Research and Innovation Action

Social Sciences & Humanities Open Cloud

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

Report on **Milestone 42**

Marketplace – alpha release

Dissemination Level	PU
Due Date of Milestone	30/06/20 (M18)
Actual Achievement Date	**30/06/20**
Lead Beneficiary/LTP	5. DARIAH ERIC
Work Package	WP7 Creating the SSH Open Marketplace
Task	NA
Version	V1.0
Number of Pages	p.1 – p.9

Abstract:

This document reports on the achievement of SSHOC Milestone 42, the alpha release of the SSH Open Marketplace. It explains the role of the first release in the development process of the application as well as the community efforts deployed to ensure the uptake of the SSH Open Marketplace by SSH research communities.

Author List

Organisation	Name	Contact Information
DARIAH	Laure Barbot	laure.barbot@dariah.eu
DARIAH	Yoann Moranville	yoann.moranville@dariah.eu
DARIAH/UGOE	Stefan Buddenbohm	buddenbohm@sub.uni-goettingen.de
DARIAH/OEAW	Klaus Illmayer	klaus.illmayer@oeaw.ac.at
DARIAH/OEAW	Matej Ďurčo	matej.durco@oeaw.ac.at

1. Introduction

Three different releases of the SSH Open Marketplace are planned during the SSHOC project. The three corresponding milestones (MS42, 43 and 44) ensure the progress and the coherence of the various tasks contributing to the creation of the SSH Open Marketplace. MS 42 "Marketplace - alpha release" is the first one and is described as following in SSHOC Grant Agreement: *"Initial application deployed, populated with initial set of data, and with initial connectivity to selected external components available for internal project review"*. This MS report describes the different aspects of SSHOC Work Packages activities that contributed to achieve this milestone: mainly based on WP7 tasks, the alpha release of the SSH Open Marketplace is also an example of a SSHOC cross-WPs collaboration, as it substantially involves the expertise and collaboration from WPs 2, 3, 6 and 9.

2. Description of the Milestone

2.1. Role of the Milestone

The alpha release of the SSH Open Marketplace allows WP7 to clarify and verify its understanding of community involvement while the application is still under development. This milestone is a good opportunity to implement several key functionalities of the SSH Open Marketplace, test the data ingestion pipeline with a first set of content and prepare the curation components and the first building blocks of the curation policy. After the deployment of a Minimal Viable Product (MVP) in December 2019, "i.e. a running application which demonstrates the connectivity between the main components and exhibits minimal functionality"[1], and before the public release planned for the end of 2020, this alpha release of the SSH Open Marketplace presents an

---

[1] cf. description provided in the SSHOC Deliverable 1.3 First Annual Progress & Activity Report [10.07.2020]

instance of the application showcasing manually curated items based on user stories and accessible for testers and for an internal project review[2].

A series of events and community involvement activities, co-organised with WP2 and WP6, was planned to ensure an adequate level of communication around the alpha release and to pave the way for the public release.

Furthermore, MS42 allows DARIAH, as WP lead, to evaluate and - if necessary - to adjust the work plans with regard to scheduling, quality of results and partners' involvement.

2.2. Means of verification

Means of verification for MS42, as stated in the SSHOC Grant Agreement, are the following: "Initial application deployed, populated with initial set of data, and with initial connectivity to selected external components available for internal project review". These different aspects are described in more detail below.

The initial version of the SSH Open Marketplace is deployed

Several key components were developed and implemented to create the Minimal Viable Product delivered at the end of 2019[3]. Based on this in multiple iterations the MVP was further developed to deliver the alpha version of the application as for MS42[4].

Two instances of the SSH Open Marketplace application are now running: a production instance released for testers and accessible at https://marketplace.sshopencloud.eu/ and a development instance[5] where next steps are implemented to prepare the beta version of the application to be released at the end of 2020.

The initial version of the SSH Open Marketplace implements the proposed UI-design and allows for a simple full-text search in the data and detail view for individual items. All this functionality of the client is based on the API of the server-side component. The API also already supports ingestion of data. For the test runs as the first real dataset tools from TAPoR (cf. next section below) were used to validate the suitability of the API definition.

Further development feedback from user-interface testing and experience ingesting further datasets induced numerous smaller adaptations and refinements to the data model and correspondingly to the API-definition.

---

[2] the following URL is shared with testers, but is not publicly advertised: marketplace.sshopencloud.eu [10.07.2020]

[3] cf. related milestone in the development instance: https://gitlab.gwdg.de/groups/sshoc/-/milestones/1 [10.07.2020]

[4] cf. related milestone in the development instance: https://gitlab.gwdg.de/groups/sshoc/-/milestones/4 [10.07.2020]

[5] accessible at: https://sshoc-marketplace.acdh-dev.oeaw.ac.at/ [10.07.2020]

The functionality of the user interface has been extended, e.g. display of images and media, as well as display of related items in detail view, filtering by different facets as additional entry points into the dataset on the frontpage, etc.

The SSH Open Marketplace populated with an initial set of data

Since the beginning of the project, sources to collect information and populate the SSH Open Marketplace have been identified and prioritised by task 7.3. In an initial onboarding phase (from January to June 2020), 1,500 individual items coming from three different sources have been ingested. This initial phase was the opportunity to experience different data formats and to design the first lines of the data ingestion workflow and the interoperability framework. The chosen sources cover heterogeneous sets of data and represent different approaches, how to distribute information on digital tools, training material and workflows especially in the humanities:

- TAPoR[6] is a gateway to tools used in sophisticated text analysis and retrieval. 1371 individual items - tools - coming from TAPoR were ingested into the SSH Open Marketplace.
- Programming Historian[7] publishes novice-friendly, peer-reviewed tutorials that help humanists learn a wide range of digital tools, techniques, and workflows to facilitate research and teaching.
- Standardization Survival Kit (SSK)[8] presents a collection of research use case scenarios illustrating best practices and the application of standards in Digital Humanities and Heritage research.

The ingestion of these three initial data sets allowed to review the data model and to gain experience on the challenges regarding data quality and FAIRness[9] of both the sources and the SSH Open Marketplace. One of the outcomes of the early phase of ingestion was an iterative workflow to adapt the data model of the SSH Open Marketplace to such experiences. Another important step was the implementation of an organised and collaborative way of analysing sources, mapping them to the data model, documenting the processing of data including the documentation of ignored data fields and building the ingestion path. These steps are a blueprint for every ingestion of a new source involving contribution by all partners.

The implementation of the ingestion path can be done in different ways. The SSH Open Marketplace provides a REST API[10] at the backend - built up by SSHOC partner DARIAH/PSNC - where data can not only be read but also inserted or updated. This approach allows technology-agnostic ingestion. Currently, the main ingestion is built up by SSHOC participant Semantic Web Company (SWC). Based upon their product PoolParty

---

[6] Tapor.ca home page http://tapor.ca/home [10.07.2020]

[7] Programming Historian home page https://programminghistorian.org/ [10.07.2020]

[8] SSK home page http://ssk.huma-num.fr/#/ [10.07.2020]

[9] See FAIR data principles, https://www.force11.org/group/fairgroup/fairprinciples [10.07.2020]

[10] REST stands for "Representational state transfer". There are many resources on REST APIs, as a good starting point this dedicated website can be consulted: https://restfulapi.net/ The Wikipedia entry https://en.wikipedia.org/wiki/Representational_state_transfer also serves as a good overview. [10.07.2020]

UnifiedViews[11] a dedicated ingestion pipeline for every source is created. This includes the gathering of the data at the source (e.g. via an API or a Git repository), the mapping of the data to the SSH Open Marketplace data model including transformation of parts of the data (e.g. connecting terms from the source to a vocabulary like TaDiRAH[12]) and the integration of the processed data into the SSH Open Marketplace via the API. This way was chosen for the bulk ingestion of the three sources. It also includes methods to check for updates or for creations of items. In contrast, SSHOC partner DARIAH/OEAW developed a script to ingest a selected set of manually enriched data. This script is coded in PHP and takes data from a spreadsheet, processes it and uses the SSH Open Marketplace API to insert the data. As such it is also a proof of concept that different ways of ingestion are technically possible. The chosen setup for the ingestion workflow therefore gives the flexibility of customized ingestion processes for all kinds of sources.
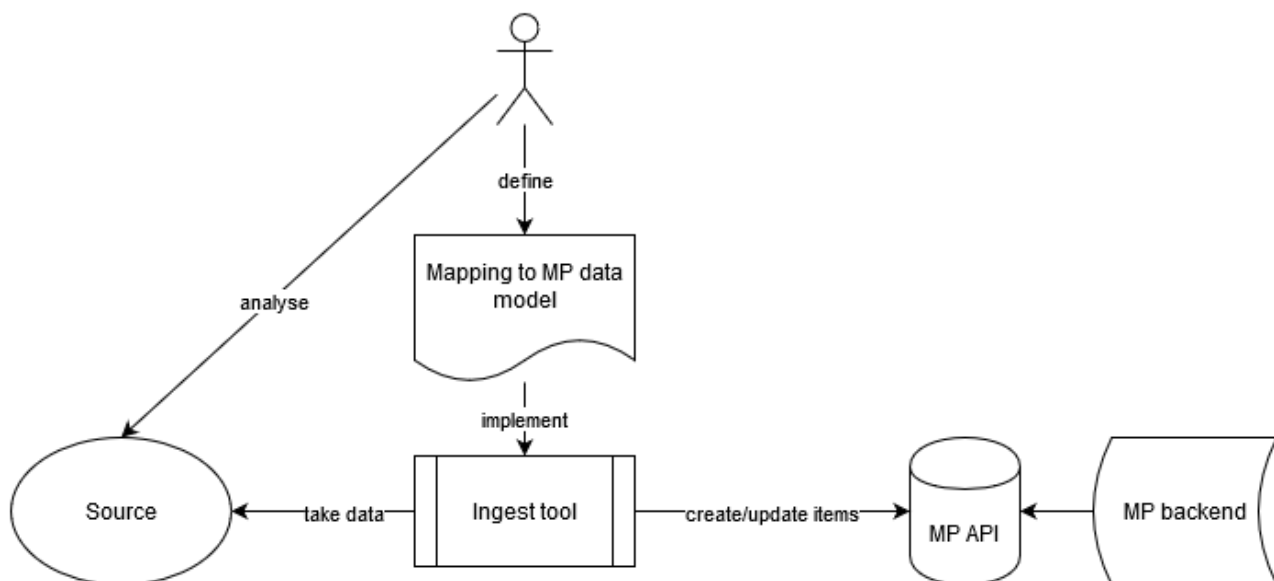

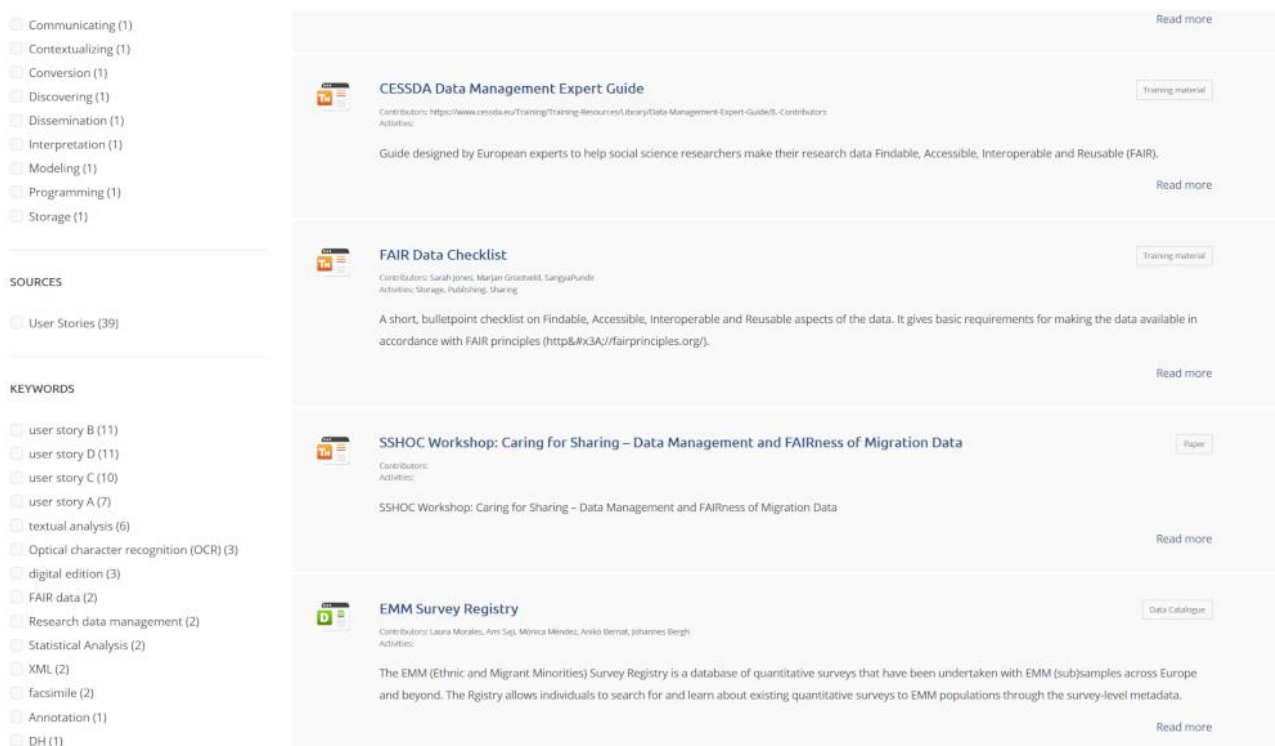Illustration: Overview on the different parts of the ingestion workflow

Illustration on public domain: CC0 1.0 (https://creativecommons.org/publicdomain/zero/1.0/

The workflow of the ingestion pipeline is applicable generically. Adaptations are necessary when it comes to process different types of data that a source holds. For the currently ingested sources the basic data types are tools for TAPoR, training materials for Programming Historian, and workflows for SSK. But in some cases there are also references to other data types, e.g. SSK features references to training material within its workflow descriptions. Identifying such references and dealing with multiple data types from a source are challenges for the analysing, mapping, and ingestion part of the workflow. It also affects curation of ingested data, especially in regard to identifying and merging similar items from different sources.

---

[11] https://www.poolparty.biz/agile-data-integration [10.07.2020]

[12] See http://tadirah.dariah.eu/vocab/index.php resp. https://github.com/dhtaxonomy/TaDiRAH [10.07.2020]

As curation components are not implemented yet, the decision was taken to not go public with an application showcasing 1500 non-curated items coming from the three mentioned sources. Instead, a manual curation exercise was set up: based on user stories creating fictional research workflows, items already ingested in the SSH Open Marketplace were identified, isolated and manually enriched. New items coming from not yet fully ingested sources were also integrated. The result of this work - done by the SSHOC partners in WP7 - was a spreadsheet that was ingested with the help of the mentioned script in a "clean" instance of the SSH Open Marketplace, showcased as the Alpha release.

Illustration: Screenshot of the alpha release version of the SSH Open Marketplace. It is possible to filter the items by one of the four user stories (see the keywords "user story A" to "user story D" in the illustration).

The implementation of an ingestion workflow was an important requirement for the successful launch of the SSH Open Marketplace alpha release. Having a stable way to insert data from different sources allows us to fill the SSH Open Marketplace with the necessary data to fulfil its mission. The next step is the development of the curation part of the SSH Open Marketplace and a seamless integration of the ingestion workflow into the curation workflow, including quality checks while analysing a new source and the further development of the ingestion infrastructure.

The SSH Open Marketplace presents initial connectivity to selected external components

The ingestion pipeline described in the previous section relies on the PoolParty UnifiedViews software to connect to external sources of data for the SSH Open Marketplace. In addition, the following external components were also set up to outsource some functionalities.

- User Management and Federated Identity (AAI)

    While the majority of the users of the SSH Open Marketplace are expected not to need authentication and being "viewers", for all the other roles identified in the curation workflow a user management is needed. Login for the SSH Open Marketplace will include a local solution as well as a federated solution, allowing users to login via their own institutions thanks to eduGAIN or via social accounts (e.g. Google, GitHub, …). At the Alpha release time, the AAI solution that is in place is EGI Check-in, on which the EOSC AAI partly relies. The reason for using EGI Check-in is that they propose a development environment that we could use directly and that the move to the more definitive solution (EOSC AAI) will simply be a matter of changing the configuration.

- Web usage analytic tool Matomo, run by DARIAH/OEAW.

The SSH Open Marketplace is available for internal project review

Conceptually the curators' and moderators' roles and community features are designed to motivate feedback and participation from the users. The challenge of this SSH Open Marketplace alpha release lies in finding the right balance between communication, consultation processes, and users' involvement. These aspects were primarily handled, discussed and solved within the WP7 community involvement task force (which also included the regular input from other WPs, particularly from WP2 and 6), and a modular approach was set up for the alpha release.

The involvement of research communities is crucial for the success of the SSH Open Marketplace. For this end the community involvement activities are aimed at two main objectives:

1. Requirements engineering: to ensure that the service matches with the research practices and requirements of the user communities, the D7.1 System Specification of the SSH Open Marketplace[13] has been created under the involvement of selected researchers. A community involvement task force had also been established to bring together and coordinate the further involvement activities. By various means (such as webinars, blog posts, tweets) the participatory design approach was presented to the community and feedback collected to further inform the development work.

2. Awareness/Dissemination: apart from constantly refining the specifications of the service by involving the potential users, the second main goal was to raise awareness for the SSH Open Marketplace and

---

[13] Laure Barbot, Yoan Moranville, Frank Fischer, Clara Petitfils, Matej Ďurčo, Klaus Illmayer, … Sotiris Karampatakis. (2019). SSHOC D7.1 System Specification - SSH Open Marketplace (Version 1.0). Zenodo.

establish it as future service in the public eye of the social sciences and humanities research communities.

These goals were pursued by various means and following, a few of them, are highlighted:

- Becoming an SSH Open Marketplace tester: a dedicated registration form was set up on the SSHOC website to invite users interested in becoming testers of the SSH Open Marketplace during its development.
- These testers will gain access to the Alpha release (the curated version of the SSH Open Marketplace) and will be asked to further contribute to its development, thanks to dedicated events or testing sessions that will be organised during the second half of 2020.
- For the preparation of the Alpha release a series of monthly news items informing on the development process, the means for community involvement and specific requests for feedback from the potential users was initiated[14]. These news items were published in close collaboration with WP2 to ensure the alignment with the communication strategy of the SSHOC project, and LIBER, DARIAH-EU and DARIAH-DE channels were also used to disseminate the posts.
- As an important and useful instrument to coordinate the consultation phase, WP2 and 7 have developed an open consultation platform (to be deployed on the SSHOC website), which answers to different communities' needs within the project. A dedicated area for consultation for the SSH Open Marketplace is included. This consultation will allow WP7 to collect feedback on topics such as the design of the SSH Open Marketplace, the sources used to populate the Marketplace, and the solutions that are under development to answer to the sustainability and the curation of the SSH Open Marketplace. Furthermore, the platform is designed as an extensible instrument, allowing WPs to adapt it to the possibly developing requirements of the consultation process.
- To announce and support the alpha release, a series of events have been organised in collaboration with WP2 and 6:
  - 17 June 2020: a SSHOC Twitter Take-over of the DARIAH' Twitter account[15]
  - 30 June 2020: a user workshop around the "agile development of the SSH Open Marketplace" involving four testers during the ICTeSSH Conference[16]

---

[14] The April post - https://www.sshopencloud.eu/news/ssh-open-marketplace-how-and-when-can-you-contribute - presents the timeline and possibility for contribution; the post released in May - https://www.sshopencloud.eu/news/ssh-open-marketplace-whats-it-you - highlights the content and sources of the SSH Open Marketplace. The posts released in June introduced the testing opportunities - https://www.sshopencloud.eu/news/ssh-open-marketplace-meet-our-testers - and announced the alpha release of the SSH Open Marketplace - https://www.sshopencloud.eu/news/ssh-open-marketplace-alpha-launch-0 [10.07.2020]

[15] The summary of this Twitter event can be found here: https://twitter.com/i/events/1272830135524372480 [10.07.2020]

[16] Recording and materials of the workshop are available here: https://www.sshopencloud.eu/agile-development-ssh-open-marketplace-user-workshop [10.07.2020]

○ 3d of July 2020: a consultation webinar was organised with DARIAH communities[17]. This webinar initiates a series that will target the different SSHOC communities.

3. Conclusions and next steps

MS42 allowed the WP7 team to identify a risk in the data ingestion process. Some measures are planned to speed up the data ingestion between July and December 2020. A support documentation for data ingestion will be created (DARIAH/OEAW, SWC, DARIAH/PSNC, CLARIN). Priorities in the identification of sources will be revised and preliminary analysis for each identified source will be conducted (T7.3). Furthermore, based on the consultation processes (consultation platform, events with SSHOC communities), a better equilibrium between scientific fields covered by the content will be achieved.

Curation discussions - including editorial aspects and sustainability options - will be central questions during the next six months, especially curation workflows need to be clearly established (T7.4) and implemented (T7.2).

---

[17] Recording and materials of the workshop are available here: https://www.sshopencloud.eu/ssh-open-marketplace-public-consultation-dariah-community [10.07.2020]