



CO-OPERAS - SSHOC Workshop: FAIR research data in the Social Sciences and Humanities

January 30th 2020, Göttingen

Stefan Buddenbohm, Raisa Barthauer, Joseph Dung



SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement No 823782.
OPERAS-P has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871069

On January 30th 2020 CO-OPERAS and SSHOC conducted a joint workshop to discuss various topics under the headline "FAIR research data in the Social Sciences and Humanities. This document outlines the main findings from the SSHOC related part of the workshop.

Setting and audience

The workshop was split in two parts: in the morning presentations and a group discussion regarding FAIR principles in the SSH had been conducted. The afternoon session was devoted to discussing the SSH Open Marketplace to involve the user community in the development of the platform (participatory design, user involvement). The marketplace session was realised as group work along pre-selected topics. The organisers tried to apply activating methods and shifting of setting: collecting feedback in groups on posters, a moderated group discussion led along selected topics, talks with subsequent discussion.



Filling posters with ideas, feedback, critic along selected or self-chosen marketplace topics

The uptake of the workshop had been unexpectedly good with 28 participants regarding the limited forerun of only two months. Initially intended as local event a large share of the participants came from abroad. The participants' backgrounds had been quite diverse with emphasis on researchers (mainly from the Göttingen research campus), information specialists (libraries) and research organisations (such as Max Weber Stiftung, Leibniz Gemeinschaft).



Some of the participants of the CO-OPERAS - SSHOC workshop, January 30th 2020

The overall comfortable and informal discussion atmosphere was well received with the participants and lead to a nearly complete consent to get prospectively involved in the development of the SSH Open Marketplace.



There was consent among the audience that the time frame of only 6 hours is too short for the discussed topics. On the other hand the limited length of the event spurred participation better - theoretically - compared to a two-day event. The German language approach of the workshop was commended by the audience because it considerably lowers the threshold to give feedback and formulate own ideas.

The event showed that there is a considerable demand within the SSH research and information infrastructure community for exchange and guidance regarding FAIR and also that there is some enthusiasm to take part in the development of information infrastructures. Although it seems difficult to interest a large audience ONLY for a participatory design workshop on the SSH Open Marketplace a joint effort or docking to larger events seems a very promising approach. On the other hand it seems reasonable to approach the growing pool of interests in the SSH Open Marketplace development for further involvement like focus group meetings or surveys.

Feedback on the SSH Open Marketplace concept and the MVP

The participants had been presented the concept of the SSH Open Marketplace (its aims, target audience, intended benefits for research) and the MVP of the marketplace with the functions, user interface and contents. The results of the discussion are grouped below along selected topics.

The most popular topics had been:

1. Content of the SSH Open Marketplace: formats, types, interoperability, actuality, particularly the processing of research data was highlighted, e.g. what kind of methods and tools, relation of "our" content to third parties
2. User involvement: means of getting involved, enabling exchange with the other users BEYOND simple content update forms
3. Context of the presented items, especially how they have been rated, used and applied in specific research undertakings



Audience following the presentation on FAIR principles in the SSH

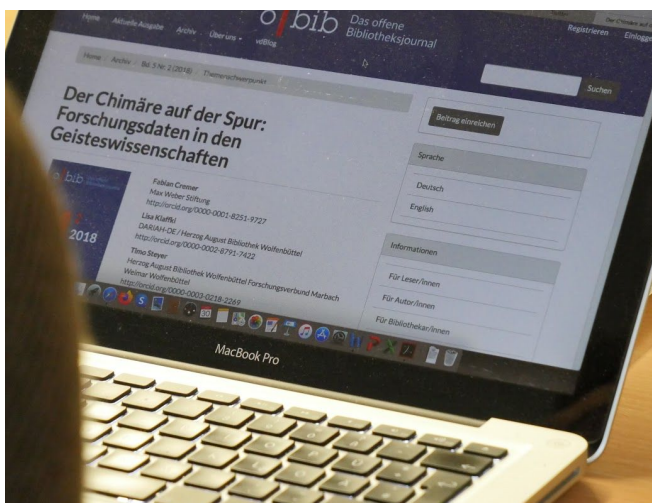
In the following the feedback is clustered along various paragraphs without a certain hierarchy in the list.

Content of the marketplace

- A recurring topic was the fluidity and lack of consistent, standardised formats for research data, particularly in the humanities. Participants expected this to be a considerable challenge for the SSH Open Marketplace and expressed scepticism towards the broad range of contents and how to display it in the search results in a convenient way.
- This challenge could be addressed through the concept of "solutions" or "scenarios" which had been well received among the audience. By linking stand-alone items in a

logical sense to one another and providing context the marketplace can be really useful for researchers. If it just stays a registry of resources, this could also be useful but an added value could really be reached through the contextualisation of its items.

- Several participants requested support for research data related activities like converting data sets.
- Actuality seemed to be (after the contextualisation) the second most important challenge for the marketplace. In terms of the user interface this could initially be covered with a “notify function” to highlight broken links or outdated information. Beside, this could mean that user-centered functions and the actuality of the content seems to be worthy of discussion itself.



Landing page/ functions/ user interface

- The frontpage of the MVP appeared too empty for some, as it did not provide enough guidance and the necessary context of what to do and how. While the single search box on the front page works for Google-styled searches, the participants highlighted the need to have displayed what other researchers are searching for. Others wanted to show editorial or

automated content presentation like, the most popular tools and datasets.

- All categories: The tools section needed a more granular based structure with as many sublevels as possible.
- Utility metrics: e.g. older versions
- Personalisation: “My tools and search settings”
- Comment or update “button”
- Ratings of content
- Special icons for the categories, e.g. tools, data, tutorials

User guidance, tutorials and user involvement

- It became clear that the users as a whole want to be involved in the platform. They want to benefit from other user experiences and do not want to have static, predefined result displays for the use of the site, which they cannot comprehend. Instead, they want to be able to flexibly adapt the search and the site to their needs. An interaction with each other and with other projects is considered important.
- The need for guided tutorials in actually using and navigating the site. Some recommended video demonstrations on Youtube on how to make good use of the features on the site.
- Some demanded more information at the About Page on: Who moderates? How? Data ingests from where? Explain the search (and ranking) algorithms?

- Essentially, a lot more transparency about the whole process needs to be clearly explained to enable better participation (already at the MVP instance).
- An idea was also to integrate a community-related function, like a forum or platform for discussions into the site, where users can ask their questions directly to the community to solve problems together (StackOverflow).
- Tutorials desired for the: A. Marketplace, B. Content Category, C. the platform itself needs a guided tutorial.



- User engagement was grouped in the following categories:
 - Gamification: Enhancing Collective Action by rewarding collaborative research activities and tool usage: A researcher familiar with Gamification and who also teaches the same subject at an institution explained how it was quite difficult for her to get her students initially motivated to use certain tools for an assignment. It was only when she let them start from 0 with an opportunity to earn more points to pass her course did they suddenly see the need to use those tools. Her point was that their motivation to use the tools was not really based on any inner desire or motivation to learn about the tools but merely a desire not to fail a course. Researchers, she argued may have a slightly different type of motivation and may not react favourably to having their performance or even their tools "ranked" in this way based on their participation in the MP. She forwarded the following question: What if the researcher has sensitive data and cannot share or contribute to a research activity? Will that researcher be penalized for being limited in that regard?
 - Personalization: The possibility of a researcher to like or follow a favorite collection of tools and get further recommendations from those. The workshop feedback on Personalization (especially if it is a factor at all in the tool/data ranking algorithm) was an emphatic example of how a Lecturer/Professor can make his

students like and follow his favorite tools (especially if those affect the visibility of those same tools in the MP).

- Visualization: New ways to Visualize the tools and data results based on the term or keyword searched to enable further filtering. The possibility of researchers visualizing their connections with other researchers based on their tool and data usage. The feedback on Visualization was the need to show a video demonstration of such features or tools in action.
- Social Network Analysis: A statistical analysis of the types of graphs a research activity generates based on what keywords are used in connection to tools and datasets clicked. How are researchers similar or dis-similar through this process based on this analysis? The feedback on this type of engagement method retained the focus on having short video interviews with researchers (example: SAGE Research Methods). Since the MP does not host content, perhaps the possibility of enabling researchers with profiles link to such interviews from Youtube possibly?
- Tasks: A researcher can pose questions to the general research community based on a dataset and/or tool and challenge the community to come up with solutions.

Result presentation

- For the display of individual results connecting with respective tools (contextualisation), was in particular demand, e.g. how can a tool be used with certain data sets. Users want to be made clear how to use a tool. Suggestions included linking tools to results, experiences and/or publications made with them (what is the tool good for?) and how-to videos in which standard cases of data entry illustrate the application.
- In this context, projects or researchers could be given the opportunity to add to their websites that tools/data sets are used (direct link).
- Provenance: the age and update of the tools must be displayed, users want to check if the tool is outdated. This was a recurring issue as the participants were aware that the marketplace as such is mainly collecting from third party sources.
- With information about the licensing the results would show users whether and how they can use data and whether they can further develop tools (also useful as filter/search category).
- In a user-specific area (Login), users could save search settings or generate their own tool list for their projects.
- A major point for the participants regarding the display of results was the lack of clarity due to the amount of offered tools ("I find too much a.k.a. nothing") and the demand for grouping similar items together (contextualisation). This seems to be connected to the draft character of the MVP.
- The need for different filter options was mentioned particularly often, such as Filter by specific disciplines; Filter by Open Source tools; Filter by tool categories; Filter by types of data; A better overview was required. This includes: adding more tags and categories to the tools; the possibility for users to rate content or tools and write reviews or comments (user enriched content) was frequently requested; more links between similar or related tools; more search categories.

- It is important for users to be able to understand the ranking of the results in the display. A fixed black box-liked algorithm is explicitly rejected. In connection with this, a hierarchical output of search results was proposed. Graphical preview of the results, especially for object based data-sets.
- A possibility of interconnection or at least a link to other projects was proposed.
- Filtering: more search categories; filter for subject-specific discipline; filter for open source tools; hierarchical output of successful results; more tool units: filter, tags, categories, filtering by tags.
- Color code search results, e.g. tools that may be outdated already.
- Workshop participants responded favorably to suggestions for User Tagging of tools, solutions and datasets. However, when an Algorithmic versus a Human Mediated ranking of tools was suggested--the issue of transparency in this area was raised. There was a demand to explain the basis of the algorithms.

Best practices or examples to look at

- <https://dans.knaw.nl/en/about/services/DataverseNL>
- <https://stackoverflow.com/>
- <https://re3data.org> (example for the display of results)
- <http://b2find.eudat.eu/>
- <https://agate.academy/de.html>
- <https://www.gesis.org/angebot/recherchieren>
- <https://colreg.de.dariah.eu/colreg-ui/>
- <http://www.meta-share.org/>
- <https://linguistic-lod.org/>
- attention to how 'bbaw.de' organise their diverse data sets and formats: XML, tables, etc., into different clusters, <http://www.bbaw.de/telota/projekte/digitale-editionen>
- <https://www.semanticscholar.org/paper/Understanding-Data-Retrieval-Practices%3A-A-Social-Gregory-Cousijn/e4148bdc374788a9cc0f5dc2068a3f8130ec71f1>
- work with personas, e.g. HRK 2015 paper of FDM an Hochschulen
- link up with multiplicator institutions and platforms, e.g. RDM at universities
- <http://tadirah.dariah.eu/vocab/index.php>
- SAGE Research Methods Platform
- <https://www.ianus-fdz.de/> - IT definitions groupings, layout
- <https://operas.hypotheses.org/projects/triple>
- <https://archwort.dainst.org/de/vocab/>



Worst practices

- No step by step guided usage.
- Relevance ranking: What becomes "relevant" for who please? "Relevance" is determined for whom please?
- Outdated Links
- Risk for losing focus: too many goals and target groups at once: Forum (Stack Overflow) + Metasearch + portal here
- Cutting/Stepping into international repos available?
- Bad display: no good/correct way to display the current MP on the smartphone
- Empty marketplace: Display some interesting contents before searching.
- User Uploads: ability to upload content possible?
- Do not ignore Librarian's expertise on metadata harvesting.

As a researcher I want (merely a wish list)

- Content: Tools, Primary data, research data, Sources of Information, Support
- Support/ Communication: Information on State of the Art, Data Librarian, Data Steward, Tutorials/Workshops, Lab
- Unsorted: Trust, Visit the best technology platform, Forum for all dummy questions, Contempt of all digitized literary works

Follow up activities

Several follow up activities to the SSH Open Marketplace session are planned or in discussion within work package 7 of the SSHOC project:

1. All participants of the workshop received a summary (also documented on the corresponding website¹ and additionally a questionnaire detailing their feedback and expectations towards the marketplace. This questionnaire is part of an overall effort to build up a pool of researchers and information specialists interested in the development of the marketplace.
2. Additional workshops facilitating user involvement are either planned or in discussion (e.g. on occasion of the DARIAH Annual Event 2020 in Zagreb).
3. Small focus group meetings bringing together not more than 5-7 persons looking specifically at the user interface, the contents and functionalities of the marketplace. This can be done quite easily with the already available MVP². Here the focus could lay more on the user experience related aspects.
4. Some participants of the workshop invited SSHOC to present the SSH Open Marketplace in their institutions and/or existing formats³.
5. Beyond this partner-specific follow up activities are worth considering. For Göttingen this can be a linking up with activities of the eResearch Alliance⁴.

¹ <https://fair-rdm.sciencesconf.org/>

² <https://sshoc-marketplace.acdh-dev.oeaw.ac.at/>

³ E.g. the FDM Jour Fixe NRW or at the IEG Mainz.

⁴ <https://www.eresearch.uni-goettingen.de/>