SSHOC Consortium Proposal, and also <https://ddialliance.org/training/why-use-ddi>). In the course of their empirical work, many individual researchers develop new data (e.g., generated via re-use and re-purposing), new analytical procedures and new tools that are of potential relevance to others in their own and in other user communities. These products of their work can be made available via the EOSC/SSHOC infrastructure in the same way as products generated by user-community organisations can be. Yet, for a variety of reasons (to be discussed in section 4) it is unlikely that such contributions from individual researcher will be voluminous in the short run, were it only because understanding what DDI is about requires already quite a bit of technical competence which is not universally distributed across the multiple SSH communities.

4. Varieties of obstacles to engagement with, use of and contributing to SSHOC, and their background

A variety of almost self-evident obstacles exist that may prevent user communities and their researchers obtaining the benefits that the SSHOC infrastructure can potentially impart on them. Being **unaware** of SSHOC – and of the aspirations that have given rise to EOSC, and thus to SSHOC as well – and its products and services is one of these. Even when being aware of its existence, being **unable** to utilise it is a second obstacle. More of such obstacles can be listed but doing so is of limited value as long as the factors leading to these obstacles remain unelaborated. Lack of awareness, for example, can originate from a number of factors, such as never having been exposed to relevant information; but it could also come about from never having had the time to absorb, let alone act upon such information; and so on. Thus, different circumstances may lead to what seems in first instance to be the same impediment (in this example lack of awareness of what SSHOC may potentially offer to one's research), while more profound, and different problems hide behind this first appearance. In order to be effective, any activity originating from SSHOC to ameliorate obstacles to the use of what it offers to user communities and their researchers needs to be attuned to these more distant factors, all of which relate to the ways in which individual researchers and user communities operate. We start therefore with some theoretical notions about how researchers behave, which will help to clarify typical motivations, attitudes, preferences, and behaviours that impede or promote the active taking up of relevant possibilities to be offered by EOSC/SSHOC. Such ideas are by necessity focussed on individuals. After all, only individuals act (or fail to do so) in particular ways. This individualistic perspective is, however, also indispensable for the wider question about the extent that user communities will engage with, make use of, and contribute to SSHOC. After all, user communities only 'act' via individuals, who hold positions that entitle them to act for and on behalf of the community they represent (see also the discussion in section 2 about the distinction between 'leading' and 'regular' members of user communities). Obviously, not only characteristics of leading members, but also characteristics of their community matter for what their leading members do (such as the magnitude,

cohesiveness, internal diversity of user communities, and so on), but in last instance only individuals act, while organisations or user-communities can only do so by way of the individuals that act on their behalf.

In the following we will discuss two different (but complementary rather than antagonistic) perspectives for reflecting on how researchers can be expected to behave: a bounded rationality perspective, and a socialisation perspective.

4.1 The perspective of bounded rationality

When reflecting on obstacles for active engagement with SSHOC the perspective of rationality assumes, first of all, that researchers operate on a rational basis. This implies that they (implicitly) assess costs and benefits of different actions that are open to them, and that they choose the course of action that is most profitable on the basis of a comparison of costs and benefits. To be useful, this idea of weighing costs and benefits requires a specification of the goals that researchers pursue, as what would be a 'cost' for the attainment of some goals, could be a 'benefit' for other ones. We assume here that researchers are motivated by two goals (that may occasionally conflict with one another; however, the occurrence of such conflicts is not pertinent to the discussion here): 1) making a living from their work in research (while maintaining a reasonable work-life balance); and 2) contribute new knowledge, insight and understanding in the subject area of their research to their research community and the world at large. The first of these goals reflects that in the current day and age scientific work is rarely self-funded, but mainly pursued in positions of employment. Therefore, what employers demand from the researchers they employ—in terms of goals, priorities and outputs—is of paramount importance for these researchers to safeguard their livelihood and for career-success, and these demands will structure their behaviour. The second of these goals implies that researchers can be expected to choose courses of action that enhance the likelihood of attaining new knowledge, insight and understanding, and the likelihood of having these contributions acknowledged and validated via publications.

The general perspective that researchers' behaviour is guided by rational comparisons of costs and benefits must, however, be somewhat qualified, even for those working in scientific research. For reasons amply expounded by Simon (1947, 1982)³ a perspective of 'bounded' rationality is more useful. In this perspective, rationality is 'bounded' by available information, and by the capacity of actors to process that information. Thus, awareness of options (or the lack thereof) becomes an important aspect of choosing a particular course of action, as well as *perceived* costs and benefits. This leads to the following factors that will impinge on (often implicit) decisions to actively engage with EOSC/SSHOC once it has become available as an infrastructure that user communities and individual researchers can engage with:

³ Herbert A. Simon, (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, New York, NY: Macmillan. Simon systematised his perspectives on the topic later in Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.

- **Awareness of the existence of SSHOC**

Knowledge of the existence of SSHOC and of what it entails is an obvious precondition for any active engagement with this infrastructure. This is probably unproblematic for user communities whose organisations and institutions are part of the SSHOC consortium, although it cannot be taken for granted for all individual researchers who are members of these user communities. For individual researchers much depends on the extent to which their research is contained within a single user community and whether that community is included in, or closely linked to the SSHOC consortium (so that exposure to relevant information will be more frequent and more endorsed by relevant others). Alternatively, when their research relates to multiple user communities, of which some are not included in the SSHOC consortium, exposure to what SSHOC may offer, and receptivity to that exposure may be much smaller.

For user communities that are not represented in the SSHOC Consortium, knowledge or awareness of the existence of SSHOC is inherently more problematic to achieve. Mere dissemination of information is unlikely to be sufficient given the ever-growing number of organisations, institutions and initiatives that all compete for the attention (and the allegiance) of active researchers and their communities. To be effective, information must connect to as-yet-unresolved needs of the intended recipients (which links up to the discussion of potential benefits, below).

- **Perceived costs of engaging with SSHOC**

The higher the perceived costs of engaging with SSHOC, the less likely that engagement will follow. Costs are of various kinds, including monetary expense, human resources (time and attention), and opportunity costs.

One kind of cost is monetary; using SSHOC effectively requires **funds**. Contributing data or methods to any platform on EOSC is not cost-free in terms of human resources (and thus: funds). Contributing already existing data and metadata, for example, requires resources for hiring and training research assistants.

A different kind of cost consists of **time** and **attention**, to be invested for learning what EOSC/SSHOC may have to offer and how to make use of or contribute to those possibilities (this can in principle be monetised, but often it is not).

The *perception* of how much time is required for learning and acquiring necessary skills is not merely a reflection of whatever is communicated about this from SSHOC. It is primarily a personal assessment of necessary follow-up activities after initial training/learning to acquire a productive level of skills. This assessment is influenced by many factors, including experiences from other occasions to achieve similar competencies, and by information from relevant peers. If the required time investment (as perceived) is larger than the amount of time available (given other tasks that have to be fulfilled), no investment will be made in learning/training. The implication of this is that any communications from SSHOC about time to be invested in learning/training should address these perceptions as explicitly and realistically as possible.

In addition to monetary costs and human resources, **opportunity costs** are important drivers of the decision whether or not to engage with the EOSC/SSHOC infrastructure. These costs consist of other

activities that could have been done instead, but that must be declined or deferred if one invests in learning/training and engagement with SSHOC. In view of the universal tendency of *temporal discounting*⁴ any investment in learning/training is at risk to be deferred when other urgent current tasks must be attended to. After all, learning/training leads to costs that are to be incurred immediately, while any benefits will accrue only in the future so that their value will be discounted at the moment of deciding whether to engage in learning/training.

This problem of temporal discounting is compounded when potentially rivaling ways to use available time are perceived to be more urgent in terms of employer demands or career benefits. In other words, when dealing with highly pressured researchers ('publish or perish') opportunity costs may lead to the learning/training of new skills being relegated to the realm of good intentions. To mitigate this, training opportunities could be integrated in events that intended trainees are likely to attend anyway (e.g., during conferences, or as a pre-conference event), thus reducing opportunity costs.

- **Perceived benefits of engaging with SSHOC**

The larger the perceived benefits of engaging with the SHOC infrastructure, the more likely it is that researchers will indeed do so. Yet, to be effective as a driver of actual engagement and usage, perceived benefits must be 1) *sufficiently specific* (i.e., not in general terms, but in terms of the specific projects a researcher works on); 2) *close enough in the near future* to avoid being heavily discounted; and 3) *address not-yet-met needs* of researchers that are unlikely to be met in other ways. With respect to this last aspect, the perceived benefits will improve when researchers are made to understand that making use of EOSC/SSHOC builds upon competencies they already may have acquired (e.g., with respect to archiving of data).

- **Perceived net benefit or cost**

The rational choice perspective, including the bounded rationality version thereof, implies that specific actions will only be undertaken if the (perceived) benefits exceed the (perceived) costs. To draw user communities and individual researchers to SSHOC it is thus not enough that potential users recognise that they would benefit from utilising the EOSC/SSHOC infrastructure. They must expect that this benefit outweighs the costs. Moreover, because costs include opportunity costs (as discussed earlier), choosing a particular action (in our case: engaging with and utilising or contributing products and services via the EOSC/SSHOC infrastructure) requires that the net gain to be derived from it should be not only positive, but *larger* than the net gain that could be derived from other, competing, courses of action. For researchers, the alternative courses of action that are most relevant in this context (and that obviously also include non-negotiable demands on their time made by their employers) are their substantive research on the one hand, and, on the other hand, investing time and effort in acquiring new skills and competencies other than those necessary to productively exploit EOSC/SSHOC. Their substantive research is particularly important, as it not only serves researchers' goal of generating new

⁴ Frederick, Shane; Loewenstein, George; O'Donoghue, Ted (2002). "Time Discounting and Time Preference: A Critical Review". *Journal of Economic Literature*. 40 (2): 351–401.

knowledge, but also because it often contributes to career advancement. Producing a research-based book, article or paper is for many researchers more rewarding –intrinsicly and qua career advancement– than acquiring the skills and competencies necessary to make use of an infrastructure. The main exception is when those skills and competencies are perceived to plausibly help in overcoming research problems that otherwise obstruct substantive progress. The other ‘competitor’ for time necessary to acquire skills and competencies required for making productive use of the EOSC/SSHOC infrastructure exists in the form of training to acquire other skills and competencies, including methodological training, the pursuit of ‘impact’,⁵ and so on.

All in all, the perspective of bounded rationality helps to identify a set of factors and conditions that make it more, or conversely less likely for researchers to exploit what EOSC/SSHOC may offer them. Awareness of these factors and conditions may help SSHOC in its outreach and communications towards user communities and individual end-users, particularly towards those that are not closely associated with the SSHOC consortium or its partners.

4.2 The socialisation perspective

People are not born researchers, but develop as such through various processes, two of which are of particular relevance for the current discussion. The first of these is about *substantive training*; this pertains to the acquisition of *knowledge and skills* that are the defining characteristics of contemporary conceptions of professional researchers in a specific field of study. The second process involved in the development of researchers is socialisation into the *norms and values* of a scientific community.⁶ These two processes are intimately interlinked, particularly for user communities, as socialisation involves also values related to learning and training in general, and in more specific ways in relation to what has to be learned. Moreover, both training and socialisation extend to the entire research career and are not restricted to a phase of initiation into a research community. The perspective of socialisation into becoming a researcher and a member of a research community is therefore of direct relevance when reflecting on obstacles and opportunities for researchers and user communities in the social sciences and humanities to actively engage with the EOSC/SSHOC infrastructure that is under development.

Norms and values that are socialised include explicit or implicit specifications of knowledge and skills that members of research- and user-communities should possess. For prospective professional researchers (i.e.,

⁵ The term impact is here used in a meaning that is of particular relevance for research undertaken in academe and that would involve ‘an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia’ (cf., <https://re.ukri.org/research/ref-impact/>).

⁶ Without elaborating this further (as doing so would go far beyond the remit of this document) it has to be emphasised that socialisation is never homogeneous. Within research communities, diversity exists about norms and values that pertain to the role of professional researcher (although less so the more the norms and values in question define the community in question), and the internalisation of such norms and values is invariably personalised by researchers themselves.

many PhD students and postdocs) these specifications are embedded in formal and informal training and mentoring. By who and in which form these specifications are shaped varies greatly across countries, disciplines and more fine-grained research- and user communities. Most often they are formally specified by degree-awarding institutions of higher education or in certification requirements, in ways which are, in turn, usually informed by disciplinary professional organisations, or by research councils and other funders, and by national regulation. However, these requirements are almost unavoidably rather generic, and will therefore generally not specify the requirement of competency to make use of and contribute to the specific products and services to be developed in the EOSC/SSHOC context. Yet, the provision by SSHOC of high-quality training materials may help in certain instances to have such training included in the implementation (operationalisation) of more generic requirements by relevant faculties and departments of degree-awarding institutions. Not only knowledge and skills are acquired through socialisation processes, but also attitudes and orientations; in the context of this document particularly attitudes relating to principles of FAIR data (findable, accessible, interoperable and re-usable) and open data practice. The promotion of such attitudes may be part of formal training but will in any case also depend to a large degree on informal processes of socialisation.⁷

More informal and smaller-scale processes of mentoring, collegial interaction and peer-reviewing are crucial in researchers' socialisation into members of user-communities that focus on specialised research areas, on particular kinds of empirical data, or on specific analytical approaches. The user communities included in the SSHOC consortium are of such kinds. Socialisation into relevant norms of competency and behaviour occurs in such communities via the informal mechanisms referred to earlier, and by exposure to productive use of such competencies by other members of the community etc. How effective such forms of socialisation are depends on factors such as the extent frequency and visibility with which elements of the EOSC/SSHOC infrastructure are actually used in a research community; the esteem that such usage generates in the community; the density of relevant interactions between the members of a user community; and the centrality of these communities in the actual research networks of individual researchers.⁸

Because of their natural socialising influence, user communities offer an ideal context for providing training into the use of the EOSC/SSHOC infrastructure for their individual members. The relevance of user communities in this respect is enhanced by the possibility to link training to specific research challenges that their members

⁷ Although somewhat outside the socialisation perspective, the development of such attitudes may also be stimulated by – as yet virtually non-existing – incentive systems that rewards those that adhere to the FAIR and open data agenda (i.e., recognition of such adherence in promotions, or through citations).

⁸ It must be kept in mind that informal forms of socialisation deriving from mentoring, collegial interaction and peer-reviewing may, under certain circumstances, *discourage* engagement with the products, services and research opportunities offered by the EOSC/SSHOC infrastructure. Mentors expressing that those under their tutelage should not let themselves be distracted by all these opportunities for training, peers expressing that time spent on such engagement has not solved their research problems, and similar encounters will have socialising effects, although in a discouraging direction.

encounter in actual practice (and thus to connect to as-yet-unresolved needs, as discussed earlier in the context of bounded rationality).

Informal socialisation is, moreover, of particular importance for the cultivation of perceptions of costs and benefits that underly the bounded rationality perspective discussed earlier. These perceptions are crucial in the stage of deciding whether or not to embark on a particular kind of action (i.e., of whether to make use of, or contribute to products and services in the EOSC/SSHOC infrastructure). To the extent that such perceptions cannot yet be based on personal experiences, they will be derived from information, suggestions, and hints from relevant others.⁹

Socialisation into competency with new infrastructures (such as being developed by the EOSC and the SSHOC consortium), into supportive attitudes and orientations, and into perceptions of low costs and high potential benefits will greatly influence the potential of these infrastructures being actively used by newly entering cohorts of scholars in the social sciences and humanities and thus, over time, by increasingly more researchers. But socialisation does not end after researchers' entry in research careers. Established members of research communities are, in principle, also receptive to socialising influences, although for them it will even be more important than for new entrants that new tools and infrastructures will meet as-yet-unresolved needs, as they may otherwise find established routines and ways of working sufficient for their needs.

4.3 Obstacles identified by the bounded rationality and socialisation perspectives

The discussion above suggests that, in the abstract, the following kinds of obstacles can stand in the way of users and user communities engaging with, making use of, and contributing to the infrastructure of EOSC/SSHOC:

- **Unfamiliarity:** insufficient, or insufficiently detailed knowledge and awareness of the existence of the infrastructure and of what it may offer;
- **Expense:** a perception of limited relevance, or of poor balance of costs/benefits, which results in the absence of any strong inclination to engage with the EOSC/SSHOC infrastructure;
- **Low priority:** a perception of high opportunity costs, which may stand in the way of engaging with the infrastructure, even in the presence of a strong perception of relevance and of a good costs/benefit balance;

⁹ Following up from footnote 7, personal and institutional incentives may well be indispensable as additional stimuli for developing desired practice. Personal incentives relate to recognition, remuneration and promotion. Institutional incentives may relate to EOSC accreditation for research institutions that achieve certain levels of contribution to EOSC (similar to, e.g., Athena Swan accreditation for recognising Universities that successfully commit to equality in the workplace, see <https://www.ecu.ac.uk/equality-charters/athena-swan/>).

- **Inability:** insufficient competence and skills for productively engaging (in whatever way) with the EISC/SSHOC infrastructure, in spite of a general perception of relevance and of a good costs/benefit balance.

5. Differential importance of obstacles for different kinds of users and user-communities

As discussed in section 2, important differences exist between user-communities and individual users that will result in variations of the relevance of the potential obstacles discussed in section 4.

The two most important distinctions referred to in section 2 were

- a. whether or not user-communities are or are not included in the SSHOC consortium, or linked to it in a clear way; and
- b. whether individual users are *leading* members of user-communities –who occupy positions that require them to act for and on behalf of a user-community– or *regular* members –whose most immediate concerns focus on their research careers and the substantive research in which they are engaged.

These two dichotomies give rise to a fourfold categorisation. One of these four categories is least likely to experience the obstacles identified in section 4: the *leading members of user-communities included in or clearly linked to SSHOC*. These user-communities and their leading members are already committed to the aspirations of SSHOC (and EOSC), and they are facilitated by obtaining resources from the consortium that enable the realisation of those aspirations. Yet, anecdotal evidence suggests that even among this category the obstacles identified in section 4.3 are more common than one might expect.

This is different for *leading members of user-communities* –and thus for those communities as such– that are *not included or clearly linked to SSHOC*. For these, unfamiliarity with the developing EOSC/SSHOC infrastructure may be a very real obstacle.¹⁰ Additionally, expense (including money, time and attention) may be an obstacle, in particular for user-communities that have not (yet) achieved a stable foundation of resources. Yet, to the

¹⁰ This obstacle is more plausible than one might perhaps expect, as demonstrated by the example of the emerging MEDem consortium (Monitoring Electoral Democracy), that represents various user-communities in the broad field of studies of electoral processes and political representation. The founding documents of this consortium, that aspires to ESFRI-status, demonstrate its unawareness of the existence of EOSC/SSHOC, in spite of MEDem's linkages to institutions that are part of the SSHOC consortium. A similar observation can be made with respect to the user-community of ethnic and migration studies (as organised in COST Action 16111 Ethmigsurveydata).

extent that user-communities not included in the SSHOC consortium are sufficiently well organised,¹¹ it can be expected that the major obstacle for their engagement with the EOSC/SSHOC infrastructure is unfamiliarity. However, for user-communities not included in or linked to SSHOC that are not yet strongly organised –i.e., that do not have established organisations, and thus also no clearly identifiable and acting leading members– all the obstacles identified earlier are likely to be present and to stand in the way of engaging (in whatever way) with the evolving infrastructure.

As discussed in section 2, user communities as such do not act, engage with, make use of, or contribute to the EOSC/SSHOC infrastructure. Only individuals can do so, and many of these individuals are *regular* members of user communities –researchers whose main concerns are the pursuit of research in their respective fields and securing a livelihood as a professional researcher. For these regular members of user communities, all obstacles discussed above are potentially of relevance, albeit with some variations depending on other factors.

Regular members of user-communities that are included in or clearly linked to SSHOC will have a fair chance of being cognisant of the developing EOSC/SSHOC infrastructure and its products and services, because of communications, information and stimulus from the organisations of their user-community. Without such support from organisations of user-communities, certain activities could hardly be undertaken at all by their regular members. The receptiveness of regular members to such activities, however, is likely to depend on the extent to which their research is mainly located within a single user-community or spread over several. Being part of multiple user-communities will diminish the relevance of and attention given to communications from each of these.

More active use of products and services, or contributing to those via SSHOC by regular members of user-communities (both those included in or connected to the consortium, as those who are not) is more likely to be restricted by the kind of obstacles identified in the previous section, with one prominent exception. To the extent that already existing facilities for data dissemination, curation, and archiving are incorporated in the EOSC/SSHOC infrastructure (e.g., GESIS, Liber, etc.), some regular members of user-communities will make use of the EOSC/SSHOC infrastructure merely by continuing to do what they already did previously, namely to use these facilities when they require the data or tools for their own research. But even in this respect it must be noted that large segments of SSH research communities are – even nowadays – deeply unaware of the existence these various infrastructural organisations. The extent to which regular members will in due course make use of other services to be provided by the infrastructure (e.g., via its Marketplace) will depend highly on familiarity with their existence (more likely for those in user-communities included in or linked to the

¹¹ Strength of organisation manifests itself in the existence of organisations that aim to represent and act for user-communities, that are seen as legitimate in that respect within those communities, and that have sufficient resources to allow remuneration for leadership positions (thereby reducing the risk that acting for the community in question has to compete for attention and priority with other demands on the time and attention of the individuals involved).

consortium), on acquired skills that make this possible and, on the priority given to acquiring and applying such familiarity and ability.

The above implies that perceived opportunity costs can be expected to be of paramount importance for use of products and services provided by the EOSC/SSHOC infrastructure. Members of user-communities that are under pressure to deliver tangible research products and to secure a position as professional researcher (i.e., young scholars; scholars without tenure; researchers in organisations that reorganise due to financial pressure) will be likely to postpone engaging with the EOSC/SSHOC infrastructure until less-pressured moments. This risk is particularly high with respect to contributing data –both primary data and data created through re-purposing and re-using of already existing material– and tools developed in the course of their individual research.¹²

A summary of the relevance of obstacles that were earlier identified for different kinds of user-communities and their members, classified as low (L), medium (M) or high (H) is provided in Table 1.

¹² The specific problem that presents itself here for regular members of user-communities is that costs of properly preparing and depositing such materials for possible use by other researchers are usually non-negligible, while benefits in the form of rewards for one's career are often absent or near-absent. The problem may be exacerbated by researchers' perception that they have already contributed to these processes by having published their work in peer-reviewed outlets, even when these do not operate on the basis of open-data policies (see, e.g., the classification of scientific journal's policies in this respect in <https://cos.io/blog/landscape-open-data-policies/>).

TABLE 1 RELEVANCE OF DIFFERENT KINDS OF CHALLENGES/OBSTACLES FOR DIFFERENT KINDS OF USER COMMUNITIES AND DIFFERENT KINDS OF THEIR MEMBERS

	Unfamiliarity	Expense (money, time, attention etc.)	Low priority (opportunity costs)	Inability (required skills, training)
Leading members of user-communities in or linked to SSHOC	L	L	L	L
Regular members of user-communities in or linked to SSHOC				
Making use of products and services offered by EOSC/SSHOC infrastructure*	L/M	M/H	L/M	M/H
Contributing products and services**	M/H	H	H	M/H
Leading members of user-communities not in or linked to SSHOC				
Making use of products and services offered by EOSC/SSHOC infrastructure	M/H	M/H	M/H	L/M
Contributing products and services	M/H	M/H	M/H	M
Regular members of user-communities not in or linked to SSHOC				
Making use of products and services offered by EOSC/SSHOC infrastructure	H	M/H	M	M/H
Contributing products and services	H	H	H	M/H

* Lower relevance for uses that involve pre-existing facilities that are incorporated into the EOSC/SSHOC infrastructure (e.g., obtaining data from data archives); medium or high for using newly developed facilities of the infrastructure

** Medium vs. high dependent on importance of the user-community for regular members, and on direction of influences from peers and mentors

List of Tables

[Table 1: Relevance of different kinds of challenges/obstacles for different kinds of user communities and different kinds of their members](#) page 23