Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782 Start Date of Project: 01/01/2019 Duration: 40 months

Deliverable 7.1 System Specification - SSH Open Marketplace

Dissemination Level	PU
Due Date of Deliverable	30/09/2019, September 2019 (M9)
Actual Submission Date	02/10/2019
Work Package	WP7 Creating the SSH Open Marketplace
Task	T7.1
Type	Report
Approval Status	Not approved yet
Version	V1.0
Number of Pages	p.1 – p.62

Abstract:

The system specification report delivered by T7.1 describes the user requirements identified, the conceptual model of the SSH Open Marketplace and its relation to SSHOC reference ontology, as well as the overall system architecture with a description of key internal and external components and their dependencies.

SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.0	16/01/2019	First draft	Matej Ďurčo (OEAW) & Frank Fischer (DARIAH)
0.1	19/02/2019	Input chapter 1	Yoann Moranville (DARIAH)
0.3	03/06/2019	Input Data Model and System Architecture	Matej Ďurčo (OEAW)
0.4	24/06/2019	Final chapter structure	Philipp Wieder (UGOE), all
0.5	19/07/2019	Input Data model	Matej Ďurčo (OEAW)
0.6	26/07/2019	State of the art, user requirements & further development parts	Laure Barbot (DARIAH), Clara Petitfils (CNRS)
0.7	18/08/2019	Data model & system architecture	Matej Ďurčo (OEAW)
0.8	19/08/2019	Adjustments and suggestions Data model & system architecture	Sotiris Karampatakis (SWC)
0.9	22/08/2019	Adjustments and suggestions all chapters, Implementation	Tomasz Parkoła (PSNC)
0.10	03/09/2019	Substantial rework of data model, structured description of individual classes	Matej Durco (OEAW)
0.11	03/09/2019	Adjustments and suggestions all chapters	Laure Barbot (DARIAH), Clara Petitfils (CNRS), Yoann Moranville (DARIAH), Frank Fischer (DARIAH)
0.12	09/09/2019	Introduction, Conclusion & executive summary	Philipp Wieder (UGOE)
0.13	25/09/2019	Peer review	Eric Balster & Maurice Martens (CentERdata), Triet Ho Anh Doan (GWDG), Nina Bakanova (CESSDA)
1	27/09/2019	Address peer review comments & final version	Laure Barbot (DARIAH), Clara Petitfils (CNRS), Matej Ďurčo & Klaus Illmayer (OEAW), Sotiris Karampatakis (SWC)

Author List

Organisation	Name	Contact Information
CNRS (Huma-Num)	Clara Petitfils	clara.petitfils@huma-num.fr
DARIAH ERIC	Frank Fischer Laure Barbot Yoann Moranville	frank.fischer@dariah.eu laure.barbot@dariah.eu yoann.moranville@dariah.eu
OEAW	Matej Ďurčo Klaus Illmayer	Matej.Durco@oeaw.ac.at Klaus.Illmayer@oeaw.ac.at
PSNC	Tomasz Parkoła	tparkola@man.poznan.pl
UGOE	Philipp Wieder	philipp.wieder@gwdg.de
SWC	Sotiris Karampatakis	sotiris.karampatakis@semantic-web.com

Contributors: Chryssoula Bekiari (FORTH), Stefan Buddenbohm (UGOE), Suzanne Dumouchel (CNRS), Martin Kaltenböck (SWC), Carsten Thiel (CESSDA), Justyna Wytrążek (PSNC).

Reviewers: Eric Balster & Maurice Martens (CentERdata), Triet Ho Anh Doan (GWDG), Nina Bakanova (CESSDA).

Executive Summary

This document delivers the results of Task 7.1 of the Social Sciences & Humanities Open Cloud project funded by the European Commission under Grand #823782. Its main purpose is the specification of the SSH Open Marketplace (SSHOC MP) in terms of service requirements, data model, and system architecture and design.

The Social Sciences & Humanities communities are in an urgent need for a place to gather and exchange information about their tools, services, and datasets. Although plenty of project websites, service registries, and data repositories exist, the lack of a central place integrating these assets and offering domain-relevant means to enrich them and communicate is evident. This place is the SSHOC Marketplace.

The approach towards the system specification is based on an extensive requirements engineering process. First and foremost, user requirements have been gathered through questionnaires. The results have been then prioritised based on the user feedback and the experience of the SSHOC project partners. Based on the requirements and thorough state-of-the-art analysis, a data model and the system design have been developed. In order to do so, and by taking into account as much previous work from other European projects as possible, the integration with the EOSC infrastructure has been a primary concern at every step taken.

The system specification is now the starting point for the development of the SSHOC MP and also a communication instrument within the project and externally. Over the course of the agile development of the Marketplace, the system specification will also be evolving and contributing to a growing number of SSHOC outcomes.

Abbreviations and Acronyms

API	Application Programming Interface
CESSDA	Consortium of European Social Science Data Archives
CIDOC-CRM	Comité International pour la Documentation – Conceptual Reference Model
CLARIN	Common Language Resources and Technology Infrastructure
CNRS	(French) National Centre for Scientific Research
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DESIR	DARIAH ERIC Sustainability Refined
DIMPO	Digital Methods and Practices Observatory
DiRT	Digital Research Tools
EGI	European Grid Initiative
EOSC	European Open Science Cloud
ERC	European Research Council
ERIC	European Research Infrastructure Consortium
ESS	European Social Survey
EUDAT	European Data Infrastructure
EURISE	European Research Infrastructure Software Engineers
FAIR	Findability, Accessibility, Interoperability, and Reusability
FORTH	Foundation for Research and Technology - Hellas
GDPR	General Data Protection Regulation
LIBER	Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries
OEAW	Austrian Academy of Sciences
OP	Onboarding Process
OSG	Open Science Graph
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
PID	Persistent Identifier
PSNC	Poznań Supercomputing and Networking Center
RDA	Research Data Alliance
RI	Research Infrastructure
ROHub	Research Object Hub Platform
SERISS	Synergies for Europe's Research Infrastructures in the Social Sciences
SHARE	Survey of Health, Ageing and Retirement in Europe
SMS	Service Management System
SPM	Service Portfolio Management
SSH	Social Sciences and Humanities
SSHOC	Social Sciences and Humanities Open Cloud
SSHOC MP	Social Sciences and Humanities Open Marketplace
SSK	Standardization Survival Kit
SWC	Semantic Web Company
TaDIRAH	Taxonomy of Digital Research Activities in the Humanities
TAPoR	Text Analysis Portal for Research
TERESAH	Tools E-Registry for E-Social science, Arts and Humanities
TRIPLE	Targeting Researchers through Innovative Practices and Multilingual Exploration
UCD	User Centered Design
UGOE	University of Goettingen
VLO	Virtual Language Observatory
WP	Work Package

Table of Contents

List of Figures

List of Tables

Introduction

The Social Sciences & Humanities Open Cloud project (SSHOC) aims at providing a cloud infrastructure where data, tools, and training are available and accessible for social sciences and humanities (SSH) users. The goal is to create a cloud ecosystem through the design, development, and maintenance of user-friendly tools and services, covering all aspects of the SSH research data lifecycle. To achieve this, SSHOC will apply a human-centric approach and create links between people, data, services, and training.

SSHOC is a Research & Innovation project within Horizon 2020 contributing to the implementation of the European Open Science Cloud (EOSC[1]). In close collaboration with the EOSC-hub project and with other European projects, SSHOC contributes to the agenda by adding SSH-specific contributions to a growing portfolio of EOSC-related services. The SSH Open Marketplace (SSHOC MP) is an integral part of SSHOC's goal to implement the SSH part of the EOSC. It will thus become part of the EOSC through its integration with its manifestations, such as the EOSC portal.

Further information about the project and its objectives can be found at https://sshopencloud.eu.

Creating the SSH Open Marketplace

Work Package 7 (WP7) of the SSHOC project is dedicated to the creation of the SSHOC MP. The work package is divided into four tasks. Task 7.1 "User requirements, Conceptual Model and System Architecture of the SSH Open Marketplace" focuses on the architecture and the system specification of the SSHOC MP that are at the core of this deliverable.

Concretely, "i*nstead of being just a list of links or database of resources, [the Marketplace] will* ***contextualise*** *and interlink tools, services and datasets offered, with screenshots, tutorials and links to training material, user stories, showcases. It will also encompass community features like user feedback, ratings, the integration of other related channels (...) allow categorisation according to multiple classification schemes, such as TaDiRAH[2] and NeMO[3], and contain qualified links to involved actors – persons and institutions who authored, contributed to, funded or host a given resource.*"[4]

Purpose and content of this deliverable

The main purpose of this deliverable is the provision of the system specification of the SSHOC MP. This specification will then be used to implement it accordingly. Furthermore, it will serve as an instrument to communicate with stakeholders like content providers or the EOSC-hub project and lay the foundations for integration activities.

The deliverable contains the following core content:

- state-of-the-art,

1 SSHOC is one of the five cluster projects funded to develop a domain specific response to the EOSC. For more information, see SSHOC and EOSC blog post here: https://www.sshopencloud.eu/sshoc-eosc , but also the EOSC timeline here: https://www.eosc-portal.eu/about/eosc, and the EOSC strategic implementation plan for more details: https://publications.europa.eu/en/publication-detail/-/publication/78ae5276-ae8e-11e9-9d01-01aa75ed71a1 .
2 http://tadirah.dariah.eu/
3 http://nemo.dcu.gr/
4 SSHOC Grant Agreement, Annex 1 (Part A) p. 55

- user stories,
- system requirements,
- the data model, and
- the system architecture of the SSHOC MP.

The publication of this document already at the end of project month 9 (September 2019) allows synchronising a first full specification early in the project with the development of the marketplace (Task 7.2), thus realising a more agile software development process. Both tasks will share the current status of the development, feedback from users, feature requests, and changing/new user stories through https://gitlab.gwdg.de/sshoc, thus representing the development and actual state of the architecture and system design.

Approach

The SSHOC MP aims directly at users and helps them to **find solutions to enhance their particular research practices**. Instead of just offering a list of tools or services, as many other marketplace-like services do, the SSHOC MPe will try to answer the question "How can I achieve a certain goal?" by offering a **discovery service** for existing solutions and approaches and integrate means to

- **contextualise** them with related information,
- enrich them with feedback and usage information from the **community**, and
- carefully **curating** the information.

Therefore, instead of a "one-stop-shop" or "app-store", you can think of the SSHOC MP as a well-stocked workshop (in its original sense), which has all the necessary tools available, but it's compelling mainly due to the helpful advice of knowledgeable people on how to actually do things. This approach will be developed in particular by following the FAIR principles[5], to improve the findability, accessibility, interoperability and reusability of digital assets.

To achieve this, WP7 has conducted a variety of interviews and derived system requirements from them. Based on those requirements, the underlying data model of the marketplace has been defined, considering well-known standards and best practices from the SSH domain and beyond. In the next step, the requirements have been translated into technical functions, which then led to the actual architecture.

The outcome of the approach and, thus, the overall system specification of the SSHOC MP is delivered through this document.

5 https://www.go-fair.org/fair-principles/

State of the Art

The idea of the SSHOC MP is being developed in a complex multi-faceted and historically grown landscape. The aim of this chapter is to provide the background for its implementation.

Broader EOSC context

As mentioned in the Introduction, the SSHOC MP is a component of the SSH part of the EOSC. EOSC is not yet a mature product, but a moving enterprise under active development. Different initiatives and projects contribute to its shaping[6], and in the following, we try to reflect on some of the concepts and components that have been circulated in various recent working documents[7] of these initiatives.

These documents provide a good overview of the current state of the discussion on how the EOSC architecture will look like and how the EOSC resources[8] will be organised :

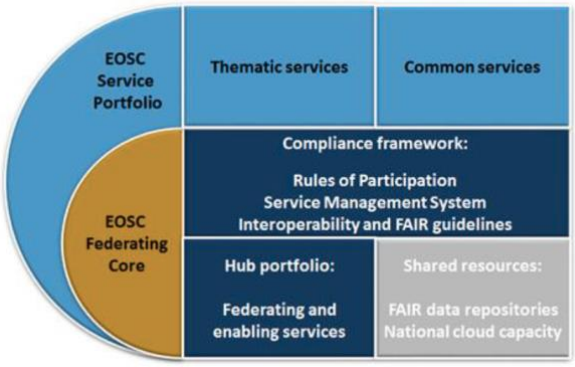

Fig.1 – The EOSC Resources, organised into two portfolios: the EOSC Federating Core (yellow) and the EOSC Service Portfolio (light blue)

This organisation presents services as (one of) the central answer(s) for researchers' needs in terms of support for the data life cycles they are dealing with. The service-oriented approach is nowadays a predominant paradigm and makes sense from the perspective of large e-Infrastructures. For the communities of practice, this means that they need to translate some of the current scholarly practices to fit into this approach. Indeed, if we want the SSHOC Marketplace to include not only services in the narrow sense, but also other assets

6 The EOSC Governance Board, the Executive Board and the Stakeholder Forum are now supported by five Working Groups – Landscape, FAIR, Architecture, Rules of Participation, Sustainability – that "ensure a community-sourced approach to the current challenges of the EOSC" (cf. https://www.eoscsecretariat.eu/eosc-working-groups). The implementation of the EOSC also relies on H2020 projects like EOSCpilot, EOSC-hub, eInfraCentral, EOSC secretariat or the clusters projects of which SSHOC is a part and which aims to connect the ESFRI Infrastructures to the EOSC.

7 Such as the final EOSC-Hub Strategy plan https://documents.egi.eu/public/ShowDocument?docid=3469 and the Briefing Paper – EOSC Federating Core Governance and Sustainability: https://eosc-hub.eu/sites/default/files/EOSC-hub%20Briefing%20Paper%20-%20EOSC%20Federating%20Core%20Governance%20and%20Sustainability%20Public.pdf

8 Based on the EOSC glossary (https://www.eosc-portal.eu/glossary), an EOSC resources represents "any asset made available *(by means of the EOSC system and according to the EOSC Rules of Participation) to EOSC System Users to perform a process useful to deliver value in the context of the EOSC*. EOSC Resources include services, datasets, software, support, training, consultancy or any other asset."

produced by the SSH communities[9] that could be relevant for end-users, we will need to work in-depth to position these other assets in the existing landscape.

The second major aspect is the efforts in context of EOSC for cataloguing the "EOSC resources" to make them findable and accessible. Indeed, a central challenge of such an all-encompassing endeavour is how to enable the user to find what they need, given the enormous scope and heterogeneity of available resources. The two most prominent efforts in "cataloguing EOSC" are the EOSC Marketplace[10] delivered by EOSC-Hub and the eInfraCentral catalogue[11], both launched within the last year. Since a few months, it is also possible to submit a service or a resource directly on the EOSC portal[12], and as cluster projects are currently designing their service offers (partially based on existing services as it will be detailed in the next subsection for the SSH communities), the offer will become broader and more balanced once more new research-oriented solutions will be contributed.

These are not the only relevant cataloguing efforts. Next to the number of discipline-specific discovery solutions, some of which have been in use for many years, and to which we return in the following sub-section, we want to highlight the OpenAIRE initiative. For one, its main catalogue[13] offers information about millions of resources, in this case effectively of different kinds – publications, as well as datasets, or software. Moreover, this information is also available in machine-readable form through an API[14], in a well-documented OpenAIRE Research Graph Data Model[15] – based on CERIF[16], a well-established format for current research information systems (CRIS). These characteristics of OpenAIRE become particularly pertinent in the light of the coordination efforts between OpenAIRE, EOSC-Hub and eInfraCentral, as witnessed among others by co-authored recent 'Common Vision for EOSC' White Paper[17], or the EOSC Portal Concept 2.0 document[18].

We sketch the envisaged position and role of SSHOC MP within the EOSC context in the section Integration with EOSC later in this document.

SSH communities

As a cluster project representing the SSH community, the SSHOC project and the SSHOC MP provide content developed by and/or useful to the SSH researchers and professionals in the domains. The Research Infrastructures and the other partners of the SSHOC project participate in identifying not only the content that will populate this SSHOC MP but also the usability component that fits best with SSH actors' habits and expectations.

To start the work, we identified previous attempts and projects – mainly in the Digital Humanities context – that have helped us to draw the first lines of this SSHOC MP. First, the work done by two non-European projects

9 We will detail in the following sections the kind of assets we would like to include in the SSHOC Marketplace but the description of work of the SSHOC project (Part B, p.10) mentions at least the following: "datasets, tools, and services (…) with screenshots, tutorials and links to training material, user stories, showcases, and other related resources".
10 https://marketplace.eosc-portal.eu/
11 https://www.einfracentral.eu/search
12 https://eosc-portal.eu/for-providers
13 https://explore.openaire.eu/
14 http://api.openaire.eu/
15 https://zenodo.org/record/2643199#.XW-ZEXuxVO9
16 Common European Research Information Format https://www.eurocris.org/cerif/main-features-cerif
17 https://www.openaire.eu/a-common-vision-for-eosc-white-paper
18 https://wiki.eosc-hub.eu/display/EOSC/EOSC+Portal?preview=/34637786/45711867/EOSC%20Portal%20Concept%202.0-v2.2.pdf

can be considered as a good inspiration. The project Bamboo and the DiRT (Digital Research Tools) directory, funded by the Andrew W. Mellon Foundation between 2008 and 2012, present an overview of the challenges and limits that we have to take into consideration starting a collaborative and community-based infrastructure project (Dombrowski, 2014). The work done between DiRT and the TAPoR gateway (Text Analysis Portal for Research)[19] to merge their contents provides one possible answer to sustainability questions that these kinds of projects raise. Such cross-project collaboration is also inspiring for the Marketplace we are working on (Dombrowski/Rockwell, forthcoming; Grant et al., forthcoming). Furthermore, the TERESAH (Tools E-Registry for E-Social science, Arts and Humanities)[20] platform, created under the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project[21] and further developed under the Humanities at Scale project[22] is "a cross-community tools knowledge registry aimed at researchers in the Social Sciences and Humanities" which provides a strong basis for the Marketplace.

In addition to that, the work done under the DESIR project coordinated by DARIAH, and here especially the "D5.4 – Implementation of a centralised helpdesk and marketplace mock-up" (Raciti et al., 2019), presents a design study for the SSHOC MP that played an essential role also for the system specifications presented in this report. The DESIR deliverable offers a detailed comparison of the different existing discovery platforms (TERESAH,[23] TAPoR[24], EGI Marketplace[25], EOSC-hub MP[26], DiRT Directory[27], Humanities Data[28], and ROHUB[29]) and highlights the most common functionalities that existing platforms encompass, like searching, filtering and categorising. This deliverable recommends for example to work on the organisation of the contents of the platforms to allow either Search Results Clustering Engine to automatically organise results of searches into thematic categories, or to address regularly the users by curated presentation of the contents (for example in the form of categories like "tool of the month" or "latest tools"). Finally, the findability aspect and the quality of the content are also presented as key elements to work on.

Another project gave us very interesting inputs. This is the PARTHENOS project[30], and especially the work done under WP4 dedicated to the standardization that has created the Standardization Survival Kit[31], "A collection of research use case scenarios illustrating best practices in Digital Humanities and Heritage research". Based on the idea that "it is necessary to stabilise knowledge on standards and research good practices"[32] in Social Sciences and Humanities, the SSK presents research scenarios as recipes for each given research workflows and accompany researchers at every step of the recipes providing information on the standards that can be followed, as well as examples. The SSK is one of the most relevant existing tools/services to illustrate the workflow, scenario, recipe ideas that we want to develop within the SSHOC MP.

19 http://tapor.ca/pages/about_tapor
20 http://teresah.dariah.eu/
21 https://dasish.eu/ and https://github.com/DASISH/TERESAH
22 http://has.dariah.eu/ and https://github.com/DARIAH-ERIC/TERESAH/
23 http://teresah.dariah.eu/
24 http://tapor.ca/
25 https://marketplace.egi.eu/
26 https://marketplace.eosc-portal.eu/
27 http://dirtdirectory.org/
28 https://humanitiesdata.com/
29 https://www.rohub.org/
30 http://www.parthenos-project.eu/
31 http://ssk.huma-num.fr/#/
32 https://ssk.readthedocs.io/en/latest/0_userDoc.html#the-ssk-a-toolkit-for-humanities-scholars

Furthermore, we should also rely on two important sources of the DARIAH environment, namely, the in-kind contribution tool and the OpenMethods Metablog[33]. The concept of an in-kind contribution is closely related to the DARIAH ERIC and is used to map the national contributions of DARIAH members[34]. An online tool (de Leeuw et al., 2017) has been designed to collect, review and disseminate these contributions that are described thanks to the TaDiRAH taxonomy, and that are also accessible via API. The most relevant contributions could be used to populate the SSHOC MP. On a different note, we should also build on the experience of the OpenMethods Metablog. This platform gathers user experiences with certain methods or tools from the Digital Humanities. With this simple blog-like interface and its light but strong curation process, the OpenMethod Metablog is one of the models we should keep in mind for the SSHOC MP.

In the SSH landscape, several catalogues of data, which are of interest from the point of view of our project, exist and are maintained by different Research Infrastructures. Those aggregators and discovery services already represent sources we can rely on to populate the SSHOC MP. Let's mention the main ones we will harvest/build on: CLARIN VLO[35], DARIAH-DE Collection Registry[36], CESSDA data catalogue[37] that includes SHARE and ESS datasets.

When it comes to Social Sciences, Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS) project[38] needs to be mentioned. This project that brought together Research Infrastructures in Social Sciences addressed key challenges for cross-national data collection and developed common technological platforms as well as shared online tools and resources to offer better coordination between the different stakeholders of the communities involved. This project developed seven tools[39] and delivered training materials[40] to support different stages of the survey life cycle. These outputs were not only a starting point for the SSHOC project since some of those tools are being extended within the WP4 but will also become content for the SSHOC MP.

SSHOC project inputs

In addition to these primary sources of information and to the requirements based on our consultations with end-users (see next section), we also take into consideration SSHOC inputs to be included in the system specifications of the SSHOC MP. This platform will integrate tools, services or resources produced by other SSHOC WPs, like the SSHOC switchboard developed by WP3; but also some of the results of WP4, such as the Sample Management System software, the Translation Management Tool, the Open source Computer Assisted Translation, the Social policy APIs, and the Aïoli platform. WP5 is working on a SSHOC data repository that will be referenced in the SSHOC MP. Furthermore, as WP5 and WP9 will produce case and pilot studies, it could be interesting to include their work as "scenario(s)" or "recipe(s)" in the SSHOC MP or as datasets when relevant[41].

33 https://ohttp://ssk.huma-num.fr/#/penmethods.dariah.eu/
34 https://www.dariah.eu/tools-services/contributions/
35 https://vlo.clarin.eu
36 https://colreg.de.dariah.eu/colreg-ui/?lang=en
37 https://datacatalogue.cessda.eu/
38 https://seriss.eu/
39 https://seriss.eu/training/tools/
40 https://seriss.eu/training/training-overview/
41 The International Ethnic and Immigrant Minorities' Survey Data Network involved in the WP9 Data Community Pilot is for example working on an Ethnic and Migrant Minorities (EMM) Survey Registry that could be integrated as a source for the SSHOC MP.

Knowledge and skills transfer represents one of the central pillars and enabling factors of EOSC. Correspondingly there are numerous initiatives and related catalogues offering training events and materials for various audiences. This kind of resources is particularly relevant for SSHOC MP as a source of information about scholarly practices. In the SSHOC project, WP6 is dedicated to collecting, organising and creating training events and materials. The idea is to integrate all output from WP6 into the Marketplace. Towards this goal, there is ongoing tight coordination between WP6 and WP7 to ensure a harmonised approach regarding the description and classification of these resources right from the beginning.

Beyond the integration of these resources, it should also be noted that the SSHOC reference ontology (T4.7/D4.18) and recommendations about (meta-)data interoperability problems (D3.1) or multilingual terminology (D3.9), as well as the SSHOC citation format (D3.2) will be used to develop the SSHOC MP.

From verbose publications to structured knowledge

The principle of Open Science, dictating transparency and reproducibility, demands a radical change in the scholarly practices, recognising that traditional publications of scientific results in articles are inadequate means for knowledge sharing.

One consequence is that research data, or even software, is being published alongside the traditional research papers and is increasingly being recognised as scientific output in its own right. It is not only necessary to reproduce the claims put forward in the scientific prose, it is also indispensable means for efficient propagation of ideas and solutions. Fellow researchers can quickly reproduce the results and use them as the basis for the next iteration of the research cycle.

Another effect of this shift in scholarly practice coinciding with the emergence of Semantic Web is the increasing representation of metadata describing the scientific outputs as semantically interlinked information, adhering to the principles of Linked Open Data, implicitly contributing to a global knowledge graph. This serves as a catalyst for efforts to exploit this structured information for better discovery and faster, more efficient, knowledge generation cycles. In the context of EOSC and for the development of SSHOC MP, the FREYA PID Graph, and especially the OpenAIRE Research Graph are important implementations of such an "Open Science Graph". In this context, we cannot forget to mention the Research Data Alliance's Open Science Graphs for FAIR Data Interested Group[42], as well as the soon to be presented Workshop on OSG interoperability[43] at the Open Science Fair conference 2019 in Porto.

A third complementary aspect to these developments to overcome the limitations of document-centric publications paradigms are efforts to extract structured information from existing publications. While it is an established practice in disciplines such as Life Sciences, it is in rather early stages in Social Sciences and Humanities (cf. Fathala et al., 2017; Auer, 2018; Constantopoulos/Pertsas, 2019; as well as the initiative Open Research Knowledge Graph[44]).

All three lines of action contribute to a new paradigm of "knowledge-based methods of scholarly communication".

42 https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig
43 https://www.opensciencefair.eu/workshops-2019/open-science-graphs-interoperability-workshop
44 https://projects.tib.eu/orkg/

Given that these developments, especially the information extraction task from research papers to populate an ontology is subject of research itself, the SSHOC MP cannot afford to rely on it. However, the idea of structured knowledge about scholarly practices is central to the Marketplace endeavour, and research papers can indeed serve as an important source of information about actual research practices, tools and methods used. Thus this field will possibly be accommodated as an experimental branch in the development of the Marketplace.

Challenges of the SSH Open Marketplace

Based on this state of the art of SSH registries and discovery frameworks, we can identify several challenges to overcome. Indeed, as we aim to provide a dedicated platform for SSH researchers, that can facilitate the digital aspects of their work, and that will be seamlessly integrated into the EOSC landscape, we need to pay attention to the representativeness of the **communities** we work for and to the conditions of integration in the existing landscape. The main challenge is to reflect the existing research practices and methods providing not just another directory but the **contextualisation** and the interconnection that is missing between/in the existing ones, so as to provide the researcher with the optimum response to his/her request. While attempting to achieve what is currently missing through the existing tools, services, datasets or training materials, the SSHOC MP also represents an opportunity to display those resources centrally, and to offer a **vitrine of the SSH research practices**. Research Infrastructures and the other partners associated in the SSHOC project provided us with the right context to bring SSH communities together and to work in a framework that offers cutting-edge interoperability solutions.

User Requirements

User requirements collection has been an essential part of our work to achieve the specifications presented after. As a community-based project, we decided to align our approach with User-Centered Design principles and agile approach to software development. We therefore first started working on identifying key stakeholders, including future users of the SSHOC MP. In order to identify user stories, we used interviews, workshop sessions and desk-based research. In the final step, user stories were transformed into initial technical requirements.

Target users

As indicated on the main pages of the EOSC portal, "The EOSC will offer 1.7 million European researchers and 70 million professionals in science, technology, the humanities and social sciences a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by federating existing scientific data infrastructures, currently dispersed across disciplines and the EU Member States."[45] Alongside this general statement, WP2 and 6 of the SSHOC project worked on stakeholder categorisations and community engagement strategy allowing us to refine the vision of the SSHOC community[46]. SSH researchers and support staff for SSH researchers (identified for example in the SSHOC stakeholders mapping as Research & e-Infrastructures, Research libraries and archives institutions or Universities and research performing organisations) were primarily identified as end-users of the SSHOC MP.

As a first step, we chose to focus our attention on the SSH researchers user group. Since WP7 is mainly composed of representatives of the Digital Humanities communities, it was straightforward to collect respective user stories, mainly using the desk-based research approach. In order to gather requirements from social sciences communities and non-digital humanists, we conducted a series of interviews with some of their representatives. To get a good representation of our target communities, we used several criteria before soliciting researchers who were invited to be interviewed:

- To cover the SSH domains, we chose to follow the ERC panels for the SSH[47] and invite at least one researcher from each of the SSH panels: SH1 Individuals, Institutions and Markets (Economics, finance and management); SH2 Institutions, Values, Beliefs and Behaviour (Sociology, social anthropology, political science, law, communication, social studies of science and technology); SH3 Environment, Space and Population (Environmental studies, geography, demography, migration, regional and urban studies); SH4 The Human Mind and Its Complexity (Cognitive science, psychology, linguistics, education); SH5 Cultures and Cultural Production (Literature and philosophy, visual and performing arts, music, cultural and comparative studies); SH6 The Study of the Human Past (Archaeology, history and memory).

45 https://www.eosc-portal.eu/about/eosc
46 D2.1 SSHOC Overall Communication and Outreach Plan and D6.1 SSHOC Community Engagement Strategy, available here: https://www.sshopencloud.eu/publications/deliverables
47 See the ERC panels for the SSH here: https://erc.europa.eu/sites/default/files/document/file/erc%20peer%20review%20evaluation%20panels.pdf. The following other nomenclatures have been considered: AUREHAL (https://aurehal.archives-ouvertes.fr/domain?locale=en) and r3data (https://www.re3data.org/browse/by-subject/). ERC panels sub-categorization has been chosen because of the numbers of sub-categories (6 vs 27 for AUREHAL ontology and 13 for r3data) and because it is a European nomenclature.

- In order to glimpse into the different phases of a researcher's life, we have decided to distinguish between two subcategories: early-stage researchers and experienced researchers.
- To take into consideration researchers without any background or experience with digital methods, we distinguished between "usual suspects" (researchers using digital methods on a daily basis), researchers aware of the digital landscape, and researchers not especially experienced with digital work.

This methodology[48] led us to plan 26 interviews, of which 22 were eventually conducted. We paid attention to country and gender representativeness as it is described in Annex 1.

Thanks to these interviews, we got a better understanding at how SSH researchers encompass the digital in their day-to-day research work, in light of their regular usage of services, tools and resources but also through their attempt to discover new ones. As presented in Annex 2, interviews were divided into two parts: first, a set of questions to get an idea of the research of an interviewee and its link to digital aspects; and then questions on the research habits and practices focusing on a) the digital aspects of the work and b) the way the interviewee finds and chooses the resources used.

Based on these materials, we were able to refine needs that were already considered in the initial notion of the SSHOC MP, as the necessity for a well-curated platform or the need to link training materials to tools and services. Interviews also allowed us to confirm needs that were identified as essential in our first discussions, such as the community dimensions to be taken into account in the project. The results of these interviews are presented in detail in the next section "User stories from the SSHOC interviews".

Beyond the needs of SSH researchers, we also started to address the needs of the second essential category of users: the support staff for SSH researchers. This has been done in particular through a workshop during the LIBER Conference: "Social Sciences & Humanities Open Cloud: What's in it for research libraries?"[49]. The key takeaway of this workshop was the role of research communities when it comes to the use of services and resources. The possibility to ask for advice or recommendations in the community has been highlighted as a very common behaviour, and research librarians who were present during this session recommended to reflect these community aspects in the SSHOC MP. Furthermore, participants also stressed the importance of having good metadata quality to improve the actual discovery process of tools, services or resources. Further collaborations with LIBER working groups, but also with other SSHOC WPs are planned during the course of the project in order to improve the involvement of support staff for SSH researchers in the Marketplace.

User stories from the SSHOC interviews

The 22 interviews conducted were transformed into two different materials. First, they were transcribed and summarised. Those transcriptions still remain stored by the assigned Data Controller (DARIAH ERIC) in order to be further studied throughout the SSHOC project – on the basis, of course, with the interviewees' approval and in accordance with the GDPR.

[48] The methodology used is based on the results of the DARIAH Digital Methods and Practices Observatory Working Group (DiMPO) and in particular on the framework for meta research into digital practices. See here some results of the DIMPO Working Group: https://www.dariah.eu/2019/07/31/working-groups-stories-11-digital-humanities-work-in-focus-multiple-case-studies-of-research-projects-across-europe/

[49] For more details, see the program of the workshop (https://liberconference.eu/programme/workshops/social-sciences-humanities-open-cloud/) and the blog post on the SSHOC website (https://www.sshopencloud.eu/news/social-sciences-humanities-open-cloud-what%E2%80%99s-it-research-libraries).

Secondly, those interviews had to be turned into user stories in order to properly express users' expectations and translate them further into more technical requirements. The user stories' translation was achieved by following the methodology adopted for the DESIR project and an agile method:

Based on the interview #	User story #	As a (role)	I want to (something)	So that (benefit)	Features (user requirements)

Table 1: User story format

Here is an example of a user story from the interviews that uses this very format:

Based on the interview #	User story #	As a (role)	I want to (something)	So that (benefit)	Exemples / feature
7	7.2	Young researcher in political science and European studies	know what kind of tools the researchers from my community use	I can identify what is probably the best tool for me as well	Researchers' views / comments; quotations and links to forum.

Table 2: Example of a single user story extracted from the interviews

Based on conducted interviews, we identified 81 user stories – all available in Annex 3. They essentially have two motives: (1) they account for what researchers expect from the future SSHOC MP and (2) they express researchers' current needs in their day-to-day work. Next, the collection of user stories was analysed in order to identify groups of user stories related to topics and specific feature sets (these groups can be seen as "epics" from an agile development perspective). In parallel, technical requirements (functionalities) were extracted from each user story and then prioritised – as explained in more detail below. As a result, we obtained a list of requirements, each with a specific priority and grouped by type of feature.

Grouping user stories

Eight main thematic groups of user stories emerged from this merging exercise. Those are either related to specific types of content the future users expected to find on the SSHOC MP, or more broadly to the platform's user-friendliness and management. Identified groups of user stories are as follows:

1. CONTEXTUALISATION or "I want to be offered a thorough contextualisation/information to answer my problem/request": This group expresses user needs for well-documented answers to their requests through high-quality contextualisation and information. Although interviewees may have had different opinions on the concrete implementation, they all agreed on this being a very important point.

2. COMMUNITY PLATFORM or "I want to be able to communicate with the SSH research community through the SSHOC platform": Here, the interviewees transcribed their will to use a community-oriented platform through several possible features, allowing them to either exchange with other researchers or to evaluate a solution (e.g., a tool) based on the community opinions/advice (e.g. rating mechanisms, peer reviews).

3. TRAINING MATERIALS or "I want to have access to dedicated and comprehensive training materials": Through this third identified group, researchers explained their will to access different training materials (e.g.,

tutorials, FAQs, user stories) when being suggested solutions to their request in order to resolve tools' related research problems, gain in experience or build up a digital background knowledge.

4. DISCOVERABILITY AND USABILITY or "I want to use a user-friendly and useful platform to find solutions": For this fourth one, future users translated their user-friendliness expectation via two main aspects: (1) through the search/browse/filtering features they would ideally like to use and (2) through usability aspects of the future SSHOC MP (UX/UI). For instance, regarding the search, many interviewees highlighted the possibility to personalise it based on specific characteristics, such as SSH (sub) disciplines, tool types, licences, languages, etc.

5. DATASETS or "I want to have centralised access to datasets": Researchers expressed their needs in terms of centralised access to datasets in a way that the future platform could provide/suggest both tools and datasets based on a user's request.

6. CURATION or "I want to use a well-curated platform": The curation of the Marketplace content was an important feature, especially for researchers already familiar with the use of digital tools in their research process (the "usual suspects", if you will). Indeed, they generally asked for a well-curated platform to find up-to-date information on the resources. For those interviewees, the curation ought to be carried through (1) the community itself (e.g., peer reviews) and (2) dedicated curators as part of the overall platform management.

7. GDPR (COMPLIANCE) or "I want to use a platform compliant in terms of data privacy policies": For this seventh group, it was identified that GDPR compliance was essential for most interviewed researchers, especially social sciences ones. The GDPR compliance of the platform had to be implemented through (1) the provision of GDPR information related to the suggested solution (e.g., a tool) and (2) the general and overall management of the SSHOC platform. On this last point, some researchers expressed their will to know who would manage this Marketplace to better trust and use it.

8. OPENNESS or "I want to use a genuinely open Marketplace": Last but not least, this item related to the expected openness of the future Marketplace, in terms of accessibility to the entire SSH community, but also in terms of the content it will provide.

The user requirements' prioritisation

As previously explained, there was a need to prioritise the 81 user requirements dispatched among the eight main identified groups of user stories. To do so, it was decided to follow the MoSCoW method[50] and to prioritise the requirements using four categories:

1. Must have: the user requirements related to the core priorities, namely the ones repeatedly mentioned by the interviewees and with a feasible technical implementation in the framework of the project.

2. Should have: the user requirements still identified as important, but which had either (1) a technical implication that leads it to be postponed to a medium-term implementation or (2) that weren't perceived by the prioritisation group as the core components of the future SSHOC Marketplace.

50 https://en.wikipedia.org/wiki/MoSCoW_method

3. Nice to have: the user requirements identified as interesting to implement and generally feasible in the long run.

4. Would not have: the user requirements identified as non-implementable either in terms of technical implications or because they wouldn't meet the SSHOC MP initial objectives, which were therefore rejected.

For the time being, the "must-have" will first be subject to implementation as they constitute the basis of the future platform. It is important to keep in mind that the others, namely the "should have" and "nice to have" aren't forgotten and/or left aside. They rather constitute medium- and long-term objectives and strategies for the platform development and will be dealt with in due time.

Here are the users' stories classified through each priority and based on the eight groups of user stories:

1. The "must have" requirements:

Item #	User requirements
1. CONTEXTUALISATION	→ Possibility to select/pick a tool based on the problematic encountered; → Possibility to find tutorials related to the tool/problematic encountered; → Possibility to be suggested alternative tools / present contents as alternatives to other more well-known ones; → GDPR compliance section to provide information on the GDPR compliance of the tools/solutions/throughout the data life cycle; → Possibility to find a section on the tool's price; → Possibility to find information on the technical functioning of a tool (How to install? How to use?); →Possibility to find information on how actively the tool is being used/developed/supported (obsolescence).
2. COMMUNITY	→ Possibility to have different types of peers' reviews: quotations, comments, views, rating; → Possibility to link to a forum to get more information around the problematic.
3. TRAINING MATERIALS	→ Possibility to access different types of training materials: tutorials, screenshots, users' stories, articles, help files and FAQs of tools, How To's (to be standard compatible / to create FAIR data / ...)
4. DISCOVERABILITY AND USABILITY	→ Possibility to use the search feature and organise researchers' request based on disciplines by (sub)categories, tools' families/functions, multilingual, tools' open access and FAIRness, keywords, resource types, data formats; → Possibility to find a classification of tools/solutions based on resource types for which they are applicable (input & output format); → Possibility to obtain the most pertinent result; → Possibility to use existing accounts to access the platform; → Access a platform free of charge; → Possibility to get a list of relevant tools to allow overview and comparisons; → Possibility to find tools/solutions through keywords and tags.
5. DATASETS	→ Possibility to find datasets (and present the conditions of access/availability).
6. CURATION	→ Possibility to find updates on the obsolescence of tools/solutions; → Possibility to get curation through peers' review; → Possibility to get an overview of relevant functions a tool supports, e.g. if a tool can import/process/export data in a specific format (e.g. TEI); → Possibility to get recommendations of workflows a tool supports.

7. GDPR	→ Possibility to get information on the platform data privacy policies' compliance (tools or any other solutions provided); → Possibility to get information on the platform management/governance; → Access a search bar/platform that respects the anonymity of users (collected and stored data is anonymous); → Possibility to find explanations on how to implement GDPR requirements during the research process.
8. OPENNESS	→ Possibility to find all the existing resources/tools/workflows to his/her problematic/request from the entire SSH community and not confined to the ERICs community only.

Table 3: "Must have" requirements

On the inclusion of datasets, some interrogations arose within WP7 regarding users' potential needs and the degree of inclusion of datasets in the SSHOC MP. After the interviews were conducted, and during the results' analysis, it appeared that users' will to find datasets within the SSHOC MP wasn't actually emphasised. This gap can be explained because of a bias observed in the questionnaire. Indeed, the interview began with a short description of the future platform that presented the inclusion of datasets, but the questions seemed too much "tool oriented". They, therefore, led interviewees to take for granted the inclusion of datasets and didn't leave enough room for comments on that matter.

This point being noted, the extent to which datasets needed to be included and how to insert them was central. It was important for WP7 that the SSHOC MP wouldn't be a catalogue of existing datasets because enough already existed. Also, it differed from one of its core objectives: providing quality contextualised information through a quality curation process. Aligned with this aim, the choice was made to include datasets in a contextualised manner: only relevant datasets that could be linked to a resource would be pointed out. Also, datasets' catalogues could be found as services in the future SSHOC MP. Finally, to overcome potential gaps, a hypothesis is currently being discussed: the TRIPLE project[51] could provide access to other datasets, and its platform would be linked as a service in the future SSHOC MP.

2. The "should have" requirements:

Item #	User requirements
1. CONTEXTUALISATION	→ Possibility to view pro/cons for tools; → Possibility to access a section/description based on FAIR principles or other similar standards.
3. TRAINING MATERIALS	→ Possibility to access training materials that are clear, simple, not time-consuming, and mostly created by SSH researchers, not by computer scientists.
4. DISCOVERABILITY AND USABILITY	→ Possibility to suggest the researcher what he/she didn't suspect (e.g. resources in other languages than the ones chosen, alternatives to most commonly employed solutions/tools) through a kind of explore function (show me something new); also give meaningful/alternative information when there is no result.

Table 4: "Should have" requirements

3. The "nice to have" requirements:

[51] The TRIPLE (Targeting Researchers through Innovative Practices and Multilingual Exploration) project will start on October 2019 will focus on reuse of data and projects in the Humanities and Social Sciences (see https://humanum.hypotheses.org/5384#en).

Item #	User requirements
1. "CONTEXTUALISATION	→ Possibility to be suggested academic literature on tools whenever possible; → Possibility to access custom-tailored scripts linked with resources in the portal (i.e. Jupyter Notebook); → Possibility to get recommendations based on the user-friendliness of the front-end user interface of tools.
2. COMMUNITY	→ Possibility to view recommendations from organisations (RDA, DDI...) or based on already used tools in other research projects; → Possibility to view the usage of tools based not only on analysis of current research papers but also on what and how a tool is used in my research community.
3. TRAINING MATERIALS	→ Possibility to share my experiences and solutions with others.
4. DISCOVERABILITY AND USABILITY	→ Possibility to access to a Q&A-tool; → Possibility to get push messages/emails if new tools/resources are registered that fit a researcher's profile.
6. CURATION	→ Possibility to access a "gap analysis"/" feedback" section where users could inform on what the platform could not provide them so that the curator later take it into account and tries to answer this and show this gaps when other users search for it.
8. OPENNESS	→ Possibility to access information that implements critical view regarding the more technologically oriented approaches to a given problem.

Table 5: "Nice to have" requirements

4. The "would not have" requirements:

Item #	User requirements
1. CONTEXTUALISATION	→ Possibility to access a list of the most common issues; → Possibility to use a Chatbot popping out "How can I help you?"; → Possibility to try out tools through a Virtual Machine to test them beforehand.
2. COMMUNITY	→ Define a minimum number of peers' review to consider evaluation as relevant and trustworthy; → Possibility to gather and comment tools in a dedicated group (social functions like follow the activity of a user/group).
5. DATASETS	→ Possibility to access datasets through one single platform; → Possibility to get lists of existing datasets; → Possibility to differentiate between "raw data" and "processed data".
6. CURATION	→ Possibility to get an overview of different versions of a tool.

Table 6: "Would not have" requirements

In this section, we elaborated on user requirements as collected via interviews from potential users, resulting in a comprehensive list of needs related to the type of content that will be available in the platform, specific features and functionalities as well as trustworthiness, provenance, or quality of the content. In the following section, we describe the corresponding data model that needs to be expressive enough to capture all the

required information. We also touch upon potential sources of this information that will be included in the Marketplace.

Data Model

Considering user requirements as well as the data models of potential sources, we devised a data model, which defines the main entities/classes, their attributes and the relations between them. It needs to be expressive enough to capture all relevant information in the available sources and support all features retrieved from the user requirements.

In the following subsection, we motivate our modelling decisions against the backdrop of primary sources as well as other relevant/related work. The main part of this section describes the entities of the data model and their relations.

Existing sources as the baseline for the data model

As mentioned in the previous section, during the initial phase of the SSHOC project, we have conducted an investigation within the SSH communities, in which we have also solicited most pertinent catalogues in the respective domains. These are expected to become key sources of information for populating the Marketplace. Thus, we need to be able to map the information in these sources onto the Marketplace data model. We have identified around 30 catalogues representing potential sources for population. We evaluated these with respect to the type of entities they provide, their size, thematic scope, responsible organisation, etc. to determine a priority order for processing these sources. A selection of the most pertinent catalogues is available below in Table 7. Subsequently, we tentatively manually extracted relevant information for individual items in these sources and expressed in terms of the data model, in order to validate (cf. Table 8 below).

Catalogues	Data?	Tools?	Recipes?	Other?	Nr of records	Age (years)
TAPoR	no	yes	no	Papers	1495	12
CESSDA Data Catalogue	yes	no	no		19,188	1
CLARIN VLO	yes	yes	no	Services	1 mio	8
Parthenos SSK – Standardization Survival Kit	no	no	yes	Resources	27	3
In-Kind Contribution Tool DARIAH	yes	yes	no	Services, Events	~ 1000	4

Table 7: Extract of the "sources mapping" created by Task 7.3

Key [type]	Label (accessibleAt)	Description	related	Required Input Features	Provided Output Features
gephi [Tool]	gephi.org	visualisation and exploration software for all kinds of graphs and networks	gephi_intro	Graph Data	
gephi_intro [InfoObj]	Introduction to GEPHI	This session will provide an overview of the software, its features, and resources for further study.	gephi	/	/

stata [Tool]	stata.com	Stata is the solution for your data science needs. Obtain and manipulate data. Explore. Visualise. Model. Make inferences. Collect your results into reproducible reports.		Numerical data	
websty [Tool]	ws.clarin-pl.eu/websty.shtml?en	Stylometric analysis tool		ZIP with txt files	numerical data, visualisations
SSK_sc_statisticalAnalysisOccupations [Activity]	Perform Statistical Analysis on Historical and Contemporary Occupations	Given a large dataset with several raw variables, a social science researcher needs to (re)code some of the data in order to properly conduct statistical analysis and modelling techniques. [...]	stata	/	/

Table 8: Extract of the examples collected from the primary sources for validating the data model

Main entities/classes

In this section, we describe the entities that form the data model of the Marketplace, including their attributes and relations between them.

Some general remarks on the data model:

- **Generic model**
  Given the heterogeneous dataspace we aim to cover, and the broad and underspecified range of information we may want to capture about the entities represented in the Marketplace, we opted for a generic data model that can be refined through configuration during runtime. Main mechanisms are: Every MP Entity can have a set of *Properties,* i.e. key-value pairs, where allowed keys and allowed values for individual keys are specified in the configuration of the system, not in the data model itself. Equally, there is a generic relation between MP Entities, *related,* that can be typed as needed.
- **Flexible classification/typing**
  The above argument also implies that the data model should not hardwire a rich class hierarchy. Therefore, most information about the described resource is captured already in the top-class *MP Entity* and the majority of classification/typing is expected to be covered by appropriate Properties (*keywords*) with corresponding Vocabularies. As an example, the various types of training materials, as expressed in the user requirements, are better expressed as a property of an MP Entity-instance, rather than being hardwired as classes in the data model. This is because a) there is no global consensus on the categorisation, b) a training material could belong to multiple categories, c) from the system point of view there is no principle difference between the different types, and it is more efficient to handle them all uniformly as *MP Entities*.
  Currently, the data model introduces only a handful of subclasses of *MP Entity*, dictated by special properties pertinent to these classes. It is also to a certain extent left for the implementation, which

types will need to be represented as separate classes, and made available over dedicated API endpoints.

- **Versioning**
  To support the envisaged curation workflow, the curator must be able to see/compare and approve/dismiss changes proposed by editors, or introduced by automatic updates. This implies a versioning system, where every change to an MP Entity, including the author, is recorded and can be reviewed. For reasons of transparency and reproducibility, a complete history of changes must be available.

In the overview (and the diagram) we used the following conventions:

- Quantifiers:
  + = one or more values
  * = zero or more values
  ? = zero or one value
- Prefixes for namespaces
  - crm: http://www.cidoc-crm.org/cidoc-crm/
  - crmdig: http://www.ics.forth.gr/isl/CRMext/CRMdig.rdfs/
  - foaf: http://xmlns.com/foaf/spec/
  - prov: http://www.w3.org/ns/prov#
  - skos: http://www.w3.org/2004/02/skos/core#

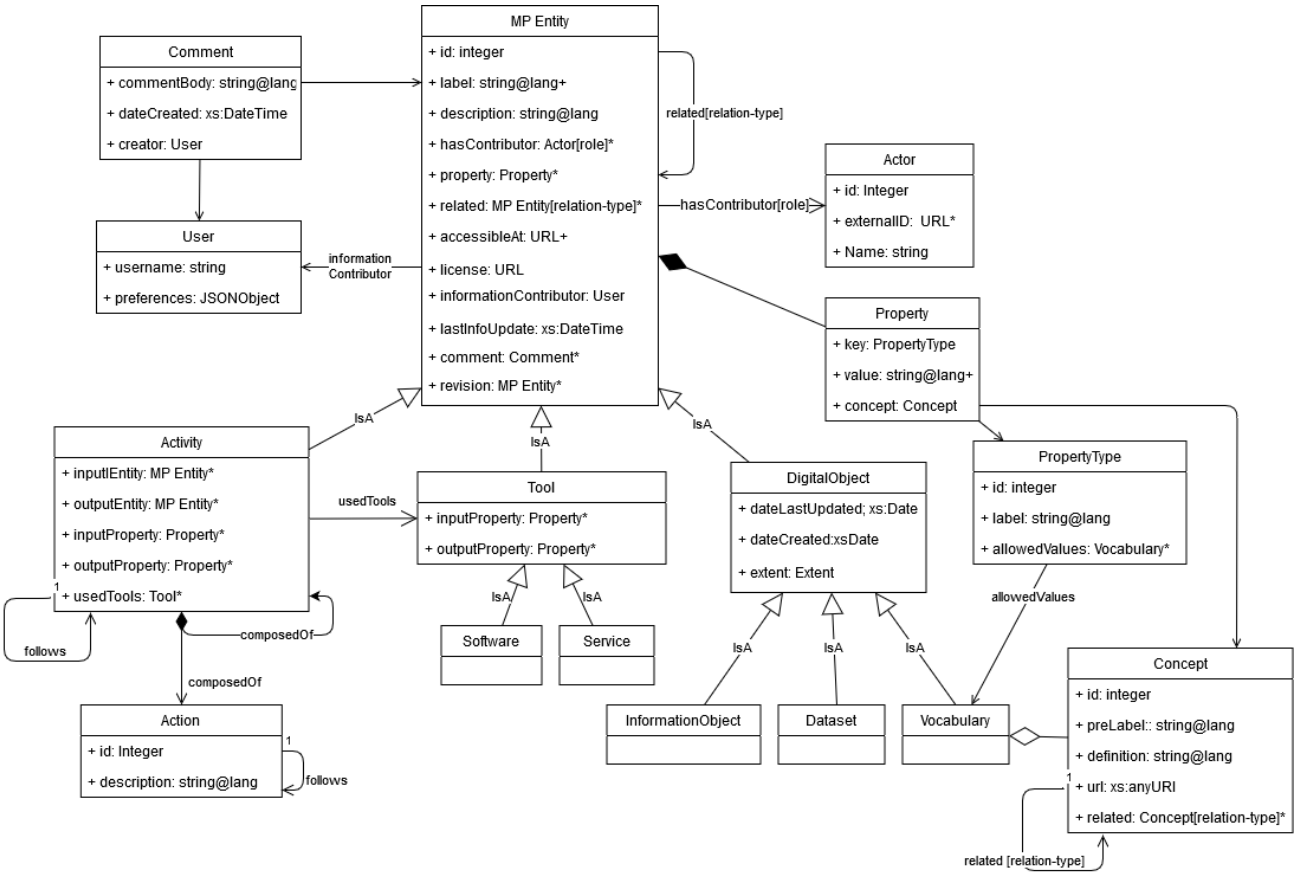

Fig. 2 Data model diagram

Class Label	MP Entity
Subclass Of	-
Superclass Of	DigitalObject, Tool, Activity
Properties / Relations	• id: integer • label: string@lang+ • description: string@lang • hasContributor: Actor[role]* • properties: Property* • related: MP Entity[relation-type]* • accessibleAt: URL* • license: URL • informationContributor: User • comments: Comment* • lastInfoUpdate: xs:DateTime • revision: MP Entity*
Scope Note	Top class of all primary entities captured in the Marketplace. There are only a few dedicated attributes like a label, description, etc., most of the information about an MP Entity is to be captured through a flexible set of Properties. The Properties are also used for any classification/categorisation/tagging of MP Entities with Concepts from any Vocabularies. An attribute to highlight is *accessibleAt* which should contain a link to the MP Entity in its original context. Another generic mechanism is the attribute *related* that allows setting relations between two MP Entities. Allowing to set the type of relation dynamically. I.e. which types of relations between entities are allowed is not hard-wired in the data model but can be defined by the administrator at runtime. A typical example of such a typed relation would be a MP Entity (e.g. a Tool) *isMentionedIn* another MP Entity (typically an Information Object, which could be a research paper or a training material). Similarly, the relation between an MP Entity and an Actor is kept generic, with the configuration of applicable "roles" of an Actor with respect to an MP Entity left to administrators at runtime. Typical roles include: *hasContributor, hasAuthor, hasFunder*. MP Entity also features a few meta-attributes, pertaining to the provenance of the information gathered about given entity: informationContributor allows to keep track of the User who entered the information about that entity, lastInfoUpdate the time of last change to the entry. To allow for the envisaged curation workflow, where human editors as well as automatic processes can propose changes to an existing MP Entity, a versioning mechanism needs to be in place. This is tentatively indicated with the attribute *revision* but will need to be refined in the actual implementation.
Mapping	• crm:E1_CRM_Entity • prov:Entity

Examples	

Table 9: MP Entity

Class Label	Digital Object
Subclass Of	MP Entity
Superclass Of	Information Object, Dataset, Vocabulary
Properties / Relations	• dateCreated: xs:DateTime • dateLastUpdated: xsDateTime • extent: <Value,Unit>*
Scope Note	This class comprises identifiable immaterial items that can be represented as sets of bit sequences, such as data sets, e-texts, images, audio or video items, software, etc., and are documented as single units. Any aggregation of instances of Digital Object into a whole treated as a single unit is also regarded as an instance of Digital Object. This means that for instance, the content of a DVD, an XML file on it, and an element of this file, are regarded as distinct instances of D1 Digital Object, mutually related by the *related* property the type *isComposedOf/isPartOf*. A Digital Object does not depend on a specific physical carrier, and it can exist on one or more carriers simultaneously. (based on definition of crmdig:D1_Digital_Object)
Mapping	crmdig:D1 DigitalObject
Examples	

Table 10: Digital Object

Class Label	Information Object
Subclass Of	Digital Object
Superclass Of	-
Properties / Relations	
Scope Note	This class comprises the intellectual or artistic realisations of works in the form of identifiable immaterial objects, primary texts, but also images, multimedia objects, or any combination of such forms that have objectively recognisable structures. The substance of Information object is signs. An Information object is the outcome of the intellectual or creative process. Such information objects do not depend on a specific physical carrier and can exist on one or more carriers simultaneously, including human memory. (based on definition of frbroo:F2_Expression)

Mapping	crm:E73_Information_Object; FRBRoo: F2 Expression
Examples	• Tutorial/How to on using a tool • Training material • Research paper describing a solution

Table 11: Information Object

Class Label	Dataset
Subclass Of	Digital Object
Superclass Of	-
Properties / Relations	
Scope Note	Identifiable immaterial items that can be represented as sets of bit sequences and whose content contains propositions about the objective world. The identity of an instance of PE18 is determined by its content in bit level encoding alongside its provenance. Any instance of a dataset may be composed of many distinct parts of other identifiable datasets. An aggregate of instances of a dataset is treated as one instance, and its parts can be documented as having a part of the relation. Datasets in practice are either volatile or persistent. (source: crmpe:PE18 Dataset)
Mapping	• crmpe:PE18 Dataset
Examples	• A parallel text corpus • Results of a survey

Table 12: Dataset

Class Label	Vocabulary
Subclass Of	Digital Object
Superclass Of	-
Properties / Relations	• hasMembers: Concept*
Scope Note	A set of concept definitions
Mapping	• skos:ConceptScheme
Examples	• TADiRAH

	• Getty Thesauri: AAT – Art & Architecture Thesaurus

Table 13: Vocabulary

Class Label	Concept
Subclass Of	-
Superclass Of	-
Properties / Relations	• Id: integer • memberOf: Vocabulary+ • prefLabel: string@lang • Definition: string@lang • Url: xs:anyURI+
Scope Note	The definition of Concepts, which are parts of Vocabularies, used to classify, categorise, tag items.
Mapping	• skos:Concept , CIDOC CRM: E55 Type
Examples	• TaDiRAH: Relational Analysis[52]

Table 14: Concept

Class Label	Tool
Subclass Of	MP Entity
Superclass Of	Software, Service
Properties / Relations	• inputProperty: Property* • outputProperty: Property*
Scope Note	Is used in (or to perform) an Activity. For practical reasons we subsume under Tool both Software and Service, even though these are conceptually two very distinct entities (in CIDOC CRM terms Software is a Digital Object, while Service is an Activity). From the user's point of view, both Software and Service can be a "Tool" serving a function to achieve a certain goal. Nevertheless, internally, we will have to keep the distinction between software and service as these require partially different properties.
Mapping	
Examples	• A Tool allowing you to visualise data (Gephi) • A Service transforming PDF files into consolidated TEI-XML (Grobid)

52 http://tadirah.dariah.eu/vocab/index.php?tema=28&/relational-analysis

Table 15: Tool

Class Label	Actor
Subclass Of	-
Superclass Of	Person, Group, Organisation
Properties / Relations	• Name (string) • Affiliation (Group) • isContributorTo
Scope Note	Entity (Person, Group or Organisation) able of intentional actions. Actor appears as contributor to an MP Entity.
Mapping	• crm:E39_Actor • prov:Agent • foaf:Agent
Examples	• Austrian Academy of Sciences • Gertrude Stein • SSHOC project consortium • CESSDA

Table 16: Actor

Class Label	Activity
Subclass Of	MP Entity
Superclass Of	Project
Properties / Relations	• follows/isFollowedBy (Activity) • composedOf/partOf (Activity) • composedOf (Action) • inputEntity: MP Entity* • outputEntity: MP Entity* • inputProperty: Property* • outputProperty: Property* • usedTools: Tool*
Scope Note	Description/Instruction on **how** to perform certain actions to achieve a certain goal. It is compositional and ordered, i.e. an Activity can be subdivided in an ordered sequence of multiple smaller Activities or "atomic" Actions. It can be of any desired or available/feasible level of granularity. However, not strict composition – one activity can be part of multiple higher-level activities. An Activity can also be described in terms of Entities used as input for the activity or resulting as output/outcome of an Activity (*inputEntity, outputEntity*). In case of a more general recipe, it is probably not a concrete entity that is input or output of an Activity, but

	rather a class of entities with specific properties. E.g. output of a NLP processing task may be a text with additional annotation layers. The attributes *inputProperty* and *outputProperty* allow capturing such properties that may be required for input entities of a given Activity or are present in the output entities as a result of an Activity. *Note:* It will be a matter of fine-tuning and adjustment, how fine-grained we will (be able to) capture Activities. Considering the Scenarios in SSK, expressing the individual Steps of each Scenario would essentially duplicate it in the Marketplace. *Note:* For pragmatic reasons,, we do not distinguish between activities that were effectively performed and recipes/plans describing a set of actions to be taken, i.e. we disregard the factual state of activity.
Mapping	• crm:E7_Activity • Recipes in methodi.ca/tapor.ca • Scenarios in the Parthenos-SSK • prov:Activity
Examples	• Performing a NLP-processing (e.g. PoS-Tagging) on a given set of texts • SSHOC project • All the scenarios in SSK • Data management plans

Table 17: Activity

Class Label	Action
Subclass Of	-
Superclass Of	-
Properties / Relations	• partOf (Activity) • follows/isFollowedBy (Action)
Scope Note	An atomic part of an Activity. A sequence of Actions forms an Activity. *Note:* The level of granularity is flexible and subject to available information and desired level of detail.
Mapping	crm: E7 Activity
Examples	• Build the model of the dictionary (one step within SSK Scenario SSK_sc_dictionaryInTei) • Create a corpus of useful resources for the dictionary (one step within SSK Scenario SSK_sc_dictionaryInTei)

Table 18: Action

Class Label	Property

Subclass Of	-
Superclass Of	-
Properties / Relations	• key: PropertyType • value: string
Scope Note	Certain characteristic of an MP Entity, allowing to capture its atomic properties (e.g. a text is annotated with PartOfSpeech according to a given tagset). *Note:* This approach allows for most flexibility with respect to what can be captured, but, at the same time, limits the possibilities to restrict/validate the input, because it is not possible to restrict which values are allowed/valid for a given key. At least not with the usual modelling means. *Note:* Ideally, this can be used to match Information Objects and Tools, e.g. provided a Document with PoS-Annotation, one could find a tool for linguistic analysis which requires PoS-tagged text as input. Another typical "Feature" would be the language of a text.
Mapping	
Examples	• Text is tagged with PartOfSpeech from STTS-tagset • Object is of mime/type RDF/XML • Text is written in German

Table 19: Property

Class Label	PropertyType
Subclass Of	-
Superclass Of	-
Properties / Relations	• id: integer • label: string@lang • Allowed values: Vocabulary*
Scope Note	A certain aspect that can be used to describe an MP Entity. This class encompasses the allowed Property keys.
Mapping	
Examples	• language • annotation layer • spatial coverage

Table 20: PropertyType

Class Label	User

Subclass Of	-
Superclass Of	-
Properties / Relations	• Name: string • Preferences: JSON-Object • Auth_mode: Enum • Role: Enum
Scope Note	Representation of an individual person interacting with the system.
Mapping	
Examples	

Table 21: User

Controlled Vocabularies

Controlled vocabularies are central to curation, aggregation and discovery tasks: fields with a limited set of possible values lend themselves ideally for faceted browsing. However, in most cases, the values in such fields in individual sources are not aligned. Controlled vocabularies provide a normalized, coherent set of values, to which the values encountered in the source data can be ideally mapped/translated, during the post-aggregation curation step. It should be noted that this is a potentially dangerous procedure because it necessarily introduces a subjective judgement of the curator on the semantic equivalency of the source value and a value from a controlled vocabulary. Nevertheless, the benefit of consistent searching and browsing experience by the user as well as higher recall certainly outweighs this caveat. Moreover, there are strategies to remedy the risk of misinterpreting the data and thus misleading the user: the basis is making the process transparent. That means, on the one hand, explaining publicly the curation procedure, the transformation steps taken, including publishing the normalisation tables, and on the other hand, allowing the user to search for and to see both the normalized and the original value in the discovery system.

Complementary approach in this regard is to try to make the normalisation happen already on the side of the data provider. This is the most sensible point of intervention. Unfortunately, this is only seldom an option, most of the time the data from the data providers is provided as- is and there is little incentive and capacity to adapt the provided (meta)data to specific needs of an aggregator. However, it shouldn't be ruled out in general, and where direct contact with data providers can be established, this should, by all means, be attempted.

When selecting vocabularies to be employed in the marketplace, we need to be aware that there is a multitude of controlled vocabularies of widely varying scopes, maturity levels and communities using them. BARTOC.org[53] alone currently features almost 3.000 vocabularies. Fortunately, we can build on the work done in task 3.5, summarised in the deliverable D3.1 Metadata interoperability, where most relevant/important vocabularies in different disciplines were solicited from a number of interviewees.

In the following table we elaborate on the potential use of the vocabularies mentioned D3.1 in the Marketplace:

[53] Basel Register of Thesauri, Ontologies and Classifications, http://bartoc.org/

Vocabulary	Description (Size, Scope)	Applicability in MP
CESSDA Topic Classification[54]	A typology of main themes or subjects of data.	Topic classification (for social science related items)
DDI Controlled Vocabularies[55]	A collection of 23 CVs used to describe specific aspects of a dataset across the data life cycle	Could be used as vocabularies of allowed values for specific Properties for Datasets
ELSST[56]	A multilingual thesaurus for the social sciences (~3.3k concepts).	Topic classification (for social science related items)
CLARIN Concept Registry[57]	232 approved concepts, almost all on the metadata category	CCR defines concepts on the level of fields or properties. Thus, it is not applicable as a vocabulary of allowed values, but could be relevant for defining allowed Properties (PropertyType)
ISO 639-1 language list	List of language codes	Vocabulary for languages and official language codes. However, we should use the more detailed version ISO 639-3.
CLAVAS[58]	Contains only the ISO 693-3 codes as vocabulary	Could be a source for ISO 639-3 language codes. However, it is not the authoritative source.
OpenGeoNames	Global authority file for geographic entities 25 Mio. items	Primary authority for georeferencing.
TGN (Getty Thesaurus of Geographic Names)	Thesaurus of names and associated information about places 1.1 Mio. geographic names	Potential authority for georeferencing.
AAT (Getty Art & Architecture Thesaurus)[59]	Terminology for cataloguing visual arts and architecture 60,000 records and 375,000 terms	Probably too broad and different scope not really relevant for the entities present in SSHOC MP.
PICO Thesaurus[60]	Thesaurus for cultural heritage domain in the context of CulturaItalia platform,	Probably too specific scope.

54 https://vocabularies.cessda.eu/urn/urn:ddi:int.cessda.cv:TopicClassification:3.0 and https://vocabularies.cessda.eu/#!discover
55 https://ddialliance.org/controlled-vocabularies and especially https://ddialliance.org/Specification/DDI-CV/LifecycleEventType_1.0.html
56 https://elsst.ukdataservice.ac.uk/
57 https://www.clarin.eu/ccr
58 https://vocabularies.clarin.eu/clavas/
59 https://www.getty.edu/research/tools/vocabularies/aat/
60 http://www.culturaitalia.it/pico/thesaurus/4.3/thesaurus_4.3.0.skos.xml

	developed in 2011 865 terms	
VIAF (Virtual Authority File)[61]	Union of many national authority files, contains Persons, Organizations, Locations; Global coverage, it does not contain subject headings.	Could be used as authority for persons and organisations.
GND (Gemeinsame Normdatei)[62]	Primary authority file for German speaking area containing records for persons, places, subjects/topics; Also part of VIAF Size: altogether around 19 Mio. entries, around 212,000 subjects entries	Could be used for topics.
IANA mime/type[63]	Globally used classification of Media Types 9 main types, 640 with subtypes, but there are many more variations	InformationObject Format Can be partially detected automatically.
TaDIRAH	Taxonomy of Digital Research Activities in the Humanities, containing 121 terms	ActivityType
NeDiMAH[64] Methodology Ontology (NeMO) Activity Types	An ontology of digital research methods in the arts and humanities, 161 activity types, 106 Information resource types and 1531 media types	ActivityType, InformationObject Format

Table 23: Vocabularies applicability in the Marketplace

In the data model the Vocabularies are represented as a separate class, with individual terms represented as Concepts. This model is intended in accordance with the SKOS scheme, which is also foreseen as primary import and export format for vocabularies. Important aspect of the SKOS model is that the central entity Concept is not determined lexically but semantically. I.e. a Concept is a representation of an "idea" that can be verbally expressed in various ways. These can be captured by multiple *labels* of a concept.

61 https://viaf.org/
62 https://data.dnb.de/opendata/
63 https://www.iana.org/form/media-types
64 http://nedimah.dcu.gr/index.php?p=navigate#

System Design

In the following section, we explicate the process of transition from user requirements to technical components and implementation, and we describe the individual components of the envisioned overall architecture of the Marketplace.

Translation of user requirements to technical solution

In the User Requirements section, we summarised the requirements as collected and digested from the user interviews. As such, these are formulated from the user's perspective. For the actual implementation, these requirements need to be translated into actual functionalities and features of a technical solution. The process of translating user requirements to functional features of the system is non-trivial and non-linear; they can't always be linked one-to-one. Multiple user requirements can be covered by one feature, as well as multiple features may be needed to cater for one user requirement. Also, changes to the priorities may be necessary for the implementation process, because e.g. features that may be less important for the user, may be necessary as a basis for further functionalities and thus need to be implemented first. Our procedure in this translation task was to record requirements identified as the result of the user interviews as individual issues (of type "Feature") in a ticketing system. From there, further discussion of adjusting, reshuffling, reformulating these features for the needs of implementation can be tracked in a structured manner within the system, for instance, groups of user requirements can be modelled as epics. The ticketing system used is Gitlab, and all future discussions and developments will be made openly accessible at the GWDG's instance[65].

Additionally, further features need to be defined and implemented that were not directly mentioned by the users but will still be necessary for a viable system. One example of such features would be functionalities related to user management.

The following table provides a tentative alignment of the eight main groups of user requirements that will be implemented in the system:

User requirements group	**Implementation**
1. Contextualisation	Data model allows linking items described in the Marketplace with each other. If the link exists, this information will be displayed in the details of an item. As a result, it will allow users to browse/traverse through the items, using the links defined between them. However, it will also depend on the Community and Curation if enough relevant contextual information will be provided.
2. Community	Features that allow active participation of the users, commenting, potentially rating, but mainly light-weight crowd-sourcing, where through means of micro-editing, users can contribute little pieces of information with the minimal threshold. This requires user management and/or anti-spam/anti-abuse mechanism, even though in general the platform will be available anonymously.

[65] https://gitlab.gwdg.de/sshoc

3. Training Materials	Training Materials can be considered part of Contextualisation. They can be attached to an item, e.g. a Tool or an ActivityPlan
4. Discoverability and Usability	The main component of the system, featuring faceted navigation (e.g. over ActivityType) as well as full-text search (if applicable to an item). This component is crucial to the user-friendliness, usefulness and thus, the adoption of the Marketplace.
5. Datasets	Users indicated that datasets should be available in the Marketplace. There are several alternative ways to tackle this requirement: (1) create full-fledged data aggregator, (2) add links to datasets that are related to items existing in the Marketplace (e.g. link a tool to an exemplary dataset that it can operate on), (3) add items in the Marketplace representing existing data aggregators. After a thorough investigation, it was clear that creating yet another huge aggregator of metadata records about data seemed to be neither feasible nor desirable. Serious aggregation efforts and corresponding catalogues are already conducted in corresponding communities (CESSDA Data Catalogue, CLARIN VLO, …) or in general (OpenAIRE, Europeana), and duplicating the millions of heterogeneous records would be, besides swamping the Marketplace, a project on its own. Therefore, we decided not to follow option (1) mentioned above. Nevertheless, we aim to implement two other options. First, we will capture (meta)data catalogues as items in the Marketplace, so that the users are at least pointed to, where they can look. Second, wherever feasible and justifiable, we plan to link Marketplace items to external datasets from existing sources. Thanks to that, we deliver a compromise – datasets are findable in the Marketplace, as users suggested, but only the ones directly connected with well-curated content of the platform. Plus, the users can find links to well recognised and established data aggregators.
6. Curation	Curation, in the sense of quality assurance, is crucial to the usefulness of the whole platform. We envisage a three-fold curation/governance strategy (although the final approach will be worked out in the course of Task 7.4): - Hired/professional moderators/curators, who are managing the continuous imports from defined sources, but also reviewing the contributions from the users - Continuous automatic checks supporting the moderators, by identifying old or missing information, dead links, etc. - Light-weight crowdsourcing through micro-editing – allowing users to suggest new or different information on a specific item, which needs to be reviewed by the moderator
7. GDPR compliance	(1) GDPR compliance will be achieved by developing appropriate Privacy Policy of the Marketplace platform and ensuring the presence of legal justification for personal data processing activities. Features of the Marketplace such as "properly informing users about the cookies the platform is using, and whenever necessary, receiving their consent" will be implemented in order to align GDPR regulations. (2) GDPR compliance of the entities presented in the SSHOC MP will be one of the descriptions features when relevant.

8. Openness	In the sense of open to everybody – the Marketplace will be usable anonymously, with some features (like commenting) restricted to registered users. However, we aim to support a very lightweight form of registration (via Federated Identity, Shibboleth, OpenID). In the sense of opening the scope of the content beyond the context of ERIC, this will depend on the curation process and the sources of information we will be able to identify and digest.

Table 24: Implementation of the user requirements

System Architecture

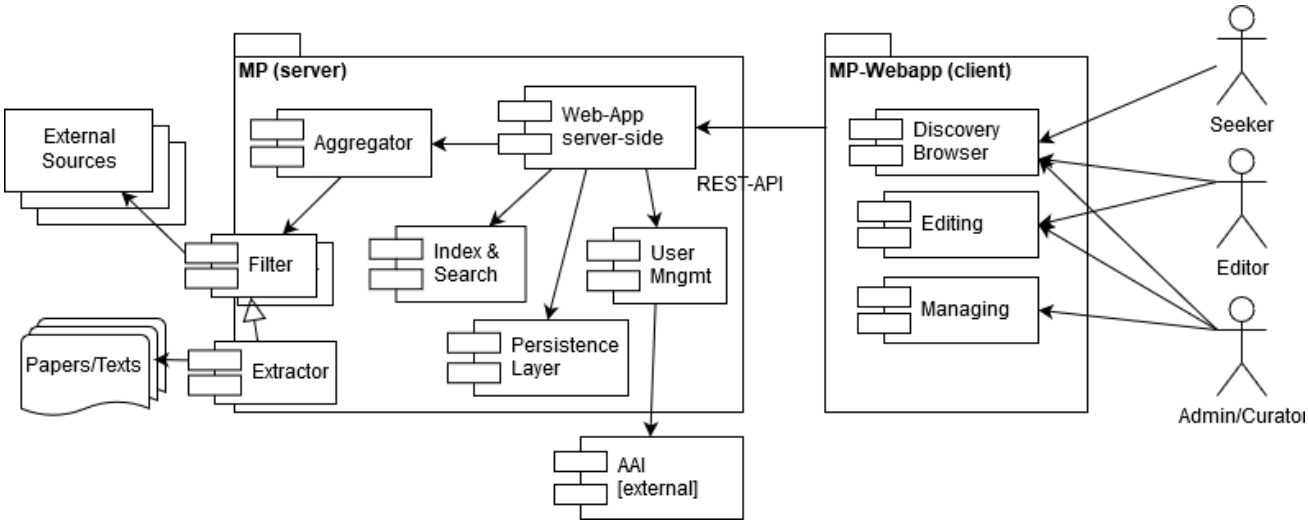

Fig. 3: System Architecture Diagram

Components:

Server-side web application	Exposes REST API, including methods used by web application.
- Persistence Layer	Stores information offered by and necessary to operate the Marketplace, e.g. information about users, metadata about entities available in the platform.
- Index & search	Provides efficient mechanisms (including a search engine) supporting discovery features of the platform, keyword and faceted search.
- Aggregator	The component is responsible for automatically harvesting information from identified sources, transforming and ingesting it into the platform. It requires custom filters, mapping and a clear policy on how to deal with updates/conflicts.
- Extractor (optional)	Can be considered as a special kind of Aggregator that takes text as input and tries to extract relevant information from it, then again ingesting it into the platform. *This is considered as an optional component serving an experimental aspect of the platform that is mentioned in the Chapter "From verbose publications to structured knowledge".*

- User management	Primarily we aim to rely on Identity Federations (Shibboleth, OpenID). Nevertheless, the system needs to have a "local" representation of the user, and also we need a fall-back to register locally if all else fails. This component also comprises a user profile that could capture user's search history or allow her to bookmark certain items or store queries, etc. Rely on federated identity (AAI/SSO). DARIAH Proxy IdP service allows managing users roles. This would be an option to outsource role management completely.
Rich client application Can be divided into following subcomponents or modules:	Implemented using React framework, communicating with the server solely through the defined REST-API (the modules obviously need corresponding counterpart methods on the server-side).
- Search/Discovery	Faceted browser and full-text search for the end-user to explore the content.
- Detail View	A detail view of each entity, gathering and presenting all existing contextual information, also allowing to navigate "sideways" to similar entities based on the contextualisation.
- (Micro-)Editing	Collaborative editing/curation of the information/content. The authoring mode, allowing to make changes to data. Only available to logged-in users, distinguishing roles. General logged in users can suggest changes which need to be approved by a moderator. Similar to wiki-data, the idea of micro-editing is that a user can suggest just one specific fact or a piece of information for an existing entity entry in a quick and intuitive fashion.
- Managing	A module for the moderators, power-users, a kind of a dashboard informing about status and history of automatic imports (aggregation) and checks as well as manual (suggested) changes (moderation of the community).
- Vocabulary Management	A big part of the Managing and curation will be dealing with / curating vocabularies. This would justify a dedicated (potentially external) tool for managing the vocabularies.
Data Lab / Notebook (auxiliary service – to be optional at best)	This is not actually considered part of the Marketplace but is expected to be an externally provided but tightly connected service that allows capturing and "replay" recipes, workflows. It is to be expected that a "data notebook" service will be offered by EOSC out of the box, however, there are already many jupyter hubs and similar services around. Describing workflows in "data notebooks" (e.g. ipynb – the python notebook), which combine prose and code and can be inspected and

	edited through a browser and executed server-side are becoming popular very fast. These seem ideal means to accompany the solutions in Marketplace, with executable code (where possible).

Table 25: Components of the system architecture

Design Decisions – Motivation and Rationale

In the following, we reflect on a few alternative design and implementation options which motivate our decisions.

Wikidata-based Architecture

An alternative overall architecture was considered, centred around a Wikidata-based Knowledge Graph and Wikibase[66]. The rationale was that information envisaged as part of the Marketplace partially is presented in Wikidata and if not, we could contribute this information, so we would essentially be curating a Wikidata-subgraph overlapping with the scope of the Marketplace. The underlying technology comes with extensive means for massively collaborative curation, complete transparency of the provenance of information and an array of tools (bots) for automatic checks of data quality and consistency, as well as solutions for exploiting and exploring the generated data, coming natively in RDF. The reasons to stay with a "traditional" Java-server/Javascript-client architecture: limitations on the data model, functional limitations brought about by the default technology stack, but also the mismatch regarding the available skill set in the project team.

Server-Client Communication REST API vs. GraphQL

Taking into account the experience and skills of the project team, Java Spring and React frameworks were selected as the technologies for the implementation of the server and client-side of the solution. Within React framework, GraphQL[67] emerges as an alternative to classical REST APIs. We decided to use classical REST APIs, to avoid technological lock-in. With REST API as a sole contract between server and client, both the server- and client-side implementation could be replaced if deemed necessary. Also, third-party clients can more easily access Marketplace information by consuming the REST API. The API will be described following the well-established industry standard OpenAPI[68].

Persistence layer

There are many technological options for implementing the persistent storage of information in the application. Next to traditional relational databases, there is a variety of "NoSQL" data models, including key-value, document, columnar and graph formats. JSON as an exchange format between server and client would justify a document-based solution like mongo-db, the expected network-like interconnected nature of the data is an argument for a graph-based database like Blazegraph, Neo4j, or GraphDB. In this respect, no final decision was made yet, but the preference is to choose also here the traditional approach with a relational database with Hibernate[69] as a flexible connector (object-relational mapper) between Java and the database system.

66 https://www.mediawiki.org/wiki/Wikibase
67 https://graphql.org/
68 https://www.openapis.org/
69 https://hibernate.org/

This primary persistence layer will be accompanied by a dedicated search & indexing engine, Apache Solr[70], which offers powerful querying and faceting capabilities.
As it is planned to provide the data also in an RDF representation, there will be a triple store, allowing querying via a SPARQL endpoint. Here, Blazegraph[71] is the default choice based on the team's experience, but we are still investigating other options like Virtuoso[72], or AllegroGraph[73] or GraphDB/Ontotext[74].

Vocabulary Management and Information Extraction

An important part of the curation process will be the management of vocabularies, both existing and custom ones. The platform PoolParty[75] developed by one of the project partners, the Semantic Web Company, is a mature product for vocabulary management & population through information extraction from texts, accompanied by the component Unified Views[76] for mapping metadata schemas as part of the aggregation process from heterogeneous sources.

For the task of vocabulary management, there is also a number of other technical solutions, some of which are even maintained and developed by some other project partners: iQvoc[77], VocBench[78], CESSDA Vocabulary Service[79] (CESSDA), ACDH Vocabs Editor (OEAW), THEMAS[80] (FORTH), OpenSKOS[81] (CLARIN).

It is yet to be decided if any of these solutions will be adopted, or if rather appropriate components will be developed as part of the overall applications. As a first step, feasibility tests will be conducted concentrating on the most comprehensive solution PoolParty.

Integration with EOSC

SSHOC MP being a discovery solution itself and aimed to become part of the EOSC; we try to sketch its envisioned position and role in the EOSC landscape. Given that EOSC itself is in the making and many aspects still need to be settled, we can only give tentative answers subject to adjustment as the EOSC evolves and as WP7's work pursues. The continuous monitoring and adaptation will be part of the work in tasks T7.3 Interoperability and T7.4 Governance:

1. SSHOC MP will become one of the EOSC services and as such will be represented in the EOSC catalogue/marketplace[82]. Considering current categorisation, it would fit under "sharing & discovery" services dedicated to the SSH communities.

70 http://lucene.apache.org/solr/
71 https://www.blazegraph.com/
72 https://virtuoso.openlinksw.com/
73 https://franz.com/agraph/allegrograph/
74 https://www.ontotext.com/products/graphdb/
75 https://www.poolparty.biz/
76 https://www.poolparty.biz/unifiedviews
77 http://iqvoc.net
78 http://vocbench.uniroma2.it/
79 https://vocabularies.cessda.eu/
80 https://www.ics.forth.gr/isl/index_main.php?l=e&c=243
81 http://openskos.org/
82 In our understanding, the EOSC catalogue will allow a presentation of all the services submitted by the providers and approved by the EOSC governance, and the EOSC Marketplace will provide an additional layer allowing to order services directly thanks to the onboarding extension.

2. Given that SSHOC MP itself is a discovery service (a catalogue), the question between the relation and potential overlap of data or functionality of the SSHOC MP and the EOSC Catalogue arises. As a comprehensive registry of scholarly practices in SSH domain, it is meant to feature tools, services, training materials, and workflows relevant for the SSH community. Assuming that the EOSC Catalogue and Marketplace will contain mainly information about services (however covering all disciplines), we expect that services for SSH will be entities relevant for both solutions, i.e. they should appear in both catalogues.

   Therefore, it is paramount to determine the flow of information between the two systems. It is yet to be determined if the information about these services will be first registered in the central catalogue (on-boarding) and propagated to SSHOC MP through some automated mechanism, or vice-versa. In any case, any duplication of effort for the providers must be avoided when submitting information about their services and resources.
3. SSHOC MP aims to make use of EOSC Federating Core, especially the Federated Identity (AAI) services and the Helpdesk.

Further Development

Implementation

Our key assumption in the design and development process is to follow best practices of agile and User Centered Design (UCD) approaches. In the context of software development, we will take into account recommendations given for the SSHOC project (based mainly on Jiménez et al., 2017) and the Technical Reference[83] jointly developed by CESSDA, CLARIN & DARIAH within the European Research Infrastructure Software Engineers' (EURISE) Network[84]. The development process will be iterative and based on well-known SCRUM methodology. The design process will be aligned with software development activities to make sure that developed features meet users' expectations. In the context of UCD, we plan to divide it into several stages and execute them in an iterative manner. These stages include investigation of the context of use, gathering requirements, designing solutions as well as evaluation. We have already investigated the context of use and interviewed potential users as well as summarised the interviews as User Stories. This research and analysis of the user needs will allow our team (including UX designer) to create appropriate mock-ups. At this stage, we will design solutions by creating sketches in the form of low fidelity mock-ups. These mock-ups will allow us to focus on the most important functionalities of the project and lead us to a functional project consisting of the most important screens. These will still be low fidelity mock-ups but covering many subpages containing the most important functions. The model will be consulted with the team (e.g. front-end & back-end developers as well as representatives of target user groups). Their comments and suggestions will be taken into account in the mock-up design process, to be finally transformed into high fidelity mock-ups. These will be verified in tests with several potential users, and if necessary, A/B tests will be conducted. Thanks to that, users will help us identify needed changes in the design (e.g. missing parts, unnecessary elements, intuitiveness). It is important to note that changes in the design will be consulted with the development team to assure the feasibility of the solution in the context of technical and time constraints. Based on the developed mock-ups, our UI designer will prepare the graphic design. There will be another round of consultation with the team and the users, and finally, its implementation will be initiated. Once the implementation is executed, the evaluation will begin in the form of final tests with the users. After that, the next iteration of the design process will begin in order to add new features or adjust existing ones according to users' feedback.

Focusing on end-users is critical in the UCD process. Therefore, we would like to involve as many users as possible and also use already planned SSHOC meetings to reach them. In this regard, we will be in close relation with WP2 and WP6, and we plan to have a user testing session in the context of the first release of the SSHOC MP planned in June 2020.

Sustainability – Governance and Curation

The question of the SSHOC MP's sustainability is addressed, among others, through T7.4 of WP7 which relates to the future governance and curation of this platform with two dedicated deliverables. To approach those topics, two Task Forces have been accordingly settled.

83 https://technical-reference.readthedocs.io/en/latest/
84 https://eurise-network.github.io/

Their work will feed upon this system specification document since it seemed difficult to discuss the governance and the curation of a Marketplace from which we couldn't yet know the precise goal and contents. However, although awaiting for this deliverable both Task Forces have started to identify some key aspects.

Regarding the governance of the SSHOC MP, it has been decided that the Governance Model's definition (and its subsequent Business Model) would be approached through hypotheses: each depending on the SSHOC MP inclusion's level within the EOSC and on the ERIC and other research infrastructures/ organisation involvement into its sustainability. The Governance Model will, of course, heavily depend on the Curation's needs – and vice versa.

As for the Curation, two main questions are being discussed. The first undergoing aspect relates to the sources (e.g. TAPoR) and resources (e.g. TAPoR's registered services/tools) that will populate the SSHOC MP. Indeed, it only makes sense to define what will compose the Marketplace before deciding about its curation process. A list of all existing catalogues and other resources has been created and is currently being used to list the available sources and resources. This list will permit to define the sources that will be harvested for the SSHOC MP.

Secondly, the extent to which the platform's curation would relate to users was approached. The community-oriented aspect has been at the heart of the SSHOC MP project since its beginning. It was therefore only natural to allow the community to take part in the platform's sustainability, in the light of its curation.

The curation will be approached through (1) automatic checks, (2) dedicated moderators/curators and (3) a community-oriented curation via micro-editing. The question remains as to how to engage users based on existing networks and in relation to other WPs (2–6).

The above-mentioned elements, to be still further refined, constitute the basis for approaching the SSHOC MP sustainability (Governance and Curation) to later defined "a set of criteria, and an assessment process to ensure qualitative content and promote longevity."[85]

[85] SSHOC Grant Agreement on T7.4

Conclusion

The purpose of this document is to deliver the system specification of the SSHOC MP, a service to be provided by the SSHOC project to its user communities. This service integrates and contextualises information about tools, services and datasets that are of relevance to the SSH communities. Furthermore, it will offer a close connection to the European Open Science Cloud through SSH-specific data and knowledge.

The document at hand delivers, in the beginning, a view on the state-of-the-art, followed by a detailed analysis and summary of the user requirements based on an extensive set of questionnaires. This requirements summary has been complemented by input from the SSHOC experts originating from their rich experience with designing, developing, and operating large service infrastructures. Clear prioritisation of the requirements defines the starting point for the service development.

The actual system specification consists primarily of the data model specification and the system design. These assets are described in depth in this document. Both the data model and the system design take as much previous work from other European projects into account as possible and ensure that the integration with the EOSC target is a primary concern at every step taken.

This deliverable has been delivered in project month 9 (September 2019) in order to synchronise as early as possible with the SSHOC development team based on a fully-fledged specification. This allows to work on an agreed basis in an agile manner, and task 7.1 will continue to develop the system specification based on the feedback from the various stakeholder groups and the results of the project-internal collaboration.

References

Auer, Sören 2018. Towards an Open Research Knowledge Graph (Version 1). Zenodo. http://doi.org/10.5281/zenodo.1157185

Constantopoulos, Panos & Pertsas, Vayianos 2019. From publications to knowledge graphs. 13th International Workshop on Information Search, Integration, and Personalization, Heraklion, 9–10 May 2019.

Dombrowski, Quinn 2014. What Ever Happened to Project Bamboo?, Literary and Linguistic Computing, Volume 29, Issue 3, September 2014, Pages 326–339, ⟨https://doi.org/10.1093/llc/fqu026⟩

Dombrowski, Quinn & Rockwell, Geoffrey. "The Directory Paradox". Forthcoming in Debates in Digital Humanities: Institutions, Infrastructures at the Interstices. Univ. of Minnesota Press. Eds. Anne McGrail et al. 2019.

Fathalla, Said & Vahdati, Sahar & Auer, Sören & Lange, Christoph 2017. Towards a Knowledge Graph Representing Research Findings by Semantifying Survey Articles. 315–327. ⟨https://doi.org/10.1007/978-3-319-67008-9_25⟩

Jiménez RC & Kuzak M & Alhamdoosh M et al. 2017. Four simple recommendations to encourage best practices in research software [version 1; peer review: 3 approved]. F1000Research 2017, 6:876 ⟨https://doi.org/10.12688/f1000research.11407.1⟩

Grant, Kaitlyn & Dombrowski, Quinn & Ranaweera, Kamal & Rodriquez-Arenas, Omar & Sinclair, Stefan & Rockwell, Geoffrey. "Absorbing DiRT: Tool Discovery in the Digital Age." Digital Studies/le Champ Numérique. Forthcoming.

de Leeuw, Lisa & Admiraal, Femmy & Ďurčo, Matej & Larousse, Nicolas & Mertens, Michael et al. 2017. D5.1 Report on Integrated Service!Needs: DARIAH (in kind) contributions – Concept and Procedures. [Other] DARIAH. 2017. ⟨https://hal.archives-ouvertes.fr/hal-01628733⟩

Raciti, Marco & Moranville, Yoann & Barthauer, Raisa & Buddenbohm, Stefan & Seillier, Dorian 2019. D5.4 – Implementation of a centralized helpdesk and marketplace mockup. [Research Report] DARIAH. 2019. ⟨https://hal.archives-ouvertes.fr/hal-02088278⟩

Annexes

Annex 1 – Interviews' distribution (gender, country of residence, research discipline)

Annex 2 – Questionnaire of the interviews

Annex 3 – List of User stories

Annex 1 – Distribution of interviews (by gender, country of residence, research discipline)

	Early stage researchers (ESR)		Experienced researchers	
	Usual suspect	Beyond the usual suspects	Usual suspect	Beyond the usual suspects
SH1 Individuals, Institutions and Markets (Economics, finance and management)	1. Female Germany/Belgium Economics	2. Male Germany Management	3. Female Germany Management	4. Male The Netherlands Political economy
SH2 Institutions, Values, Beliefs and Behaviour (Sociology, social anthropology, political science, law, communication, social studies of science and technology)	5. Male Germany Sociology	6. Male France Information and communication science 7. Male Belgium Political science and European studies	8. Female Spain/France Comparative politics	9. Male France Sociology 10. Male France Political science
SH3 Environment, Space and Population (Environmental studies, geography, demography, migration, regional and urban studies)	11. Female France, Cartography	12. X	13. X	14. Female France Geography
SH4 The Human Mind and Its Complexity (Cognitive science, psychology, linguistics, education)	15. Male Austria Linguistics language documentation	16. Female France Psychology	17. Female Austria Linguistics	18. Female France Psychology
SH5 Cultures and Cultural Production (Literature and philosophy, visual and performing arts, music, cultural and comparative studies)	19. Female United Kingdom Digital curation	20. Male Austria Literature studies	21. Female France Literature	22. Male Austria Media studies
SH6 The Study of the Human Past (Archaeology, history and memory)	23. Female France History and digital humanities	24. X	25. Female Norway Archaeology	26. X

Annex 2 – Questionnaire of the interviews

1. PROFILE OF THE PERSON INTERVIEWED

1. Could you specify your gender?
2. Could you specify your age?
3. What is your current profession?
    a. What is your research position?
    b. Since when are you working as a PhD candidate/researcher? (experience)
4. What's your research domain(s)/discipline(s)?
    a. Can you briefly describe the main lines of your research?
5. What is your country of residence?
    a. Where are you working? (country)
    b. What's the name of the institution you are working for?
6. Does the digital (significantly) impact your research? How do those digital aspects impact your research?

2. RESEARCHERS' DIGITAL HABITS WHEN CONDUCTING THEIR RESEARCH: CURRENT USE OF TOOLS/SERVICES/RESOURCES AND ATTEMPTS TO DISCOVER NEW ONES

Digital landscape

7. In the next questions we will talk about services, tools, data and data lifecycle to be sure that we are on the same page, I would like to share with you some definitions and examples:
    - a (digital) tool is "used for specific purposes in order to accomplish certain tasks or actions"[86]. In this interview we will also use the following terms to speak about tools:
        - a software is a tool accessible through an online channel (computer and/or a mobile device), but that **needs to be executed/run on the side of the user**. Examples: Gephi to analyse and visualise networks or AntConc to analyse texts.
        - a service is an application **accessible and directly usable via the internet**. Examples of services: GitHub, Zenodo...
    - resource: "all kinds of information, that can be the product of or used in a scholarly activity"[87] (broad sense: research data/datasets). A resource can have different forms: text, image, video recordings, audio recordings, interactive resources (quizzes, interactive maps and diagrams, etc)[88].
        - data lifecycle: description of the different steps of the research process and what happens to data at each step

Were you already familiar with these concepts before our interview? If yes, do you have another understanding of these concepts than the definition we are using?

8. How experienced do you feel you are in the use of those digital tools to conduct research? What about your community?

[86] cf. Nemo classes definitions & examples and description
[87] it's actually the scope note for an "Information Resource" as it is described in the Nemo classes definitions
[88] This definition comes from HAL, D5.1 Report on Integrated Service! Needs: DARIAH (in kind) contributions – Concept and Procedures : https://hal.archives-ouvertes.fr/hal-01628733v2/document

Digital research habits

9. Do you use specific tools to conduct your research?

a. If yes: What are the three main tools/services you're using? Names and types/functions of services/tools	b. If no: why? (Examples of potential answers: Not in my habits; Lack of training; I find those too complex → Why?) c. If yes, but really common tools (like word, excel or emails client)
10. How often do you use those tools/services (Daily, weekly, monthly basis…)? 11. Did you receive any training for the use of this tool? a. If yes, delivered by whom? b. If not, how did you learn to use it? 12. What are you doing when you need help using this service/tool? Do you have a resource person/community? Do you find answers online and with tutorials or forums? 13. During your daily use of those tools/services, is there a point when those tools/services impede on your research much more than they support it? Could you tell us about the last time you noticed it? How do they impede on your research (examples: lost time trying to find a solution, not the appropriate tool, etc)?	14. Do you think some tools could facilitate/improve your research process? What would be the ideal tool to facilitate/improve your research? 15. What would you need to have a better understanding of potential tools or services that could be useful for your research?

16. If you are familiar with the idea of research data life cycles: Which part of your research data life cycle[89] isn't covered well with
    a. the tools and services you use?
    b. the resources you use?
17. If you are familiar with the FAIR[90] data principles:
    a. Do the tools/services you use support you in creating FAIR data?
    b. Do the resources you use support you in creating FAIR data?

Digital discoveries

89 https://www.ddialliance.org/training/why-use-ddi
90 https://www.go-fair.org/fair-principles/

18. Let's say you need to find new software/services to conduct your research because the one you just mentioned is not sufficient enough. How would you proceed?
19. In your experience, how do researchers in your community choose their tools? (My study community recommended it / already used it; I researched it online; I was trained to use that tool; it was a part of my current education training course...)
20. How long did it take you to decide to use a tool the last time you had to do it? Did you compare several software/services? Were some criteria more important than others (paying service, free software, institutional incentive, research community incentive...)?
    a. In case the researcher decided to create his own tool to answer to his need, try to ask deeper to what services/infrastructures he had to talk to and how did he find information about these.
21. Was it difficult to find the appropriate tool/resource? (time-consuming, boring…) → How/why?
22. As we explained at the beginning of the interview, we are working on an SSH Open Marketplace to help SSH researchers with the digital aspects of their work. Do you know some similar services? If yes, which ones?
23. With regards to your own experience when in need to discover new resources/services/tools, which mistake shouldn't the SSHOC Marketplace reproduce? What would you (ideally) expect of this SSHOC Marketplace?

Conclusion

24. Would you be interested in being part of a "trusted user group"/follow-up user group for the Marketplace?

Annex 3 – List of User stories

Based on the interview (number of the interview)	User story number	As a (role)	I want to (something)	So that (benefit)	Test case and/or Input-/Output Data
1	1.1	young researcher in economics	find additional data to complete my study	work with a complete and coherent dataset	aggregator for data with filters by regions or disciplines or topics
4	4.1	experienced researcher in political economy	a platform (SSHOC MP) really easy-to-use, simple for non experimented researchers and well-established (used by the entire research community)	a researcher not experienced, digitally-speaking, can have an easy access the platform	intuitive platform, user-friendly
5	5.1	young researcher in sociology	test the different tools/services that could answer my need (because even if I lose time, I favour the comfort of use)	I can have an overview of the different options, try them all and decide by myself	there should be a selection (but not too restrictive) of tools/services to answer to a specific need.
5	5.2	young researcher in sociology	find alternatives to the most "famous" tools/services	use services/tools that answer to my principles (Open, free...)	present contents as alternatives to other more well-known ones
5	5.3	young researcher in sociology	use up-to-date solutions	I can see if new tools/services have been released	updated contents and recent comments and rates that ensure the quality of the platform
6	6.1	young researcher in information and communication science	be better informed on the basic "technical" functioning of a tool	I can not only analyse the results offered by a given tool but also understand the main lines how I obtain those very results/how the tool obtained those.	training, forums, etc.
6	6.2	young researcher in information and communication science	see a platform (SSHOC MP) very easy to use, sorted through specific categories and with a lot of choices/solutions.	I can easily access the information that I need	be suggested several categories that you can cross to find a pertinent result: disciplines/type of tools/price, etc
6	6.3	young researcher in information and communication science	be clearly informed on a suggested tool price and, when possible, be informed on how I could get it financed by my lab	I can find a way to get it	a box/comment section explaining that
6	6.4	young researcher in information and communication science	see a platform (SSHOC MP) that also includes a "gap analysis" component	when you couldn't find a solution to your specific problem, the platform takes it into account and	a dedicated space on the platform to do so

				will later provide the solution	
6	6.5	young researcher in information and communication science	see a platform (SSHOC MP) here to help researcher both by providing solutions, but also by allowing researchers to leave reviews	one can find the appropriate solution to his/her specific issue	forum, tutorials, boot popping out "How can I help you?"
7	7.1	young researcher in political sciences and European studies	have a better view of what a tool can or cannot do for me save some time when starting to use a new tool or service	I can save some time when starting to use a new tool or service	description and contextualisation of tools and services, also pointing out the limits.
7	7.2	young researcher in political sciences and European studies	know what kind of tools the researchers in my community are using	I can identify what is probably the best tool for me as well	researchers views or comments, quotations and links to forum
7	7.3	young researcher in political sciences and European studies	find some specific answers about the use of a tool	I can quickly fix a specific issue I have with a tool or service	links to forum, or tutorials. Or lists of the most common issues with a tool or service
8	8.1	experienced researcher in comparative politics	easily find what I'm looking for	I can save time in my search for new services	user-friendly and easy-to-use platform / "for dummies" approach / extremely intuitive also for people with no computer science training or low digital skills
8	8.2	experienced researcher in comparative politics	easily understand the environment of the Marketplace	I know what tools/services can be used to respect standards recommended by other organisations within the research landscape (RDA, DDI...)	give information on the context of use and provide links to organisations, projects... connected
8	8.3	experienced researcher in comparative politics	know about the starting cost of a tool	I can decide to use it or not depending on time and help I'll get before being able to "master" it	provide information on the level of complexity or access required, and links to training materials and/or use cases
10	10.1	experienced researcher in political science	have a real choice and an effective peer review with regards to the solutions that could be proposed to me (SSHOC MP)	I can be reassured in the choice that I make regarding digital solutions and more broadly speaking in the platform that I use	important number of peers' reviews on a tool, wide use of this platform within the SSH the community.
10	10.2	experienced researcher in political science	find solutions in terms of GDPR compliance with my research process	I can more easily know how to be GDPR compliant when conducting my research	GDPR section/explanations regarding its integration to a research process
14	14.1	researcher in geography	have easy access to the SSHOC MP	I don't bother in identifying several times and can have	no multiple identifications required

				quicker access to the information that I need	
14	14.2	researcher in geography	use a platform respectful in terms of data privacy settings	I can trust the platform that I use	a search bar that doesn't collect and store the information I entered; that respects my anonymity
14	14.3	researcher in geography	use a platform easy of access and free of use	I can easily find what I need and come back later on to use this platform	user-friendly and free platform
14	14.4	researcher in geography	use the SSHOC MP to find free solutions to my problem	I can put into practice those suggested solutions	propose free solutions (tools)
15	15.1	early stage researcher in linguistics/language documentation	analyse a resource as a non-expert.	I can interpret the resource and understand how this resource was built (which tools were used, how they were combined).	point to a resource (or describe the resource) and get as much information about it as possible (by combining information that is available in the Open Marketplace).
15	15.1	early stage researcher in linguistics/language documentation	get a hint that for my resources and/or my research method, there are ethical considerations.	if I'm not aware of this ethical consideration, I start to respect them.	searching for material on resources/research methods that do have ethical implications should lead to hints on this. It can also be that these ethical considerations are only valid on the national level (e.g. when publishing personal information like names of people).
15	15.11	early stage researcher in linguistics/language documentation	get a standard and easy to run a workflow for my research approach.	I don't need to invest much time to learn new digital skills.	show a preferred and easily applicable way to support a research approach, e.g. choosing actions that the workflow should fulfil and give back the most often used tools to fulfil this.
15	15.12	early stage researcher in linguistics/language documentation	have recommendations on what I can build up on top of my tool set.	to get new insights into my resources by seeing potential extensions for my workflow. to extend my skills based on my experiences.	choose a tool and get information about other tools that can be combined with my tool.
15	15.13	early stage researcher in linguistics/language documentation	see what a tool can do with my specific resources.	I can quickly compare different tools and estimate if the tool does what I like it to do.	choose the resources I work with and show the tools that can handle them in a way that it is comparable based on the functions of the tools.
15	15.2	early stage researcher in linguistics/language documentation	find out if there is already a script that helps me in my research work.	I don't need to do repetitive work manually or code a script on my own.	giving information about the action, I like to process and the environment I have access to. It should give me a list of scripts that I can re-use.

15	15.3	early stage researcher in linguistics/language documentation	get an overview of all the functions that I can do with a tool.	I'm aware of all of the aspects that are possible to do with a tool so that I can discover new ways to process my resources.	choose a tool and get all of the functions that the tool support and all of the workflows, where this tool is used.
15	15.4	early stage researcher in linguistics/language documentation	see different solutions for a research workflow.	I can optimise my research workflow, e.g. by adapting a better- suited tool.	describe (a part of) the research workflow and get all of the available solutions.
15	15.5	early stage researcher in linguistics/language documentation	be pointed to the best place where I can ask my question on a tool/method and get an answer.	I have a better chance to get an answer.	describe the domain of the question and get social contact points, e.g. a question on TEI should point to the TEI website, TEI mailing list, TEI discussion forum. Sometimes it can also point to a general Q&A platform like Stackoverflow
15	15.6	early stage researcher in linguistics/language documentation	get information about infrastructures, where I can perform my research action/process my resources/ run my tools.	I can do the processing.	describe the action to process and/or the tool to be used and get a list of infrastructures/services that can be used, e.g. I need an Apache Solr, which service is recommended.
15	15.7	early stage researcher in linguistics/language documentation	get trained data for a specific research question.	I can build a model for machine learning approaches.	choose research community and/or research scope and get information on how to get to training data, either platforms with data that fit the scope or researchers from the field that I can ask for their training data.
15	15.8	early stage researcher in linguistics/language documentation	see if there are changes between different versions of a tool.	I can estimate what an update means for my workflow.	choose a tool and see what changed between different versions of this tool. It would be interesting to see voices from the research communities there (and not only a changelog).
15	15.9	early stage researcher in linguistics/language documentation	find out where experts for a digital research method are.	I can get in contact with them, e.g. doing a course at their institute or ask them for consultancy.	get a list of experts, institutions, contact points, courses, etc. to a digital research method. In general, give a picture of the community that can be contacted/joined. maybe allow following a user in what she/he is doing (like in Facebook/Twitter/etc.)
16	16.1	young researcher in psychology	have access to information about research digital tools	I can discover new tools relevant for my research and thus gain in experience	simple training/information, clear and to the point for my specific case, not time-consuming

16	16.2	young researcher in psychology	use a pertinent search engine/service	I can discover relevant resources related to my research topic	a search engine service allowing to select specific categories to get a very pertinent result. A search engine that would understand the meaning of my research.
16	16.3	young researcher in psychology	use a "multilingual" search engine service	I can find resources/datasets in other unexpected languages	entering a research topic in a given language (French, English) and get relevant results in those languages and others if possible (Spanish).
16	16.4	young researcher in psychology	be clearly informed on the discussed/suggested digital tool's price	I can prioritise the use of free ones, following an open access policy	one of the descriptive criteria of a contextualised tool should be the price
16	16.5	young researcher in psychology	exchange with other researchers from my discipline	I can discuss my research topic/research methods	forum
17	17.1	experienced researcher in linguistics	find out where to get already processed/structured resources.	I can re-use this resource. I can compare these resources with my approaches. I don't need to do the processing.	list platforms where processed/structured resources can be gathered (differed by resource type/research community) respective give me the information, how to differ on a platform between raw and structured resources.
17	17.2	experienced researcher in linguistics	get informed what tools are preferred to apply on specific resources. I also like to have information which data formats a tool can process, import, and export.	I can try out this tool for my resources. I can point research colleagues to a list of tools they can use with their data.	list tools that can process a resource, distinct by the format of resources, state of resources, research community recommendations. I may prefer a curated list of tools used in my research community (there are sometimes blog posts/inventories from a research community that give such background information: give me a link to them or integrate them in the search result).
17	17.3	experienced researcher in linguistics	have a direct link to help files of tools.	I can consult this help files if I need to solve a problem.	have a direct link to the different helping resources of a tool. If there is no help resource, it should also be stated. If there are different versions of a tool with different help files, give me this information.
17	17.4	experienced researcher in linguistics	find out if there is/is not a tool for a specific action like conversion between two data formats.	if there is a tool, I can process the action. if there is no tool I know about that and either need to develop a new tool or choose a different data format.	choose the action "conversion", enter input and output format and get information either that there is a tool or that there is no tool.

17	17.5	experienced researcher in linguistics	get support in making raw data becoming FAIR data.	I can work for my research with structured FAIR data.	I have not well structured "raw" data (e.g. lines of text, word files) and I may know which output format I like to have (e.g. XML-TEI) but I'm also open for other formats. In the end,i,ilt should fit the standards of my research community. I like to get either direct support or documentation (like checklists) on how to turn my data into FAIR data.
17	17.6	experienced researcher in linguistics	find out – based on recent research papers – which tools are used and for what they are used in my research community	I can try out this tool on my own. I can get in contact with the authors of the research paper to discuss with them their use of a tool.	get a list of tools that are mentioned in recent research papers in my community. Link also to the research paper and give me the context of the paper (abstract, research question, for what is the tool/are the tools used for).
17	17.7	experienced researcher in linguistics	have an up to date description of the functionality of a tool for my research community.	I can see at a glance if this tool suits my research needs. I find a handful of tools that covers many useful functions for my research approach so that I don't need to handle many tools. I don't have outdated information on a tool.	give a brief description for what a tool is used, what are the advantages/disadvantages. I don't like to have advertising text; I like to have honest information from my research community.
17	17.8	experienced researcher in linguistics	know if I can try out/test a tool beforehand.	I can easily test if the tool does what is said about it without investing a lot of time for setup and without the need to buy it before testing it. Ideally, there is a button where I can start the tool in a virtual machine, try out basic functionality before I decide to download/buy it for my workflow.	give information on trial versions of the tool (especially if it is a commercial tool) or test environments where I can try out the tool.
18	18.1	experienced researcher in psychology	a platform (SSHOC MP) thoroughly organised between the different disciplines that compose social sciences and humanities	each discipline's issue/component can be distinctively addressed	thorough organisation in disciplines/sub-disciplines.
18	18.2	experienced researcher in psychology	know who is managing this platform (SSHOC MP) and be aware of its data privacy policy	I can be totally reassured and start using this platform	public ownership preferably, transparent and careful in its data management policy

19	19.1	young researcher in digital curation	have a quick answer when I don't know how to deal with a tool	I can save time	provide direct contacts and avoid chatbot – link to youtube videos – professional help pages (cf. Adobe VS Microsoft)
19	19.2	young researcher in digital curation	understand the minimal conditions I have to follow to do something when I'm not a specialist	I can respect existing standards like a professional	provide simple contents (e.g. checklist of basic questions) for identified topics of interest for a research community
20	20.1	(early stage) researcher in literature studies	have a collaborative working environment.	our team can evaluate and comment tools. our team can communicate new findings on the platform. our team can organise useful resources.	create a team account, invite people and let them add interesting tools/papers/tutorials so that everyone from the team can find and easily access them. It should also be possible to differ between categories like tools/resources to evaluate, tools/resources to watch, etc. Having a way to prioritise these tools/resources and compare them.
20	20.1	(early stage) researcher in literature studies	register my research data platform.	others can contact me for contribution. others can link to/re-use my data.	have an input mask where all necessary information on a research data platform can be entered, so that researchers working with/on similar data find it.
20	20.11	(early stage) researcher in literature studies	have tailored and manageable information based on my research profile.	I get all the important information at a glance without having long lists where I need to scroll through.	a tidy workspace and not too much information there. possibility to define researcher profiles.
20	20.12	(early stage) researcher in literature studies	have a dynamic way to explore new tools/resources.	I find better solutions than the ones I currently use. I can expand my horizon with surprising solutions.	besides a standard workspace with all information based on the researchers' profile, have an explore function, where solutions outside the profile are shown (kind of people going there are sometimes also going to this tool, that is not in your profile).
20	20.13	(early stage) researcher in literature studies	be asked different questions and based on this get an answer, e.g. we recommend this tool to you.	come to an answer without in-detail know-how beforehand.	have Q&A-trees where people start with a general question and end with detailed questions so that an answer can be delivered. This can also be a bot.
20	20.2	(early stage) researcher in literature studies	share my learning successes.	others learn faster and don't repeat my mistakes. others can comment on my approaches.	have a way to interact with others, e.g. adding comments (differ between questions, learning success, links, etc.) and apply them to tools/resources or to general topics.

				I get recognition for my learning success.	
20	20.3	(early stage) researcher in literature studies	find someone who helps me in developing or bug fixing a tool or in converting a resource to a new tool.	I can further use my tool/resources in regard to new approaches I like to implement.	either choose an existing tool or add information about the individualised tool (maybe a self-development or a commissioned work) and have a way to communicate which support is needed and whom to contact (also add conditions for the job).
20	20.4	(early stage) researcher in literature studies	find out the effect on my research workflow, if I change a part of it, e.g. replace a tool.	based on the effort it takes, I can decide if I change a part of my research workflow.	have information on the effect of a tool on the research data life cycle, e.g. this tool makes it easy to export data into a long-term preservation repository vs this tool is not bound to such export. more sophisticated: users can model their research workflow in the platform and based on this, see the effects of a replacement, e.g. if you replace this tool with this one, you will need another tool, because the new tool does not have a specific function of the old tool.
20	20.5	(early stage) researcher in literature studies	understand why I should prefer a standardised solution (e.g. format, workflow) to my solution.	I'll be convinced to use a standardised solution.	describe a solution and compare it to similar solutions, where the most common/standardised solutions are highlighted.
20	20.6	(early stage) researcher in literature studies	play around with a tool.	I can try out if this tool fits my needs better than the one that I currently use and if it still supports all of my workflow.	give information on how to try out a tool. ideally, have a sandbox of the tool, allowing to work with the resources of the researcher.
20	20.7	(early stage) researcher in literature studies	get in contact with a local institution that supports me in establishing my digital workflow.	I get individual recommendations from a local institution with the possibility to have a face to face meeting. I can initiate cooperation, e.g. for a funding application.	declare a geographic scope, the research community and the digital workflow and get a list of institutions that have the expertise and can be contacted.

20	20.8	(early stage) researcher in literature studies	have an exchange with researchers outside my research community.	I get different perspectives on my research approach/digital workflow and discover alternatives.	having a recommendation system to get in touch with researchers that are not in the research community of the researcher but do have similar workflows.
20	20.9	(early stage) researcher in literature studies	be pointed to platforms, where data similar to my research data is collected.	I can link to/re-use this data. I can contribute to this platform (instead of building up one by my own).	describing my research data should lead to a list of platforms, where such data is collected. It should also show at first hand, how this data can be gathered and/or how someone can join this platform.
21	21.1	experienced researcher in literature	be clearly documented on a tool's possible update/obsolescence when I look for a new one	I don't lose time when trying to find new tools to support my research	within the description of the tool, but also through a good curation of the platform
21	21.2	experienced research in literature	be able to look for the tool that I need based on a classification sorting out the tools by their "families"/functions	I can find the tool that I actually need based on the function I need it to fulfil	tags, identification by keywords, categorisation
21	21.3	experienced researcher in literature	to have access to some documentation about the tools that I want to start using / that I'm using	I can learn to use it properly	tutorials, screenshots, any other documents explaining how to use a tool
21	21.4	experienced researcher in literature	to have access to some documentation (tutorials, screenshots) explaining how to use a tool that isn't written by computer engineers	I can really understand how to use it since the tutorials written by engineers tend to lack some "basic" information	priority to tutorials/documentations written by the SSH community itself
21	21.5	experienced researcher in literature	have access / be able to discover tools with a good front-end user interface's development in priority	I can use it more easily but also show/recommend it to my peers/community	recommendation based on the "user-friendly" aspect of a tool
22	22.1	researcher in media studies	build up basic knowledge in using digital tools.	I can see the benefit in using digital tools for my research.	have a basic introduction on the integration of digital tools in research workflows and the benefits of using digital tools.
22	22.2	researcher in media studies	get tips and advice on selected tools that support my research process.	I can shorten my research process and minimise mistakes.	based on the profile of a researcher, give tips and advice on tools that other researcher's successful use to support their research process.
22	22.3	researcher in media studies	formulate a problem that occurred in a project where I need a solution.	I can find and apply a solution and solve the problem quickly and easy.	support problem-oriented search like: I have digitised material that has no OCR, what can I do to get a text version?

22	22.4	researcher in media studies	get push messages on tools/solutions that fit my profile.	I'm pointed to new tools that I can try out.	when there is a new tool registered, and it fits the researcher profile, send a push message/e-mail.
22	22.5	researcher in media studies	have a contextualised and meaningful response when I search for something without a result.	I know why I didn't get a result. I can reformulate my search.	if there is no result on a search query, give background information on the reasons (e.g. we don't collect information on tools from your research community) and propose alternative queries, that give results.
23	23.1	young researcher in history and digital humanities	have an easy access/discovery of user-friendly tools relevant for the "collecting" and "processing" phases of the research data life cycle	I can cover those aspects much more easily and that our research needs can be covered	need to develop them or at least to be able to discover them (user-friendly aspect + data life cycle)
23	23.2	young researcher in history and digital humanities	have access to a genuinely open Marketplace proposing all kinds of existing solutions in the SSH community and not only limited to the ERIC community	I actually can be recommended all existing solutions: tools, tutorials, resources, etc.	not limited only to the ERICs' community existing tools/solutions and not addressed only to this very community
23	23.3	young researcher in history and digital humanities	have access to a centralised platform offering a catalogue of all existing solutions depending on a research's specific aspects	I can find the appropriate solution to my specific problem through one unique platform	a single platform offering the most exhaustive solutions through catalogues of datasets, tools, tutorials for the SSH
23	23.4	young researcher in history and digital humanities	be able to precisely define my specific research problem on a platform	I can be proposed the most pertinent solutions	a profile section organised by categories and allowing to define your situation and your potential research problems precisely.
23	23.5	young researcher in history and digital humanities	have a centralised access to datasets	I can be sure to be able to find all the data that I need through one platform	a platform centralising all existing datasets / their catalogues
23	23.6	young researcher in history and digital humanities	be offered a contextualized solution to my research problem	I can sort out my research problematic	contextualisation : a tool related to an academic article to tutorials, etc.