Agile development of the SSH Open Marketplace: User Workshop

Laure Barbot¹, Tracey Biller², Daan Broeder³, Ron Dekker⁴, Matej Durco⁵, Irene Vipavc⁶, Marieke Willems^{*2}

Abstract. For the release of the Alpha version of the SSH Marketplace, we arranged a two-hour slot on 30 June 2020 during the ICTeSSH 2020 Conference. This report provides the context for the development of the Marketplace and presents the feedback received from testers who had been given prior access to the site, and from workshop attendees who were polled via Mentimeter during the event.

1. Introduction and goals

Prior to its Alpha release in June 2020, the SSH Open Marketplace had been in development for over a year. A preference for the agile approach and best practice in user experience design meant we had involved end users as much as possible via workshops, interviews, brainstorming sessions and prioritisation meetings.

With the release of the Alpha version, we wanted yet more user feedback and engagement, so we arranged a two-hour slot during the ICTeSSH 2020 Conference which later became an online event. Our aims were to:

- Raise awareness of how the SSH Open Marketplace can help researchers in their daily activities and supplement the existing EOSC services.
- Gather feedback from selected testers and the SSH research community on the functionality and content of the SSH Open Marketplace.
- Commence recruitment of additional testers to work ongoingly with the development team.

The workshop took place on 30 June 2020, and was attended by over 246 ICTeSSH participants from all over the globe.

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

¹ Digital Research Infrastructure for the Arts and Humanities (DARIAH)

² Trust-IT Services

³Common Language Resources and Technology Infrastructure (CLARIN)

⁴ Consortium of Social Science Data Archives (CESSDA)

⁵ Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH)

⁶ University of Ljubljana, Faculty of Social Sciences

^{*} Corresponding author: m.willems@trust-itservices.com

2. SSH Open Marketplace - development context

Within the larger context of the European Open Science Cloud, SSHOC (Social Sciences and Humanities Open Cloud) is one of five projects dedicated to gathering and pooling domain-specific scientific data and computational tools and making the resources available to researchers.

EOSC is about connecting Research Data with (e-Infrastructure) Tools & Services, following the FAIR principles – not just for research data, but also for software and the way science is carried out. EOSC is about breaking down silos and providing seamless access to research data, tools and facilities.

EC President Ursula von der Leyen said at the 2020 World Economic Forum in Davos, "We are creating a European Open Science Cloud now. It is a trusted space for researchers to store their data and to access data from researchers from all other disciplines. We will create a pool of interlinked information, a 'web of research data."

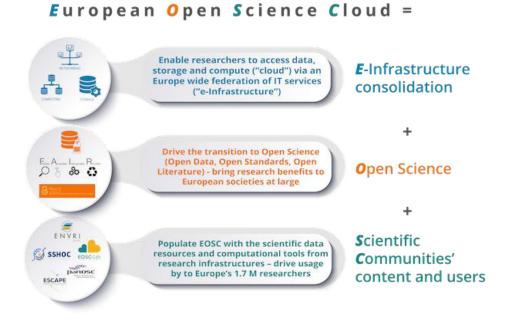


Figure 1: ESFRI thematic cluster view on EOSC. Source: Blomberg, Niklas, & Petzold, Andreas. (2020, January 30). ESFRI thematic cluster view on EOSC. Zenodo. http://doi.org/10.5281/zenodo.3631247

2.1 What is SSHOC?

The overall objective of the Social Sciences and Humanities Open Cloud (SSHOC) project is to realise the social sciences and humanities' part of European Open Science Cloud (EOSC).

The project aims at realising the transition from the current landscape with disciplinary silos and separated e-infrastructure facilities into a cloud-based infrastructure where data are FAIR, and tools and training are available for scholars - especially from those domains in the social science and humanities that have adopted a data-driven scientific approach and that have an interest in the innovation and integration of their methodological frameworks.

Specific objectives are:

- Build the SSH Cloud
- Maximise reuse through Open Science and FAIR principles
- Interconnect existing and new infrastructures
- Governance for SSH-EOSC

All SSH ESFRI Landmarks and Projects, as well as relevant international SSH data infrastructures and the association of European research libraries participate in this project. This will ensure an inclusive approach. Moreover, the consortium has the expertise to cover the whole data cycle: from data creation and curation to optimal re-use of data, and can also address training and advocacy to increase actual re-use of data.

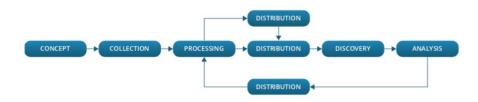


Figure 2: Data cycle. Source: DDI Alliance

2.2 Why give researchers seamless access to high-quality data?

- 1. **Increase efficiency and productivity** through easy access to data and data management tools and services,
- 2. Generate new findings and improved solutions through the opportunity to share data. At Davos-2020, EC President Von der Leyen said there are "hidden treasures and untapped opportunities in the data we generate. ...When every researcher is able to better use their own and others' data, she will come to new insights, new findings and new solutions".

3. **Foster scientific innovation through t**he opportunity to collaborate across disciplines and borders.

2.3 Opening up the SSH domain: connecting researchers with trusted resources through technology

Because of the speed of technological developments in data-acquisition - and research infrastructures over the last decade(s) we have seen an enormous increase in available data and to a lesser extent in data services. This is also true of the SSH domain even if it lags still behind in the uptake of such technologies.

Compared with earlier times when much information was exchanged personally between colleagues, this abundance of resources poses new challenges in how researchers can find suitable resources, both data and services, for their research.

New discovery services that make use of extensive inventories of available resources, created to address such challenges, need to solve curation and trust issues.

- How much confidence can a researcher have in a tool or data-set which is handed down through such a discovery service?
- Does the discovery service eliminate low-quality offerings?
- Is the inventory on which the discovery service is based curated by trustworthy parties?
- How do we measure trustworthiness of such services of often, for the researcher, anonymous origin?

The SSHOC project as a collaborative effort by all the SSH stakeholders is an excellent platform to address these challenges in the following ways:

Increasing the number of certified data and service centers through the CTS certification process. This will make the offerings of such centers more explicit with regards to quality maintenance procedures and governance for instance.

Offering well-curated and tested mappings between resources and services via the SSHOC Switchboard - which is based on technology created by one of the SSHOC partners. The Switchboard will be built into different SSH data catalogues and can recommend lists of suitable services based on a user- selected data resource. Since the current number of available on-line SSH services is still relatively small, the curation effort involved is still limited and can be accomplished by a relatively small number of experts.

Specifically, regarding the SSH Open Marketplace which is a discovery service for all types of research-related resources, development has been informed by previous experience in creating broad service and tool registries. As such, the challenges of ongoing curation and proper governance are duly noted.

2.4 What the SSH Open Marketplace is and what it is not

The SSH Open Marketplace is a discovery portal for tools, services, datasets, research papers, training materials and workflows relevant to Social Sciences & Humanities research communities. Because its creation and development are based on three major concepts - Curation, Contextualisation & Community - the SSH Open Marketplace won't be "just another catalogue" but a service fostering serendipity, supporting SSH communities in the digital aspects of their work, and putting users' needs at the core of its development and operation.

The SSH Open Marketplace automatically ingests and connects metadata collected from various sources, enriching the metadata and linking items through a community-driven curation process.

During an initial onboarding phase, three major Digital Humanities sources were ingested: TAPoR, the Standarization Survival Kit, and The Programming Historian. An extended population phase has begun at the time of writing and will enable better coverage of the diversity of Social Sciences and Humanities disciplines.

In order to ensure the sustainability of the SSH Open Marketplace after the end of the SSHOC project, a shared governance model relying on existing ERICs is under investigation. Furthermore, as the continuous curation of the portal is one of the challenges for sustainability, the SSH Open Marketplace plans to build on existing networks and to implement community features to foster interactions for the benefit of content quality.



Figure 3: Marketplace Mock-Up

3. End-user views of the SSH Open Marketplace

3.1 Feedback from testers

A week prior to the Alpha release, the Marketplace was made accessible to four European testers with different backgrounds and expertise. The testers had been selected via an open call published on the SSHOC website and in social media. They were introduced to the community in a news item published before the event.



Figure 4: SSHOC Announcement selection of SSH Open Marketplace Testers

Agnieszka Szulińska graduated from <u>Cardinal Stefan Wyszyński University</u> in Warsaw with an MA degree in Polish Philology (specialisation in editing). She is currently a PhD candidate at <u>the Institute of Polish Language Studies of the Polish Academy of Sciences</u> and a member of <u>The Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences</u>. Her research areas include: digital humanities, language of video games, sociolinguistics. <u>ORCID-ID</u>

Vanessa Hannesschläger completed her studies in German Literature at the University of Vienna with a doctoral thesis on the works of Peter Handke. Vanessa is a researcher at the Austrian Centre for Digital Humanities and Cultural Heritage. She is involved in CLARIN and DARIAH and co-chairs the DARIAH WG "Ethics and Legality in Digital Arts and Humanities" (ELDAH). Vanessa is actively pursuing Open Science. She was a Wikimedia Germany Open Science Fellow in 2017/18 and won the Open Minds Award in 2017. She is a member of the Open Science Network Austria (OANA) and teaches Open Science at the University of Vienna.

Maurizio Toscano is a researcher at the University of Granada (Spain). His research interests include Digital Public History, Data modelling and GIS. His current doctoral research focuses on Web Information Systems for research in History, History of Arts and Archaeology. Maurizio received his Bachelor's degree (Hons) in Cultural Heritage and

Master's degree (Hons) in Computer Application in Archaeology in Italy, before moving to Ireland where he worked 10 years in the private sector. Since 2014, Maurizio has participated in several FP7 and H2020 projects (MEMOLA, REACH, DESIR) as partner coordinator and work package leader. Since 2019, he is a member of the Board of Directors of Asociación de Humanidades Digitales Hispánicas.

Dr. Kasia Karpinska is scientific manager at ODISSEI (Open Data Infrastructure for Social Sciences and Economic Innovation). In this capacity she is interested in innovative tools available for researchers. She obtained her PhD from Utrecht School of Economics in 2013. She previously worked as a researcher at Erasmus School of Social and Behavioral Sciences (EUR) and Sociaal en Cultureel Planbureau (SCP). During her academic career, Kasia gained ample experience in data collection and data management.

During the workshop, the testers were called on individually to share their findings and comments with the other attendees.

3.1.1 UX Design, Navigation and Search experience

Agnieszka Szulińska

The alpha version of Marketplace.sshopencloud.eu works well. There are no major problems with navigation and content display. At this stage you can see small differences in the display of content between Chrome and Safari (in favor of the former). The mobile version is not yet available. However, it is quite understandable with the alpha version. A great kudos for adding the contrast option.

Breadcrumbs functionality as a navigation is good practice in web design and seems to be very significant in this type of 'project-catalogue' website. The site works fast, I do not wait long for the search results to load.

As a literary researcher and linguist, I have noticed the names of keywords. It seems that some of them are too general (for example *Images* or *Textual analysis*), while others are too detailed (*Tokenization tool for Arabic and Chinese* - also called topic, not as keyword). Also, for better readability, the number of keywords for one entity should be reduced to a maximum of 4-5.

As a creator of digital scientific editions, I searched the site for the TEI standard in that topic. And although I received some results, their connection between them seemed to have as a common point only TEI in the name. Probably the contextualisation and entity linking function will work better when more content appears on the site.

As for the layout, I suggest a more diverse pair of fonts than Ubuntu and Open Sans, due to their similarity. It would be desirable to maintain graphic consistency, i.e. if graphics with certain colours, human figures appear on the homepage - and only on it, then you have an impression of emptiness on subsequent subpages. Also, the icons of modules at the moment are over-detailed, and they aren't really distinctive one from another.

In general, my impression stays positive, because everything could be improved at this stage and there are no critical problems on the structure level.

Vanessa Hannesschläger

The website SSHOC marketplace is generally well structured, the main elements visible at first glance and navigation is intuitive. The paratexts are easily understandable and not too long (which is a benefit on the start page, in the about pages I would suggest more extensive descriptions). How to use the search and filter functions is self-explanatory, though I experienced a certain uncertainty about what to type in the main search field.

Even though I appreciate the filter functionality, I believe that there are too many keywords available. This might be improved if a differentiation between frequently and less frequently used filters was implemented (e.g. with larger and smaller font sizes like in traditional word clouds). I also felt that the "last added" section might become useful for future frequent users of the platform, but currently it serves no real purpose.

Maurizio Toscano

The unified search box is certainly a valued design choice, in line with the clean and uncluttered UX. At the same time, it can be intimidating for the user that goes to the Open Marketplace without a clear idea of what is looking for. Introducing Autocomplete suggestions, based on terms relevance and popularity, can help users put together more meaningful queries. If no results are shown for a certain query, a message suggesting to browse content as an alternative to free search can be useful.

Accessibility features, visible on mock-ups but not implemented yet, are a great addition for all users, as the color scheme used is very light in contrast and fonts relatively small.

A fixed menu bar during scrolling, together with a "back to top" link, can make long pages navigation easier.

At launch, replace the homepage "Last added" with a "Recommended" or "Editors' picks" section.

Useful additions for the search results page can be the possibility to pick and compare specific results, using metadata from the details sidebar, and the option to switch search results between list and grid (3x3) views.

In terms of filters, I suggest to update live the counts between brackets, accordingly to selected options.

In general, I would like to see personalised features for registered users, like the possibility to bookmark favourite resources, save searches, and retrieve reported issues.

Kasia Karpinska

In general, the Open Marketplace website is very user friendly. The main page does not include too many functionalities and is therefore easy to navigate. The main categories are informative and distinct, which allows the users to orientate in the content offered by the Open Marketplace.

The input for the website is still rather limited (compared to the envisioned final product). It raises the concerns that the website will be less comprehensible when more

content will be ingested. This point applies mainly to keywords. They are already royally applied, with more content to be categorised, their number would increase. Too many keywords might obscure the content of the Open Marketplace instead of making it easily searchable.

Small changes in the layout of the website, e.g. larger font size and smaller size of the pictures that are being displayed would further improve user experience.

3.1.2 Content & curation of content

Agnieszka Szulińska

There was not a lot of content added in the alpha version. Especially *Workflows* and *Datasets* modules had only one entity each. Basically, four modules were separated: *Tools, Training Materials, Workflows and Datasets* is correct, although the borderline for instance between *Workflows* and *Training Materials* is quite fuzzy and in the future one case may fit both modules.

Descriptions of individual entities could be more detailed and added by creators and contributors of SSH Open Marketplace. A great quality leap forward for the whole site would be to add examples of use for Tool or Datasets modules or feedback on Training Materials and Workflows used by real users. Therefore, when describing Oxygen, there could be a list of digital projects that used this tool, e.g. digital editions, dictionaries, language corpora.

It would also benefit greatly the whole Open Marketplace, if descriptions would be available in at least two languages (English and mother tongue of the entity). It could boost multilingualism movement; however, it might require rebuilding the whole browser on the website.

It is a positive aspect that each entity has "contributors" instead of "authors" which sounds more suitable when it comes to such cases. Add functionalities which enable the community to curate SSH Open Marketplace will be helpful.

Vanessa Hannesschläger

To have an unclouded experience of the marketplace, I deliberately did not take a look at the user stories developed by SSHOC so that I would be a prototypical "uninformed user" who does not know the intentions and expectations of the marketplace makers. As such, I felt that the description page for each content item, which provides a very brief description, a picture and a prominent link to the actual content, is well-arranged and allows for intuitive use. The link to the actual content is clearly visible and easily accessible; I find it a very good idea to avoid lengthy descriptions and instead offer immediate direct access to the content at hand.

One major issue is the question of multilinguality. Currently, there are content items in other languages available in the marketplace, which is somewhat problematic. For the example at hand, the "Deutsches Textarchiv" (DTA), it does make sense to provide information about a German language resource in German, however a user who does not know German and comes across this resource might be irritated as this user cannot identify the reason for the German language description.

Maurizio Toscano

The content display page is well structured and balanced in terms of metadata provided, links and images. The details sidebar is key to offer useful information to the users and help them decide if the content matches what they are looking for. Additional data can be: license type (FOSS, Free, Freemium, Subscription, One-time License); difficulty level (for training material and tools); time required (for training); latest release (for tools); provides certification (for training).

On the Related content section, I would like to see only content that is strongly, and not only somehow, related to the current resource, for example: training material specific for the tool on display; tools that are alternatives to the one on display; examples of datasets that are suited to be manipulated with the tool on display.

About content curation, it would be important to avoid highlighting outdated content, i.e. abandoned or experimental software or obsolete tutorials. A possible solution is to hide them by default, with a filter to turn them on in search results (for anonymous users) or even a fixed option on the user profile (for registered users).

Personalisation can be the key to build "not just another catalogue". In order to do that, OMP could work as a recommender system, with proactive functionalities. On account creation (optional), users are presented with options to describe what kind of DH researchers they are (new, experienced, advanced), their background and research topics, and select some tool they already know, in order to provide personalised content suggestions, for example alternative software or scenarios related to their interests (i.e. a Netflix for Digital Humanities).

Kasia Karpinska

As a social scientist, my main interest in the OpenMarketplace is the *dataset* functionality. Currently, only one dataset is included in the tool and it is therefore hard to comment on how this functionality operates and what could be improved. The functionality offers a link to the data provider and the metadata offered by the provider. There is no information on the variables that are included in the dataset.

The functionality as it is presented now resembles other tools, such as the Google Data Search or CESSDA Data Catalogue. Is it not yet clear what is a unique element of the OpenMarketplace that distinguishes it from the other tools.

The developers intend to ingest the datasets from the CESSDA Data Catalogue. This raises the question of how the functionality will work with ingestion of a large database, both in terms of technical capability but also in terms of search functionality. Possibly new selection tools would be needed to easily navigate through a large number of datasets.

Tools and training material functionalities offer a good overview, especially for new users but again, this content could also be found via other search engines.

It is not clear how the contextual links between different elements are established.

3.1.3 Curation, trust & governance

Agnieszka Szulińska

The whole website has a high level of credibility.

I am delighted with scrutiny in informing which data about users are collected (there is even the name of the tool used to analyse those data). I could also find information about the website's code copyright and direct link to GitLab to check the code by myself.

The idea of curation, community and contextualisation as three pillars of the SSH Open Marketplace is germane and solid in Digital Humanities. Although community features will (probably) be added in the beta version, the fact that testers were invited to such a project is promising in that aspect.

The most astonishing thing is that after clicking contact there is a blank page. It is not possible to find names of the authors/contributors involved in building SSH Open Marketplace (this might be added when you read that text). On the other hand, we have info about the media owner, publisher, responsible for content and editorial office, service provider etc.

In general, this project also gains my trust because of cooperation to some extent with DARIAH infrastructure.

Vanessa Hannesschläger

Information on the funders, the project context, as well as on the implementing institutions are clearly stated. It would however be desirable to indicate the names of the persons involved and make more prominent links to EOSC and the SSHOC project (e.g. by including the respective logos, which might also be done for the institutions involved).

In terms of curation, I would find "last updated" information useful. Even though the marketplace is intended to stay active for the foreseeable future, if activity does decrease at some point, it will be helpful for users to know for how long content and paratexts have not changed. Also, it will make it easier to reference the website. As for user contributions to curation, a "contribute" button and a comment section for each content can improve active user involvement.

Maurizio Toscano

The Open Marketplace looks like a very well oriented initiative, aiming at providing something new without reinventing what is already available. Anyway, this approach can prove to be challenging In the long term, as it is based on the existence and maintenance of external content providers. Some of them seem to be more consolidated initiatives than others. Somehow, without changing its approach, the OPM and its promoters should consider playing a role in the long-term sustainability and governance of these main content providers.

In terms of players involved in the OPM governance, it can be useful to include an infographic to visualise their role and interconnections. To increase trust, a non-prominent disclaimer can be added when suggesting commercial software, declaring not being affiliated or receiving revenues from links.

Kasia Karpinska

It is remarkable and laudable that maintaining the platform is seen as a community effort. Transparency about who is moderating the content of the Marketplace and along which criteria, is crucial to ensure the trust of the users in the tool.

A selection of defined Key Performance Indicators (KPI) would help to monitor the usage of the tool. It would also allow monitoring whether different sub-disciplines of SSH are equally well represented among the contributors and if the content relevant for each sub-discipline is equally well kept up to date.

Also, more transparency about governance structure is advisable (the information presented now is rather limited).

3.2 Feedback from Workshop Participants

That the event was moved online gave us the perfect platform for engaging an even larger audience to evaluate their needs and recommendations.

We used the <u>Mentimeter</u> tool twice during the workshop, the first time after the main presentations and the second time one before the end.

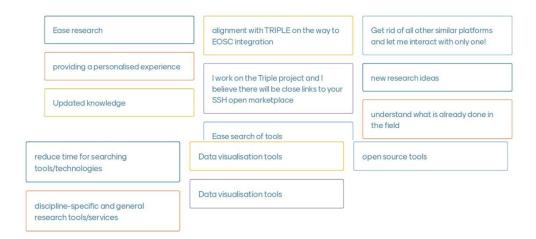
Attendees were asked about their country of residence where most reported being from Europe, with United States, South Africa, India, and Russia also being mentioned. As expected, most attendees were either researchers (15%), came from universities or academic institutions (31%), research libraries and archives (20%), or research infrastructures (19%). Representatives from the private sector, citizen scientists, funding, and policymaking organizations were present as well.

Several research domains or areas of data management expertise were mentioned. This raised our expectations that Marketplace will indeed be used by a diverse community.



We also asked about ways in which the SSH Open Marketplace could support their work and received several proposals.





Services and tools that attendees would like to see in the SSH Marketplace are a combination of what they need to look for in several other platforms and databases or datasets that would ease their research.



We were also interested in ideas on how to motivate users to contribute to the curation of the marketplace resources. Mentioned ideas went in two directions. The first one was citability and being acknowledged, and badges were mentioned several times. The second was about community building, where cooperation and curation are crucial.

SSH Open Marketplace is being built for the community and we want the community to trust resources in it. Attendees proposed transparency and open and clear information on governance in this respect.

4. Conclusions and next steps

The workshop setup with four dedicated testers having sufficient time to get acquainted with the tool proved extraordinarily informative and effective. The testers came up with many useful observations and suggestions, which will decisively inform further development of the SSH Open Marketplace.

Even where they tackled issues already known to the development team, it is helpful to get an additional external view. In general, it is a challenge to find the right balance between how far to go with development and how soon to show users (and which) even preliminary products with limited functionality. In this specific case we believe we found the sweet spot.

To highlight a few major aspects of the discussion: Some of the suggestions for UX design and navigation were already planned before, or will be included, e.g. autocomplete for search, **personalisation**. Complex search (with search operators) is technically feasible and will be considered for implementation at a later stage, though it is being very seldom used (based on experience from similar projects).

Regarding content & curation, the issue of representing **datasets** in the marketplace has been discussed often. There is no plan to duplicate all the records from big data catalogues (like CESSDA Data Catalogue or CLARIN VLO), but the catalogues themselves will be represented as items in the Marketplace (as Discovery Services) and manually selected datasets interlinked as samples for tools or used in training materials.

The similarity/overlap of **workflows** with training materials has been rightly pointed out by the testers. This is acknowledged by the team. The rationale for introducing workflows - aka recipes, or scenarios meant to describe sequences of steps to accomplish a task - is to emphasize the "how" question, which the SSH Open Marketplace is meant to offer answers to mostly.

Important to highlight the relation between data quality and user experience. This is especially manifest with faceted browsing, considered the most intuitive way to explore the dataset, that can however can be severely hampered by (bad) data quality, i.e. due to proliferation of facets or bad facet coverage. Therefore, curation and editorial work are critical efforts to improve navigation and search experience. As one measure/strategy to tackle data quality community curation is considered. However, experiences from previous endeavours show that it is difficult to motivate people to contribute without direct reward. Therefore, the team was also very glad to see a few hands raised in the audience indicating their willingness to contribute to the curation efforts. Still the question, how to incentivize contributions will remain a challenging one.

There is also a need to review the information **about** the marketplace, e.g. who are the contributors, how is the curation process, the legal aspects, and this will be done for the public and beta release planned for December 2020.

All presentations can be found on ZENODO: DOI: 10.5281/zenodo.3928148

Recordings can be found here: https://youtu.be/AZTrzHtil3Q

5. References

Gotz, Andy, Petzold, Andreas, Asmi, Ari, Blomberg, Niklas, Lamanna, Giovanni, & Dekker, Ron. (2020, February 19). ESFRI cluster projects - Position papers on expectations and planned contributions to the EOSC. Zenodo. http://doi.org/10.5281/zenodo.3675081

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.

Blomberg, Niklas, & Petzold, Andreas. (2020, January 30). ESFRI thematic cluster view on EOSC. Zenodo. http://doi.org/10.5281/zenodo.3631247

Keynote by President von der Leyen at World Economic Forum 2020. https://ec.europa.eu/commission/presscorner/detail/en/SPEECH 20 102

Ron Dekker, Laure Barbot, Daan Broeder, Matej Ďurčo, & Marieke Willems. (2020, June). Agile development of the SSH Open Marketplace: User Workshop. Zenodo. http://doi.org/10.5281/zenodo.3928148

Funding

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