



# Contextualizing Research Tools & Services Through Workflows in the SSH Open Marketplace

RESEARCH PAPER

]u[ubiquity press

LAURE BARBOT 

MAJA DOLINAR 

EDWARD J. GRAY 

CRISTINA GRISOT 

KLAUS ILLMAYER 

MICHAEL KURZMEIER 

BARBARA MCGILLIVRAY 

\*Author affiliations can be found in the back matter of this article

## ABSTRACT

This paper provides an overview of workflows in the Social Sciences and Humanities Open Marketplace discovery portal. Workflows are defined as step-by-step research scenarios pointing to resources that can be used at each stage for someone to reproduce these scenarios. Born as TEI objects in the Standardization Survival Kit, a mapping exercise was needed to transfer them into the SSH Open Marketplace and ensure their legacy. Dedicated forms are now available to create and edit workflows which provide an easy to use interface for researchers to be autonomous in the creation and design of their workflows. This paper reviews the history, the technical, and community dimensions needed to describe research workflows in the Social Sciences and Humanities. It also describes the challenges faced and the new developments envisioned so that these workflows can better reflect and serve SSH researchers.

## CORRESPONDING AUTHOR:

**Laure Barbot**

Digital Research Infrastructure  
for the Arts and Humanities  
(DARIAH-EU), Paris, France

[laure.barbot@dariah.eu](mailto:laure.barbot@dariah.eu)

## KEYWORDS:

research workflows; open research; open methods; SSH Open Marketplace; open science

## TO CITE THIS ARTICLE:

Barbot, L., Dolinar, M., Gray, E. J., Grisot, C., Illmayer, K., Kurzmeier, M., & McGillivray, B. (2024). Contextualizing Research Tools & Services through Workflows in the SSH Open Marketplace. *Journal of Open Humanities Data*, 10: 22, pp. 1–12. DOI: <https://doi.org/10.5334/johd.192>

## (1) INTRODUCTION

The Social Sciences and Humanities Open Marketplace,<sup>1</sup> short SSHOMP, a major output of the EU Horizon 2020 funded project “Social Sciences and Humanities Open Cloud” (SSHOC),<sup>2</sup> is a discovery platform for new and contextualised resources from the Social Sciences and Humanities (SSH). The idea of SSHOMP is to support researchers in finding and comparing digital tools and methods for their research and increase the findability of datasets, tools, services and training materials. For this, metadata on different types of resources is collected either manually by contributors to the SSHOMP or by ingesting them from a data source (cf. section 3). After an initial development phase funded by the above-mentioned project, three European Research Infrastructures Consortia (ERIC) – CESSDA,<sup>3</sup> CLARIN<sup>4</sup> and DARIAH<sup>5</sup> – decided to financially support the SSHOMP discovery service, under the newly created SSH Open Cluster (i.e. one of the domain specific branches of the European Open Science Cloud (EOSC)).<sup>6</sup> This institutional and infrastructural support is essential to ensure both the uptake and the sustainability of the service. This paper does not enter into details on these aspects, but rather explains what it was possible to develop and maintain in such a context.

Workflows are a specific content type in the SSHOMP and are defined as “Sequences of steps that one can perform on research data during their lifecycle. Workflows can be achieved by using diverse tools, resources and methods, and the useful resources are connected to each step.” (*Social sciences and humanities open marketplace 2023*). Workflows explain a series of actions to be performed with the help of tools, services, datasets. Workflows are an innovative type of resource proposed on the SSHOMP compared to other similar discovery platforms or catalogues. Within the SSHOMP, workflows are built on use cases that bring into light recommended tools, formats and methods, and are reusable for other researchers. Workflows represent thus “the living memory of what should be the best research practices in a given community” (*Social sciences and humanities open marketplace 2023*). As such, workflows are a good way to point to tools and services that researchers are using. This will help others in understanding which digital methods may serve their needs the best. The focus of this paper is these workflows in the SSHOMP: we will provide the background on their creation, explain how they are implemented and discuss some of the challenges met by the SSHOMP team. Section 2 discusses the genealogy of the concept of workflow and how workflows have been implemented, as well as challenges encountered along the way, taking the example of the Standardization Survival Kit (SSK) service, which was ingested into the SSHOMP. Finally, we discuss the multilinguality of the SSK and the challenges when mapping this to the SSHOMP. In Section 3, we define SSHOMP workflows, how the ingest pipeline and the data model support them and we highlight the inherent community-driven dimension of the workflow collection across humanities. In section 4 we outline the future work related to non-linear workflows, multilinguality and the reusability of workflows.

## (2) WORKFLOWS GENEALOGY

### (2.1) WORKFLOW CONCEPT

Describing research workflows can be seen as a way to put the focus on the research processes rather than on the research products, contributing to opening up methods in the Social Sciences and Humanities. Tasovac et al. (2023:3) stress that “DARIAH-EU is urging the integration of process-oriented systems of evaluating science into the currently dominant product-oriented models”. Open methods is an important dimensions of Open Science, often relegated to the

---

1 <https://marketplace.sshopencloud.eu/> (last accessed date: 12/12/2023).

2 The “Social Sciences and Humanities Open Cloud” SSHOC has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement #823782, see European Commission: CORDIS, <https://cordis.europa.eu/project/id/823782> (last accessed date: 12/12/2023).

3 Consortium of European Social Science Data Archives, <https://www.cessda.eu/> (last accessed date: 12/12/2023).

4 Common Language Resources and Technology Infrastructure, <https://www.clarin.eu/> (last accessed date: 12/12/2023).

5 Digital Research Infrastructure for the Arts and Humanities, <https://www.dariah.eu/> (last accessed date: 12/12/2023).

6 See <https://science-clusters.eu/> (last accessed date: 12/12/2023).

sidelines because they are more difficult to grasp and evaluate, but as highlighted by Leonelli (2023) “Open Methods is not a matter of recording and sharing every detail of a research procedure, but rather a reflection on which research components and techniques are most salient to the outcomes, and should thus be accessible and reproducible”. Research workflows of the SSHOMP fall under this approach. They were developed within the SSH domains, while we see similar endeavours in other fields, for example the Elixir FAIR Cookbook (Rocca Serra et al. 2022) that shares “recipes on the FAIR components” for Life Sciences. The concept of workflows was originally connected to the IT domain and the possibility of automation of repeatable patterns, but taken generally as a sequence of activities it can be applied to any domain, especially when discussing the digital aspects of the research methods.<sup>7</sup> Indeed, beyond the classic epistemological distinctions between hard sciences and SSH that often oppose hermeneutics to experimental methods, or qualitative to quantitative data analysis, we see common needs and challenges when it comes to opening the black boxes of methodologies across domains. Once we acknowledge Leonelli’s statement cited above, it is clear that the goal is not to record every detail of the decision process of a researcher, which in some cases would be impossible, but rather to critically document the perimeter chosen for the data collection and the tools and standards used to process them. In that sense, workflows, as we define them here, are more “narrative” workflows giving space to reflect on the choices made and to generalise them already having a peer in mind, rather than bundles plugging one tool after the other to fully automate a process. This critical approach could be seen as one of the “specificities” of the SSH workflows, but it is not the only one and section 4 of this article outlines non-linearity, multilinguality and the meaning of reproducibility in the SSH domains as some of the elements to also take into account when attempting to “capture” SSH research methods.

Workflows are a special kind of item in the SSHOMP. They explain an action to be performed with the help of tools, services, datasets, etc. Such an action can be abstract, like a generic method to be performed on data, or very practical, like a detailed explanation of how to process data so as to produce a specific output. One example is a workflow that describes how to arrange the digitization of a physical object. This can be either explained at a basic abstract level or in greater detail in reference to a specific physical object. All different levels of granularity are allowed on the SSHOMP, as long as they concern the SSH domains.

A workflow consists of a series of steps. We start with an action and then explain this action on the basis of necessary steps to perform this action. The workflow may comprise preliminary steps, e.g. how to obtain input data, followed by the action itself which can be separated into different sub-steps, e.g. preparing a physical object for digitization or setting up digitization software, and also possible follow-up steps, e.g. referring to another workflow that explains how to apply text detection on digitized material. The steps of a workflow may refer to any items on the SSHOMP. Therefore, to create a workflow, usually new items have to be brought into the SSHOMP – like tools or training material that were not already registered in the SSHOMP –, and be referred to in the workflow. Workflows are thus a good starting point when adding new content for the SSHOMP, which was one of the main arguments for including workflows as a dedicated item type.

There are two further aims behind integrating workflows in the SSHOMP.

First, workflows give researchers an entry point for discovering tools that they are not aware about. By identifying a suitable action one can find the relevant workflow without knowing beforehand which tools and services or keywords to look for. The workflows are thus a kind of vertically oriented search interface, if we interpret the simple search bar that asks for some background information to formulate search terms as a horizontal way of searching. Additionally, workflows offer their creators a good option to point out features of tools and services in a richer way, compared to the metadata oriented way of expressing an item in SSHOMP, especially when it comes to relating items together. It is possible to show the power of a toolchain, where a complex task can be fulfilled by using a clever combination of tools.

---

<sup>7</sup> On these aspects, see the exhaustive work around the Canonical Workflow Frameworks for Research (CWFR) published in *Data Intelligence*, Volume 4, Issue 2, Spring 2022 *Canonical Workflow Frameworks for Research*, Editors: Peter Wittenburg, Alex Hardisty, Limor Peer & Amirpasha Mozzafari <https://direct.mit.edu/dint/issue/4/2> EISSN 2641-435X.

A second aim is to create more context for the items on the SSHOMP. As workflows use the same set of metadata fields as the other items on the marketplace, this information can be used implicitly for enriching the items that are referenced in a workflow. For example, if we describe a digitization workflow and we have a step on how to set up digitization software which gets the activity term “digitization”, the mentioned tools can be also related to this activity term. By applying such connections, implicit enrichment of other items on the SSHOMP is possible and more context is available, as the referred item will also link back to the workflow. Researchers looking for a dedicated tool can then see what other researchers are doing with it by looking at the workflows that are related to this tool. The more workflows are created, the more context will be created.

In contrast to other items on the SSHOMP, there are not many sources with existing workflows that could be harvested. The common case therefore is that researchers create workflows manually inside the SSHOMP. One important exception is the Standardization Survival Kit (SSK), a project where the idea of workflows was introduced and then transferred to the SSHOMP. The initial set of workflows on the SSHOMP was ingested from this source. The next section focuses on the SSK and multilinguality of workflows in the SSK.

## (2.2) STANDARDIZATION SURVIVAL KIT (SSK)

The project “PARTHENOS”<sup>8</sup> introduced some relevant services which were later on taken over by the SSHOC project. The Standardization Survival Kit (SSK)<sup>9</sup> was such a data source for the SSHOMP. The SSK was sketched as an online service, aimed at raising awareness in using standards when performing digital research activities supporting researchers in choosing the best fitting standard for their activities (Riondet et al. 2018b). The challenge was to find a way to create a recommendation system even for researchers who were not aware of a specific standard. Thus the idea of scenarios came up, which represent a digital method separated in different steps. Tools and technical papers referenced in these steps recommended already existing standards like file formats or data models to reuse. Riondet et al. 2018a:56 highlight the need “to provide potential users with an awareness of the appropriate standards and the advantages to be gained by adopting them (...) [and] to present the cognitive tools to help them identify the optimal use of standards through the selection and possibly customisation of a reference portfolio”. Researchers could then apply these mentioned standards to their research approaches.

The software of the SSK service was custom made, with the SSK scenarios themselves expressed in TEI (Text Encoding Initiative)<sup>10</sup> and hosted on GitHub.<sup>11</sup> TEI is a standard in the Digital Humanities community and due to its expression in XML format, it is easy to read and process. TEI guidelines allow researchers not only to encode text but also to add metadata to it. The scenarios had text passages for each different step and additional references and metadata as context, e.g. which standards to use for the step. Additionally, through a connection to a dedicated Zotero library<sup>12</sup> documents, training material and articles could be referenced.<sup>13</sup>

Amongst many features, TEI enables an easy way to implement multilinguality. By using a language attribute it is possible to have translations next to each other. This feature was also used in the SSK scenarios<sup>14</sup> and was implemented on the SSK website. TEI encodes such

---

8 PARTHENOS project (running time 2015-2019), <http://www.parthenos-project.eu/> – funded by European Commission H2020, grant agreement ID 654119, <https://cordis.europa.eu/project/id/654119> (last accessed date: 12/12/2023).

9 More information can be found in the deliverables of PARTHENOS Work Package 4 on “Standardization” responsible for creating the SSK, see PARTHENOS: Project’s deliverables, <http://www.parthenos-project.eu/resources/projects-deliverables#1523355659817-a8613542-c526>, the documents are available in a dedicated Zenodo community, see Zenodo: PARTHENOS Hub repository, <https://zenodo.org/communities/parthenos-hub> (last accessed date: 12/12/2023).

10 TEI: Text Encoding Initiative, <https://tei-c.org/> (last accessed date: 12/12/2023).

11 GitHub: ParthenosWP4/SSK, <https://github.com/ParthenosWP4/SSK> (last accessed date: 12/12/2023).

12 Zotero: SSK-Parthenos, <https://www.zotero.org/groups/427927/ssk-parthenos> (last accessed date: 12/12/2023).

13 A tutorial and the data model of the SSK is available at <https://ssk.readthedocs.io/en/latest/> (last accessed date: 12/12/2023).

14 See documentation on xml:lang in [https://ssk.readthedocs.io/en/latest/2\\_ssktei.html](https://ssk.readthedocs.io/en/latest/2_ssktei.html) (last accessed date: 12/12/2023).

translations at the level of text editing.<sup>15</sup> As the scenarios are hosted on GitHub, anyone capable and ready to translate a scenario could do this by simply adding the corresponding TEI language tags holding the translation. The drawback of this approach is that it needs at least a short introduction to TEI, GitHub and Zotero. As the SSK software did not implement an editorial interface, contributors needed to use their text editors in a way that created TEI-compliant encoding, which has been proven to be a barrier for entry to add new scenarios. If this did not work, it was necessary to have an editorial team that spent some time on corrections. In terms of multilinguality it was a challenge to deal with only partly translated texts and wrong encodings of translations.

The SSK scenarios became the blueprint for the SSHOMP workflows. Thereby the option emerged to ingest the SSK scenarios to the SSHOMP. This became especially an issue when the SSK software became unmaintainable. Nevertheless such a switch was not straightforward especially because it meant moving away from the open standard TEI to the custom data model and the relational database of the SSHOMP. With this came some changes to features of the SSK. First, communicating the use of standards is not in the focus of the SSHOMP. Secondly, the TEI approach allowed creators to reuse steps of scenarios in other scenarios, which is no longer possible in the SSHOMP. Furthermore, the TEI handling of multilinguality is not covered in the SSHOMP (see discussion in section 4), which does not have a similar language attribute and does not give the option to attach translation to text fields of the SSHOMP data model.<sup>16</sup> On the plus side, the SSHOMP comes with user-friendly input forms and the integration of the SSK scenarios in the richer digital ecosystem of SSHOMP comes with a larger audience and a better way to link workflows to other items. This also helped in simplifying the techstack of the SSK, where instead of three tools – GitHub, Zotero, SSK software – only the SSHOMP is used. These arguments describe the main features of the SSHOMP and illustrate why ingesting a data-catalogue-oriented website into the SSHOMP makes sense, even when losing some functionality.

### (3) SSH OPEN MARKETPLACE

#### (3.1) WORKFLOWS IN THE SSH OPEN MARKETPLACE

The SSH Open Marketplace supports five content types:

- Tools & Services
- Datasets
- Training Materials
- Publications
- Workflows

These content types are curated in line with the SSH Open Marketplace Core Principles:<sup>17</sup> Contextualization, Curation, (Meta)Data quality, Focus on primary sources, Technical interface. This means a high level of interconnectedness between resources. For example, publications are only accepted into the Marketplace when they are related to a tool or service. Similarly, Training Materials are expected to be related to Tools and Services existing in the Marketplace catalogue.

The Workflow content type may be described as a meta-type since it describes interactions between different Tools and Services and Datasets in order to achieve specific outcomes. Workflows include a number of steps and may include different tools, as in [Figure 1](#). Their scope can vary from domain-specific tasks ([Mörth et al. 2019](#)) to large-scale projects (e.g. [Gautier et al. 2019](#)). The scope of a workflow is determined by the workflow creators, with the SSH Open Marketplace editorial team<sup>18</sup> only curating overall quality and adherence to the Core Principles.

---

<sup>15</sup> There are even specialised text editors for TEI that support editors with the features of TEI like creating translations, see [https://wiki.tei-c.org/index.php?title=Category:Editing\\_tools](https://wiki.tei-c.org/index.php?title=Category:Editing_tools) (last accessed date: 12/12/2023).

<sup>16</sup> This will be discussed further in section 5.3. It would be possible on a technical level to find solutions for these missing features in the SSHOMP, but there was not the resources to implement these.

<sup>17</sup> <https://marketplace.sshopencloud.eu/about/data-population#ssh-open-marketplace-core-principles> (last accessed date: 12/12/2023).

<sup>18</sup> The Marketplace editorial team is a group of 17 researchers from SSH backgrounds. The team ensures the day-to-day maintenance and the data quality of the SSH Open Marketplace.

**Workflow steps (9)**

**1** | Create a digital corpus of musical compositions Collapse ▾

Select resources to be included in the corpus. After collecting original musical sources, transcribe them adding critical editorial signs and normalizing, where applicable, ancient poetic texts to modern usage. To get directly MEI files, use MEISE (MEI Score Editor).

**Related items (6)**

<b>MEI Wiki</b> Here are instructions to get MEISE running and a guide to start working with MEI encodings. <a href="#">Read more</a>	<b>MEI Score Editor -MEISE</b> No description provided. <a href="#">Read more</a>
<b>MEISE Native note editor for the MEI encoding format</b> The MEI Score Editor (MEISE) is an eclipse-based music notation editor for the musicological XML-based encoding format of the Music Encoding Initiative - MEI ( <a href="http://music-encoding.org">http://music-encoding.org</a> ). It is developed at the... <a href="#">Read more</a>	<b>Music Encoding Initiative Guidelines</b> The MEI Guidelines, which include the data dictionary, are intended to serve as a reference tool for music encoders. Through the use of natural-language definitions and examples, this document assists users of MEI in... <a href="#">Read more</a>

[Show more](#)

**2** | Change Format into MEI files if necessary. Expand ▾

**Figure 1** Sample workflow showing instructions and resources associated with the first step.

A concrete application of the workflow type to demonstrate research processes and bridge different aspects of the research process may be seen in the workflow *Digitizing Textual Material*.<sup>19</sup> The nine workflow steps cover legal considerations, scanning methods, imaging specifications, image capturing, quality control, data management and post-processing. For each step, different relevant items are linked and accessible through the Marketplace interface (see [Figure 2](#)). These items include publications relating to legal considerations, training materials relating to post-processing and quality control and tools relating to image capturing.

**Workflow steps (8)**

**1** | Assess legal issues Collapse ▾

Before you start working with material, looking into its legal status is a strict requirement. By doing this, you find out what the originator of the material - be it a picture, a document or an object - allows you to do with the object in terms of (re-)use.

**Related items (2)**

<b>Open Data in a Big Data World</b> Four major organisations representing global science, the International Council for Science (ICSU), the InterAcademy Partnership (IAP), The World Academy of Sciences (TWAS) and the International Social Science Council... <a href="#">Read more</a>	<b>The Value of Open Data Sharing: A CODATA Report for the Group on Earth Observations</b> The report on the Value of Open Data Sharing was first prepared for the GEO-XII Plenary by the GEO Participating Organization CODATA (the ICSU Committee on Data For Science and Technology). Through showcasing... <a href="#">Read more</a>
---	--

[Show more](#)

**2** | Determine scanning method Collapse ▾

A thorough examination of the object provides insight in which scanning method is the most appropriate one given the type of material, its dimensions and fragility.

**Related items (1)**

<b>Digitizing Entire Collections: Chapter 4, For Scanning Technicians</b> Page Contents Characteristics of Archival CollectionsPreparing and Reviewing CollectionsHandlingSpecial Problems <a href="#">Read more</a>
--

**Figure 2** Detail of workflow steps and linked resources.

<sup>19</sup> <https://marketplace.sshopencloud.eu/workflow/GAvmqk> (last accessed date: 22/01/2024).

In this example, the workflow offers a more complete picture of the research process than any single tool or publication could have provided. The workflow can expand the scope beyond individual tools and provide users with far more context than unconnected listings of tools could.

In contrast to the other content types, workflows show the tools in motion and describe possible pathways of interaction between the different tools and data sources. These interactions are organised in steps, which guide the user through the workflow. A workflow can have as many steps as needed, and each step can refer to different resources such as Tools and Services and Datasets. From a data management perspective, this means that workflows are more complex to integrate into the Marketplace, as all referred tools and data sources must also be present in the Marketplace. These mapping challenges will be explored in the next section.

### (3.2) MAPPING CHALLENGES

The technical architecture of the SSHOMP prioritises the integration of data from various sources, as manually looking for items for the SSHOMP would cost too much time. By implementing an ingest pipeline that automatically harvests online sources based on mappings it was possible to gather more than 7000 items for the SSHOMP in a relatively short time.

An important technical part for such an approach is the backend part, which is only accessible via an API. This makes the mapping and import of data sources easier, since data only has to conform to the specifications of the API, and from there on is automatically translated into the database schema. It also extends the SSHOMP for the use of analysis tools, e.g. doing moderation tasks such as merging duplicate entries. The access to the API is open, it is documented on the SSHOMP website and it allows for everyone interested in looking into the data.<sup>20</sup>

For the ingest pipeline one project partner – Semantic Web Company<sup>21</sup> – provided a workflow built on their digital infrastructure.<sup>22</sup> Later on a second partner – Poznan Supercomputing and Networking Center – granted access to their ingest workflow called Data Aggregation and proCessing Engine (DACE),<sup>23</sup> which is the one in use today. Both workflows are based upon common ETL (extract/transform/load) approaches. First there is a mapping between the data model of the source and the one of the SSHOMP. Then the data exchange point is defined from where to get the source data. For SSHOMP we have around a dozen sources with many disparate formats that we need to deal with, e.g. CSV files, JSON and XML API responses. The mentioned SSK had a combination of TEI-files on GitHub and references to a Zotero library, both being mapped. After getting the data, the next step is the transformation of the different data types that are in use at the source. Not every field is translatable to the destination data model and some need a conversion. Finally, the transformation results are loaded into the SSHOMP.

The many challenges when mapping and transforming data from the source, especially when it concerns workflows, can be illustrated by the example of the ingestion of the SSK into the SSHOMP.

The SSK focuses on communicating standards whereas for the SSHOMP references to other items are at the centre. This means that the resulting workflows in the SSHOMP do not highlight the use of standards as much as in the SSK scenarios. This made a post-ingestion curation necessary so that the workflows fit better to the mission of the SSHOMP, e.g. collecting information on too abstract steps where only the relation to a standard was highlighted in the SSK.

Another challenge was the mapping to the controlled vocabularies of the SSHOMP. The SSK used different ones as the SSHOMP, e.g. when referencing research disciplines. This was solved by creating time-consuming manual mappings.

---

<sup>20</sup> SSH Open Marketplace: API Documentation, <https://marketplace.sshopencloud.eu/about/api-documentation> (last accessed date: 12/12/2023).

<sup>21</sup> <https://semantic-web.com/> (last accessed date: 12/12/2023).

<sup>22</sup> <https://www.poolparty.biz/> (last accessed date: 12/12/2023). The technical framework is called PoolParty UnifiedViews.

<sup>23</sup> <https://gitlab.pcss.pl/dl-team/aggregation/dace> (last accessed date: 12/12/2023).

Finally, one challenge was the identification of ingested items from other sources. The relation to such items was expressed in the SSK via a Zotero library holding manual created entries, where different interpretation of titles made it hard to find concordances. There should be no duplicates in the SSHOMP which was challenged hard by the SSK scenarios ingestion. In the end, it was also necessary to rely on manual curation.

These challenges are not limited to workflows alone, but the complex composition of workflows and the intention to create contextuality with them, make them a special kind of item in the SSHOMP especially when ingesting them from a data source. Thus, it seems to be a more effective way of creating them by hand.

### (3.3) COMMUNITY DRIVEN WORKFLOWS

Workflows in the SSH Open Marketplace are a way to combine two of our essential pillars: contextualization and community. We have found that in-person workshops are among the most efficient ways to popularize workflows and convince scholars to upload their methodology – in a new format which is not necessarily well-known. Indeed, having access to a member of the Editorial Board on site to answer any relevant questions, related to metadata or using the Marketplace, and immediately approve suggestions, is a force multiplier that allows researchers to work more rapidly and holistically on their workflow, avoiding lengthy delays as one waits for feedback on a particular question. In addition to onsite presence of at least one Editorial Board member, experience has shown that these workshops must also be of sufficient length to ensure that researchers have time to document, reflect upon, and write their workflows as they would for any other scientific publication.

For we consider that SSH Open Marketplace Workflows are truly scientific publications, as they show step by step how one researcher approaches a scientific problem – valuable insight for the reproducibility of research and fully inscribed in the context of Open Science. As well, by placing the focus on the tools and methods one uses, it encourages researchers to share important parts of the research life cycle that are too often left unsaid in traditional publications such as journal articles or books. Promotion of these workflows, as well as hosting workshops on the creation of workflows, is part of the work of the Marketplace Editorial Board.<sup>24</sup>

As well, these workflows – as any SSHOMP entry – can evolve over time, as do digital humanities research processes, ensuring that scholars can keep their methodology up to date. It was important for the SSHOMP team to ensure that entries could be updated, and that versioning could be made possible to account for the changing nature of research. When a workflow is edited, the item identifier remains stable, as does the item URL. Also, by sharing their methodologies, researchers are opening themselves up to collaborative opportunities where other researchers can evaluate their work and make suggestions for improvement. While not quite full peer review as in journals like this one, workflows on the SSH Open Marketplace allow for researchers to receive feedback from the community on their research processes. This is particularly aided by the user-friendly interface of the SSH Open Marketplace, which gives it a distinct advantage over other options and lowers the barrier for entry.

Based on our experiences, researchers seem to be convinced by the pertinence of SSH Open Marketplace workflows, with over twenty different workflows added since the end of the project.<sup>25</sup> For example, the workshop hosted in Paris in November 2022 with French national research infrastructure IR\* HumaNum's Consortiums — large, transdisciplinary working groups that are funded to produce tools and document best practices and research methods<sup>26</sup> — led to over 55 entries being created or updated. Supported by two members of the Editorial Board, these researchers were able to add different tools, training resources, publications, and indeed, workflows. Split across two days, researchers had the opportunity to reflect on how to present their research methods, and plan ahead for which resources they wanted to add to the SSH Open Marketplace. Indeed, the Consortiums and the IR\* Huma-Num team were so

---

<sup>24</sup> For example, workshops were accepted for the Dhd 2023, Clarin annual conference 2023 and Dhd 2024.

<sup>25</sup> As of 08 December 2023, there are 48 workflows in the SSH Open Marketplace, with 27 existing from the Standardization Survival Kit (described above) and another handful related to internal use by the Editorial Board (for example, the Workflow on how to create workflows in the SSH Open Marketplace). <https://marketplace.sshopencloud.eu/search?categories=workflow&order=label>.

<sup>26</sup> <https://www.huma-num.fr/les-consortiums-hn/>.



impressed with the output, that they decided to continue the exercise in the future and require that Consortiums place their materials in the SSH Open Marketplace, fulfilling their mission to communicate their research productions and processes in the European context.

## (4) OUTLOOK AND FUTURE WORK

### (4.1) REAL WORKFLOWS ARE NOT LINEAR

Research workflows are not always linear. Instead of following a sequence of steps towards a clearly-defined end result, they meander, iterate, react to unforeseen problems and produce new data. The formalism of the current SSH Open Marketplace Data Model does not allow for such forks, loops and break points in a research workflow. While this helps to keep workflows accessible, it does not necessarily allow to represent practical examples and might hinder reusability of workflows. In a sentiment analysis workflow, for example, the input text may have to undergo a number of cleaning and preparation operations depending on type and source of the text. Where in practice a researcher might iterate the cleaning process using a small subset of text, a workflow in the SSH Open Marketplace may hint at this option, yet would be unable to describe it using the logic of linear steps.

A future expansion of the Marketplace data structure can include the introduction of conditions and flow controllers between the individual steps. Using this new logic, workflows in the Marketplace would be able to express more life-like workflows such as *repeat the data preparation process with a small sample until desired quality is achieved. Then, move back to the start and prepare all data according to the settings from the sample*. The workflows could also incorporate conditional pathways, such as *if x arises, refer to step y*.

While these additions have the potential to make the workflows more complex overall, they would also allow researchers to document their work more accurately in the Marketplace.

### (4.2) MULTILINGUALITY


The SSH Open Marketplace comes for the moment only with an English user interface and the majority of the records are in English. There are both advantages and disadvantages for a monolingual discovery platform. For instance, among the advantages are the increased findability of resources when users use English keywords, and the use of English controlled vocabulary. The counterpart is the need to translate into English the names and the descriptions of resources which are not in English knowing that the resources themselves may not be useful to users who do not speak the language in which the resource is.

To assess the impact and potential extension of the SSHOMP through multilinguality, it is important to understand different levels of multilinguality in the context of the SSHOMP. As the marketplace is an aggregator and discoverability platform, its original content is the metadata, not the resource itself. It is of course possible to (automatically) translate the interface and embedded metadata, as in [Figure 3](#).

However, the resource itself, as well as virtually all documentation and accompanying resources, would still be in English. It is hard to imagine a use case where a user would find it useful to see translated metadata while the resource itself is only available in English. For a more unified user experience, it would be necessary to also provide the resources in languages other than English. As a resource aggregator, SSHOMP cannot provide such translation work at the moment.

However, the Editorial Board encourages the addition of records about resources in other European languages, while providing recommendations to increase the search experience of non-English records. First, the name of the entry or the label should be in English. This would facilitate the discoverability of the resource because if the name or entries were in different languages than English, the users would not know which languages their search terms should be in, creating a very confusing user experience. If the English name of the resource does not exist or if it does not make sense to translate it, the recommendation is to include a description of the resource (in English) in the title (e.g. Portal xx, Corpus xx). Second, the description of a resource can be in another language than English but adding at least a short description in English is recommended to increase the discoverability of the resource by users with all

Heim / Werkzeuge und Dienstleistungen / OpenCOLLADA

 **OpenCOLLADA** ↗ In die Zwischenablage kopieren

COLLADAMax und COLLADAMaya sind neue Implementierungen eines 3ds Max- oder Maya-Plugins zum Exportieren von Szenen oder Teilen davon in eine COLLADA-Datei, veröffentlicht unter einer MIT-Lizenz. Im Gegensatz zu anderen bestehenden COLLADA-Exportprogrammen speichern diese neuen Plug-Ins das COLLADA-Dokument nicht in einem Zwischendatenmodell, sondern schreiben es direkt in eine Datei. Dies führt zu einer drastischen Reduzierung des Speicherverbrauchs und zu einer deutlich besseren Leistung. Weitere Informationen zu den Plug-Ins und deren Erstellung finden Sie in den README-Dateien in den Verzeichnissen COLLADAMax und COLLADAMaya. Informationen zum Erstellen von OpenCOLLADA unter Linux und Mac OSX mit SCons finden Sie in der BUILD-Datei. HINWEIS: Die in diesem Verzeichnis enthaltene Lösung COLLADA.sln dient nur Entwicklungszwecken. Um die NextGen-Plug-Ins zu erstellen, verwenden Sie bitte die Lösungen in den Verzeichnissen COLLADAMax und COLLADAMaya.

[Gehen Sie zu Werkzeug oder Service ↗](#)

### Einzelheiten


Quelle : [SSK Zotero-Ressourcen](#)

**INFORMATIONSGEBER**

Systemimporteureur  
Laure Barbot


Zitieren als: OpenCOLLADA. Abgerufen am 7. Dezember 2023 von <https://marketplace.sshopencloud.eu/tool-or-service/5mpQoK>

### Ähnliche Artikel ( 2 )

 **Migration ins Zielformat**

Aufgrund der Vielfalt der Digitalisierungsgeräte und -techniken kann ein breites Spektrum an Datenformaten erzeugt werden. Daher ist es wichtig, den Inhalt und die Struktur der erstellten Dateien zu überprüfen. Dieser...

[Mehr lesen](#)

 **Georeferenzierung der vermessenen Daten: Verwendung eines GNSS-Empfängers zur präzisen Positionierung von Bodenkontrollpunkten**

Die Verwendung gut materialisierter Bodenkontrollpunkte (GCPs) und die Verwendung desselben Satzes bei allen durchgeführten Flügen erfordern den Einsatz topografischer Instrumente nur für die erste Erfassung, da die GCPs...

language backgrounds. Third, for the same reasons of discoverability discussed above, the use of English keywords is strongly encouraged. Finally, the language of the resource should be specified in the dedicated metadata field.

**Figure 3** Automated German translation of a SSHOMP entry through Google Translate.

When it comes to workflows, what remains to be clarified is what a multilingual workflow is and how it should be created. For instance, does it entail allowing for content in different languages, or also translating the wording of the steps and of the controlled vocabularies? In addition, a connection should be kept between the different translations of the same workflow. Multilingual workflows would also require not only to be able to support multilingual entries (from a technical perspective) but also to translate them and to map the entries.

From a technical perspective, implementing multilingual content requires changes to the data model and the underlying database. As one possible solution, multilingual content can be supported through the use of nested API calls. For example, the current “description” field is a string value. When replaced by a API call such as

```
/api/tools-services/{persistentId}/description?language=fr
```

where fr (French) is the code for the requested language, multiple descriptions can be shown. Similar solutions would have to be found for most other data fields and the frontend part of the SSHOMP website would also need to support this. This brief technical example shows some of the adjustments necessary to support multilingual entries in SSHOMP.

While we acknowledge the importance of multilinguality, we leave this aspect for the SSHOMP for future work.

### (4.3) IS THE VALUE OF WORKFLOWS IN THEIR REUSABILITY?

The intrinsic value of workflows within the SSHOMP is predominantly anchored in their reusability. Reusability, in this context, pertains to the capability of a workflow to be applicable in diverse research contexts or projects. This characteristic hinges on the foundational design of the workflow, which must be sufficiently general to ensure its applicability across a spectrum of scenarios. Within the SSHOMP framework, workflows are meticulously crafted as sequences of steps, adaptable to a variety of research data and projects. This inherent adaptability and flexibility are pivotal in enhancing their reusability.

Conversely, the concept of reuse involves the practical implementation of these reusable workflows within new or differing project contexts. In the realm of SSHOMP, workflows transcend theoretical constructs, finding active utilization by researchers. This practical application is instrumental in streamlining and standardizing research processes, thereby reinforcing the practical utility of these workflows.

The promotion of workflow reusability presents a blend of advantages and challenges. Among the advantages, reusable workflows are instrumental in significantly diminishing the time and effort required to initiate new research endeavours, offering a pre-established procedural template. They play a crucial role in harmonising research methodologies across varied projects, a factor essential for ensuring the comparability and reproducibility of research outcomes. Furthermore, the encouragement and dissemination of reusable workflow creation cultivate a robust community ethos among researchers, facilitating the exchange of knowledge and collaborative enhancement of research methodologies. Additionally, reusable workflows serve as a catalyst for innovation, prompting researchers to refine and adapt established methods to novel research contexts, potentially catalysing methodological breakthroughs.

However, the encouragement of reusability is not devoid of challenges. Firstly, the development of universally applicable workflows necessitates a profound comprehension of commonalities in research processes across diverse projects, a task that can be complex and demanding. Secondly, maintaining a consistent standard of quality and relevance for these workflows across a broad array of research scenarios is an ongoing challenge. Lastly, the comprehensive documentation of workflows for effective reuse and the efficient communication of their potential applications demand considerable resources.

In conclusion, the significance of workflows in SSHOMP is fundamentally derived from their reusability. Despite the challenges associated with their creation and maintenance, the benefits they offer in terms of efficiency, standardization, community collaboration, and innovation render them an invaluable component in the research process. The emphasis placed on workflows within SSHOMP is a testament to the increasing recognition and appreciation of these benefits within the academic sphere.

## FUNDING INFORMATION

The “Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies” PARTHENOS project has received funding from the European Union’s Horizon 2020 project call H2020-INFRADEV-1-2014-1, grant agreement #654119, see European Commission: CORDIS, <https://cordis.europa.eu/project/id/654119>.

The “Social Sciences and Humanities Open Cloud” SSHOC has received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-04-2018, grant agreement #823782, see European Commission: CORDIS, <https://cordis.europa.eu/project/id/823782>.

## COMPETING INTERESTS

Barbara McGillivray is editor-in-chief of the *Journal of Open Humanities Data* but did not take part in the editorial process or decisions pertaining to this manuscript.

## AUTHOR CONTRIBUTIONS

Klaus Illmayer – Writing – original draft

Laure Barbot – Writing – original draft

Barbara McGillivray – Writing – review & editing

Cristina Grisot – Writing – original draft

Michael Kurzmeier – Writing – original draft

Edward J. Gray – Writing – original draft

Maja Dolinar – Writing – original draft

## AUTHOR AFFILIATIONS

**Laure Barbot**  [orcid.org/0000-0002-6008-7959](https://orcid.org/0000-0002-6008-7959)

Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU), Paris, FR

**Maja Dolinar**  [orcid.org/0000-0002-8307-6486](https://orcid.org/0000-0002-8307-6486)

University of Ljubljana, Slovenian Social Science Data Archives: Ljubljana, SI

**Edward J. Gray**  [orcid.org/0000-0002-5201-1014](https://orcid.org/0000-0002-5201-1014)

Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU), Paris, FR; IR\* Huma-Num (CNRS UAR 3598), FR

**Cristina Grisot**  [orcid.org/0000-0003-0684-4442](https://orcid.org/0000-0003-0684-4442)

Swiss National Data and Service Center DaSCH (DARIAH-CH), Basel, CH

**Klaus Illmayer**  [orcid.org/0000-0001-7253-996X](https://orcid.org/0000-0001-7253-996X)

Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences, Vienna, AT

**Michael Kurzmeier**  [orcid.org/0000-0003-4925-5197](https://orcid.org/0000-0003-4925-5197)

Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences, Vienna, AT

**Barbara McGillivray**  [orcid.org/0000-0003-3426-8200](https://orcid.org/0000-0003-3426-8200)

Department of Digital Humanities, King's College London, London, GB

## REFERENCES

- Gautier, H., Joffres, A., & Besson, V.** (2019). "Collaborative Digital Edition of a Musical Corpus." *SSH Open Marketplace*, n.d. <https://marketplace.sshopencloud.eu/workflow/bKeFY4> (last accessed date: 12/12/2023).
- Leonelli, S.** (2023). *Philosophy of Open Science* (1st ed.). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/9781009416368>
- Mörth, K., Riondet, C., & Tadjou, L.** (2019). "Create a Dictionary in TEI." *SSH Open Marketplace*, n.d. <https://marketplace.sshopencloud.eu/workflow/4qFarh> (last accessed date: 12/12/2023).
- Riondet, C., & Romary, L.** (2018a). *The Standardization Survival Kit: For a Wider Use of Metadata Standards within Arts and Humanities*. *Archives et Bibliothèques de Belgique – Archief- En Bibliotheekwezen in België*, 106, 55.
- Riondet, C., Seillier, D., Tadjou, L., Romary, L., Sanz, M., Illmayer, K., & Bassett, S.** (2018b). "PARTHENOS D4.3 Standardization Survival Kit – Final". *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.2578960>
- Rocca-Serra, P.**, et al. 'The FAIR Cookbook'. Apr. 2022. [Online]. Available: <https://github.com/FAIRplus/the-fair-cookbook/>
- Social sciences and humanities open marketplace.** (2023). About the data population. Webpage. <https://marketplace.sshopencloud.eu/about/data-population#types-of-content> (last accessed date: 5/12/2023).
- Tasovac, T., Romary, L., Tóth-Czifra, E., Ackermann, R. C., Alves, D., Chambers, S., Cosgrave, M., Denoyelle, M., Garnett, V., Gautschy, R., Gray, E., Malínek, V., di Meo, C., Perkis, A., Reinsone, S., Rišler-Pipka, N., Scharnhorst, A., & Viola, L.** (2023). The Role of Research Infrastructures in the Research Assessment Reform: A DARIAH Position Paper. <https://hal.science/hal-04136772>

### TO CITE THIS ARTICLE:

Barbot, L., Dolinar, M., Gray, E. J., Grisot, C., Illmayer, K., Kurzmeier, M., & McGillivray, B. (2024). Contextualizing Research Tools & Services through Workflows in the SSH Open Marketplace. *Journal of Open Humanities Data*, 10: 22, pp. 1–12. DOI: <https://doi.org/10.5334/johd.192>

**Submitted:** 14 December 2023

**Accepted:** 30 January 2024

**Published:** 29 February 2024

### COPYRIGHT:

© 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Journal of Open Humanities Data* is a peer-reviewed open access journal published by Ubiquity Press.