Addendum

to the document on

Collaborative Use Cases between

SSH Open Marketplace (WP7),

Switchboard and Virtual Collection

Registry

July 2021

ADDENDUM TO COLLABORATIVE USE CASES DOCUMENT

About this document

This document is an addendum to the Collaborative Use Cases between the SSH Open Marketplace, Switchboard, and the Virtual Collection Registry, published via Zenodo in November 2020[1] [=source document].

Summary

[Note: Revised summary from source document.] The CLARIN Language Resource Switchboard (short: Switchboard) serves as an established and valuable asset to inform the researcher on tools and services for her or his respective research data.

Apart from looking at an already existing set of research data, moreover, many users would like to search for tools and services from the angle of the research method, a certain technology, interoperability or even just a research question. Such search interests may be served by the SSH Open Marketplace (MP). The MP not only strives to provide information on individual items (tools, research data, tutorials, software) to the searching and browsing user, but also context. Contextualized items in the MP allow for a search serendipity which contributes substantially to the service experience. The Switchboard promises to be a suitable means to convey such serendipity. For this purpose, the source document outlined possible user stories in favor of its integration into the MP.

Beyond the Switchboard-MP relation the document also considered scenarios for the relation of both components to the CLARIN Virtual Collection Registry (VCR). The VCR allows the researcher to create individual and persistent collections including records from a broad range of sources. Such sources may be repositories exposing research data (the common VCR use case= 'bibliography of research datasets'), but possibly also from other sources such as the MP.

---

[1] Buddenbohm, Stefan, Broeder, Daan, Eisner, Marthe Irene, Illmayer, Klaus, & Durco, Matej. (2020, November 5). Collaborative Use Cases between SSH Open Marketplace and the Language Resource Switchboard and Virtual Collection Registry. Zenodo. http://doi.org/10.5281/zenodo.4442320

The overarching goal of the T3.6 activity within WP3 of the SSHOC project is to achieve - wherever useful and technically possible - an integration between CLARIN and DARIAH components. This includes the MP, the VCR and the Switchboard.

This document - the addendum to the source document - serves as conclusion of the work within the SSHOC project and summarises the implementation status of the proposed use cases from the source document. Some of the use cases have been implemented during the project, others have been suspended or even discarded. The team of T3.6 sees this kind of documentation as important information for the future work on the services and for potential external players and partners, who are informed on the conceptual discussions in CLARIN, DARIAH, and SSHOC in this regard. Particularly the latter aspect is important as the service owners - here: CLARIN, DARIAH, SSHOC - encourage the re-use, adaption and broad uptake of their services in the EOSC or other research infrastructure contexts.

Author List

Organisation	Name	Contact Information
CLARIN ERIC	Daan Broeder	daan.broeder@di.huc.knaw.nl
UGOE	Stefan Buddenbohm	sbudden@gwdg.de

Implementation status

This document reflects the implementation status as of July 2021. The use cases listed below may be extended, and additional new use cases may be possible, however this will not be done within the SSHOC project. As stated before the tasks members encourage the re-use, adaptation and broad uptake of the mentioned services.

More specifically the goals of T3.6 were:

> - extend the visibility and use of Switchboard, VCR as well as MP
>
> - utilise Switchboard or VCR to provide context on data resources, including items available in or via the MP
>
> - bond of language resource-affiliated researchers to the MP
>
> - introduce new services working with new types of data to the Switchboard that are useful for the broader SSH and overcome the language resources bias; promote the integration between the SSHOC work packages WP3 and WP7

The authoritative description of the uses cases is to be found in the source document:

> *Buddenbohm, Stefan, Broeder, Daan, Eisner, Marthe Irene, Illmayer, Klaus, & Durco, Matej. (2020, November 5). Collaborative Use Cases between SSH Open Marketplace and the Language Resource Switchboard and Virtual Collection Registry. Zenodo. http://doi.org/10.5281/zenodo.4442320*

A - Creation of a Virtual Collection with Records from the MP

This use case describes the use of the VCR to include records from the MP. The virtual collection does not have to be entirely composed of MP records, they may form only a subset.

Presentation of content in the MP: The MP provides the user with information on useful resources in the SSH research domains. Usually these resources will be individual items (e.g. tools, tutorials, software) or collections of resources grouped together (e.g. in the form of a workflow or research process). Particularly the second scenario seems to be attractive as it offers search serendipity to the user and sets individual items into an explanatory context. However, with this complexity also comes the requirement for adapting and storing these collections.

<u>Scenario</u>: If a researcher finds a useful set of items in the marketplace he or she might be interested in persisting the set as a collection, to adapt to specific needs and to save it for later work. This calls for an integration of the VCR.

<u>What is a collection in this context?</u> The collection can either be composed of workflows from the MP (similar to the SSK scenarios[2]), which are annotated by the user OR can develop into a completely new collection, e.g. a search result list, which gets refined by the user and needs to be stored. An idea for a virtual collection revolves around teaching. Such a virtual collection could serve as a basis for teaching on a specific subject and collect, for instance, text, tools, tutorials, and exemplary datasets. A strenght of the VCR comes with its flexibility for annotating the collection elements.

An important aspect of the VCR affecting also the other user stories is related to its collection approach: One of the services' strengths, but also a weakness of the VCR is that VCs can be overarching different repositories and registries. This implies that the VC metadata needs to remain relatively agnostic[3] with regard to the purpose and idiosyncrasies of a particular repository. For example, if a user wants to describe workflows and to connect this description with data collections in a VC, the metadata should be part of a VC constituent rather than being part of the VC metadata (which is the case when the VC refers to the resource's metadata/landing page rather than the resource itself).

<u>Requirements and challenges</u>: Stored collections belong to the user and may also be actionable and able to invoke the Switchboard. In order to realize this, the first option requires an individual user space in the MP, where these collections can be stored. Due to the current state of development it is not clear if such individual user spaces will be created. Alternatively, these collections could be stored within the VCR itself and get a PID. The second option requires a linkage between the MP and the Switchboard. At the current stage it is unclear how the MP resource will be citable. A PID seems to be difficult as the MP's resources will get updated frequently. Nevertheless, it is likely that the researcher, in order to share with colleagues and to cite in own documents, expects the possibility to refer to (at least certain) MP entities. It is not always required that the entities are 'frozen' depending on a versioning policy, the PID could refer to the latest version. This is a technical requirement to be solved on the MP's side.

<u>Benefits</u>: A linkage between the three components (Switchboard, VCR, MP) seems reasonable because the MP will be an attractive source for creating collections motivated by the quantity and

---

[2] E.G. http://ssk.huma-num.fr/#/scenarios/SSK_sc_linguisticAnnotationOfCorpora
[3] Seen from a repository point of view, this aspect also applies to the dhrep, the DARIAH-DE Repository (e.g. https://repository.de.dariah.eu/1.0/dhcrud/21.11113/0000-000B-C8F1-3/index) which uses DCsimple for the describing metadata of its content. Technically this allows for a broader range of functionalities, from a research point of view, DCsimple is a substantial compromise.

diversity of its content. On the other hand this may motivate users to enhance the marketplace's contents and extend the usage of the Switchboard and VCR.

Feasibility: Regarding the other user stories, the creation of virtual collections containing MP items seems to be one of the more obvious and useful ones. However, the feasibility for its implementation is dependent on the possibility to reference MP items in a persistent way. This implementation effort has to be yielded on the MP's side. Alternatively the VC could be created on the basis of the original PIDs of the items provided by the items originators but this is clearly a compromise.

Implementation status

> PARTLY POSSIBLE/IMPLEMENTED
>
> This use case depends on the persistent citability of the records in the MP. Without persistency the virtual collection containing records from the MP will become porous over time. At this point the records in the MP own individual IDs which will stay stable over record revisions (due to updates or enrichments), but these IDs don't equal PIDs.
>
> With the willingness to take this risk, the researcher may create VCs with records from the MP and the use case may be decribed as partly possible. An additional implementation effort was not necessary up to this point.

B - Inclusion of the VCR as Feature in the MP

This use case outlines the integration of the VCR as a feature directly in the MP. Although the use case A covers the user's requirement to create virtual collections containing records from the MP it is possible to think of the VCR as genuine MP feature.

Scenario: This scenario addresses the demand of users to create collections within the MP, including MP records, but also including records from other sources.

Requirements and challenges: The implementation effort for this use case relates at least to two main issues: the depth of integration of the VCR in the MP. This should be done in a way that the user experience is as seamless as possible and the user experiences the function as genuine to the MP. The second issue relates to non-MP records. The unique character of the VCR is to gather records from various sources in an actionable way. Without this the virtual collection would be just a conventional list of records. The actionability of an individual record depends on its metadata.

As long as the necessary metadata for a record is held by the MP or is available in an easy way through external sources, the user experience will be sufficient.

Benefits: Although it may be possible to create a virtual collection in the VCR directly, which may include MP records it may be a threshold for the user to do so. It is a threshold for the user to leave the MP and switch to a separate web service for the creation of a virtual collection. From a usability point of view it is more convenient to create the collection directly in the MP. Apart from the usability, such a function might offer valuable information for the MP. Thought from the angle of contextualisation the user-created virtual collections might indicate relations between records that are so far not considered. A user-created collection might also hint at valuable sources that have not been ingested into the MP so far.

Feasibility: The implementation effort of this use case relates to the necessary level of awareness between the two services, MP and VCR. At least four levels of awareness can be identified (coming at different costs):

- If the creation function stays outside the MP and is being done in the VCR, reading of the MP API may be sufficient. In this scenario the VCR only retrieves information from the MP through its API and apart from this the whole process and user interaction stays within the VCR. This is covered by use case A.
- More beneficial from a user's point of view would be to create and present the virtual collection within the MP. This requires implementation effort as it should be possible to include sources outside the MP. This could be achieved by - for instance - that the MP has a button to create/ add selected items or search results to a virtual collection via the VCR API. The intrusion in the graphical user interface of the MP GUI could be minimal.
- Even more beneficial would be making VCs visible in the MP GUI Then they can be shared with other users, adapted and extended. This resembles the "shopping-basket use-case", where from different repositories and tools search results can be added to a virtual collection. Aim is to minimise the impact on the application (here the MP). So in this case making relevant VCs visible in the MP GUI is feasible, but adaption should be left to the VCR.
- Beyond the above-mentioned implementation effort the question of the non-MP records arises. In which way can non-MP records be presented if they are part of a virtual collection? The MP won't hold any describing metadata for such records. Displaying such records as wildcards may be dissatisfying for the user but possibly the compromise to implement this use case.

Implementation status

> NOT IMPLEMENTED/SUSPENDED
>
> The shopping basket function of the VCR is potentially useful for this use case. A user may store MP search results in a VC for future referencing and also use that VC to store queries from other sources (for instance training material, registries).
>
> Currently a direct integration of this shopping basket function in the MP is not planned due to the priority of other development work. However this may be a promising feature on a future roadmap of the MP.

C - Invocation of the Switchboard from the MP based on MIME Type

This use case describes a possible relation of the MP to the Switchboard.

If resources in the MP come with MIME-types - as it should be the case for certain types of research data - it could be conceptually possible to invoke the Switchboard from the individual resource site of the MP. This invocation could be implemented in a similar way as it had been done with the TextGrid Repository here:

https://hdl.handle.net/11858/00-1734-0000-0005-1421-2

The Switchboard itself can already determine the data-type of a resource if it gets the resource URI.

Scenario: The user can quite conveniently click on the 'Switchboard invoke button' and gets the instant results from the Switchboard through a pop-up. The user does not leave the MP in this scenario. The displayed results are based on the file's MIME type and will be a selection from the Switchboard tool inventory:

https://switchboard.clarin.eu/tools

https://beta-switchboard.clarin.eu/ (beta version of the Switchboard)

Requirements and challenges: A main challenge will likely be the composition of content in the MP. At the current stage of development the MP aims at tools, tutorials, software, and collections, but not primarily. The described use case relies on the existence of resources coming with certain MIME types. This is a technical requirement resulting from the Switchboard side as the identification of presented research data. The subsequent recommendation of selected tools relies on the MIME type. For instance, research data with a text related MIME type will lead to the recommendation of text processing tools. To get useful recommendations, this mechanism relies

on a well selected tool inventory in the Switchboard and on the transmission of MIME types, which have to be as specific as possible.

With regard to the MIME type: the MP doesn't need to know the mediatype because Switchboard has its own mediatype/language detector. Mediatype and language are not mandatory, and in fact currently ignored on the Switchboard side.[4]

Another challenge may be related to the way the MP collects and stores its content. As a reminder: The MP does not host content - the resources. The MP just presents describing metadata ON the resources. The resources themselves remain at third party locations, usually the creators or holders. For the Switchboard invocation this poses a technical problem: To be able to identify the dataset by its MIME type, the Switchboard needs access to the dataset OR needs the MIME type in the metadata. Both options seem feasible as the MP metadata includes both the MIME type and can include a link to the original dataset.

Benefits: The benefit of this use case is the additional uptake of the Switchboard by users browsing the MP, who so far are not aware of the Switchboard. For the MP it may offer an additional function, although not feasible for all types of MP resources. This scenario is beneficial, because it raises awareness for the CLARIN research infrastructure in general and a wealth of potential useful tools. However, currently this may only be applicable for MP users with a research background in language related research topics or linguistics. Note that Switchboard extensions are under consideration that allow filtering the list of applicable tools on the basis of a domain indicator that would then limit the list to those tools related to a specific domain eg. DARIAH or Cultural Heritage.

Feasibility: An implementation seems feasible with regard to the transfer of metadata between the two services. As soon as it is technically possible to transfer metadata other scenarios may become possible as well. Most likely, the transfer of MIME type would be a good starting point. This could also be done with suitable resource links to which the Switchboard gets pointed. Reason for independent type checking of the Switchbaord is the sometimes misleading metadata (wrong MIME types) resulting in bad user experience. So a resource link is more preferrable than a landing page.

---

[4] For a comprehensive documentation of this topic consult the Switchboard Github repository: https://github.com/clarin-eric/switchboard-doc/blob/master/documentation/IntegrationProvider.md#2-redirecting-the-client-browser-to-the-switchboard-site
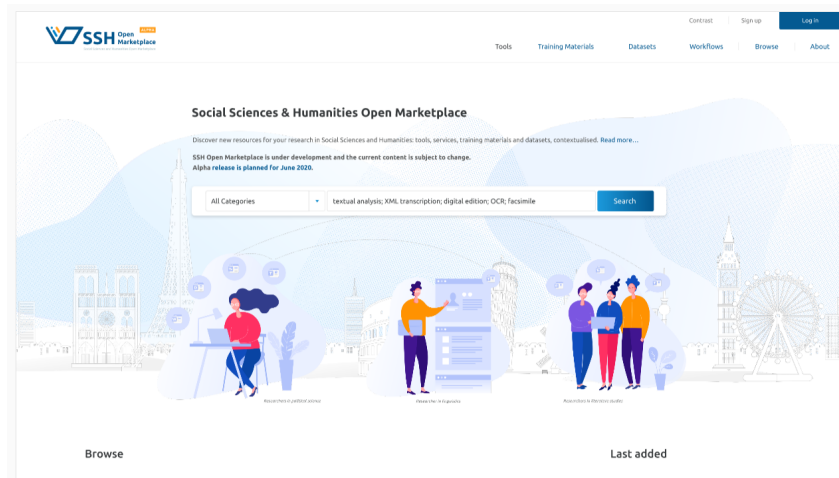
Figure 1: Using the SSH Open Marketplace search slot (Mock up by Justyna Wytrążek)

The implementation in the MP could be created in two ways: Either staying in the MP environment after the Switchboard invocation or leaving the MP and turning the user to the Switchboard. As CLARIN, DARIAH and SSHOC are closely collaborating in lifting services and resources within the cloud, the technical costs for implementation could tip the scale in this use case. Apart from the technical aspect - the transfer of metadata from the MP to the Switchboard - this scenario requires the existence of links to suitable resource and MIME types or other types of metadata - describing a resource - on the MP side.
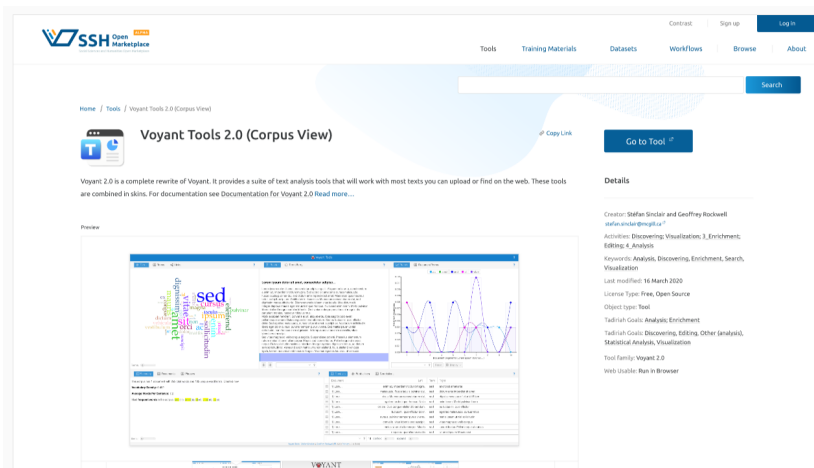

Figure 2: Detailed item view (Mock up by Justyna Wytrążek): on the right side of the item view the VCR invocation or Switchboard invocation could be implemented.

Implementation status

> **NOT IMPLEMENTED/SUSPENDED**
>
> This use case is hampered by the composition of records in the MP. The MP contains links to a broad range of content types, e.g. training material, data sets, software, services, or documentation material. Only for content types with a traceable MIME type the use case is reasonable as the look-up function of the Switchboard relies on the MIME type of the presented resource.
>
> This use case would fit perfectly for a research data repository and only to a lesser degree to the MP.
>
> As note: T3.6 implemented a look-up function for suitable tools for a specific research data type with other services apart from the MP.

D - Invocation/Call up of MP from Switchboard with a Search String

This use case describes an invocation of the MP from the Switchboard and can be seen as reciprocal to the above mentioned use case B. Technically it is not an invocation, but merely sending a search string (e.g. 'show me similar tools like XY') from the Switchboard to the MP.

Scenario: In this scenario the user visits the Switchboard, for instance, browses through the tool inventory. The user may have a certain type of research data in mind but the focus lies on the tools presented by the Switchboard. Beforehand, the user may also have uploaded a dataset to the Switchboard and now browses through the list of recommended tools. For example, the user uploaded a TEI file and wants recommendations beyond the available tools of the Switchboard for this resource type.

In both cases the user finds a call up link at each individual tool within the Switchboard which could be named 'Show me similar tools at the MP'. Currently it is possible to go through all of the tools at the Switchboard and compare them manually, but with this invocation of the MP other relevant resources could be presented to the user in an easy way. In this regard it is important to know that the Switchboard does not hold any user friendly metadata to the tools. The Switchboard concept relies on the upload of a dataset, identification of MIME type, and/or recommendation of tools. The documentation and description of the tools is not available within the Switchboard, but usually accessible on the tools' holders websites.

Requirements and challenges: The use case comes with requirements regarding the curation of the MP's resources. Apart from the MIME type mechanism described in use case B, manually curated metadata also seems to be feasible to lead the user to useful recommendations. This

could be done on the MP's side with a respective information in the 'relations' sections of the resource.
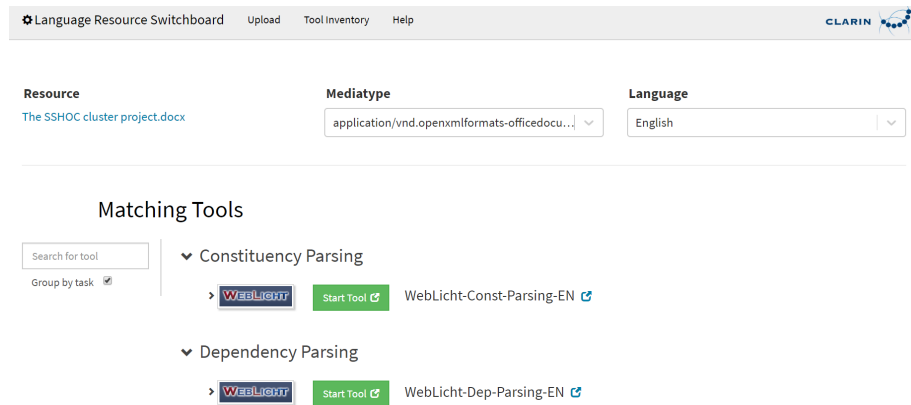

Figure 3: Using the Switchboard upload function: The function to invoke the MP from the Switchboard could be implemented as a button similar to "Start Tool" as above. The invocation button should be presented to the user after uploading a research data set and also in the tool inventory.

Both options require at least the transfer of metadata (=the search string e.g. 'show me similar tools') from the Switchboard to the MP. This requires that both services use a common vocabulary to describe resources. If a presentation of suitable tools within the environment of the Switchboard is desired, the transfer of metadata also has to be performed back to the Switchboard. However, this is not necessary for the use case to be functional.

Benefits: This scenario benefits in raising awareness for the MP. For instance, a researcher using the Switchboard with a background in linguistics may be introduced to the MP's broad range of resources which could be useful for any research discipline. The MP could be seen as a generic background information system for the Switchboard. Where the Switchboard is providing direct actionability of resources, it does not offer much background about data-types, methodologies and user tool evaluations, , this gap could be filled by the MP.

The MP may provide background information with regard to specific services or some complex recipes to further process certain data-types. This information could be made accessible via the Switchboard ie. as a 'help' function.

This use case may also serve as an opportunity to involve the SSHOC interoperability hub of WP3.

Feasibility: The effort for implementation is a technical one - the transfer of metadata between the two services - and the availability of useful metadata on the side of the MP, which is necessary to respond to the incoming request from the Switchboard.

Implementation status

IMPLEMENTED

A call-up of the MP from the Switchboard is under preparation. Taking the contextualisation of the MP records into account, this use case is not that difficult. A user coming from the Switchboard lands, for instance, on the detail page for the tool , he or she initially looked up at the Switchboard. The MP in turn shows related or similar tools and the user may inquire them[5].

E - Switchboard as Ingested Source in the MP

This use case describes the inclusion of the Switchboard tool inventory on the individual tool level in the MP. The Switchboard tool inventory is seen as an additional source for ingestion by the MP.

Scenario: In this scenario the user browses the MP for relevant resources with regard to his/her research questions. The user informs himself/herself on tools, tutorials and workflows. Here and there tools from the Switchboard will appear in the result lists, wherever a feasible relation to the search string of the user is present.

Requirements and challenges: The implementation effort connected to this scenario is overseeable. The Switchboard tool inventory[6] consists of 73 tools for language related research data (November 2020). This collection has to be ingested into the MP. It remains unclear from where the MP may retrieve the describing metadata for the tools. As said before, the Switchboard concept is threefold:

1. Users uploads a dataset,
2. Switchboard identifies the MIME type,
3. Switchboard suggests tools corresponding with the MIME type.

Any documentation or further metadata on the tools are not delivered by the Switchboard directly, but directs the user to the websites of the tools' owners.

---

[5] Important note on constraint: The implementation note describes a backlink from the Switchboard tool list to tool descriptions in the MP. Currently the MP tool descriptions are often coming from the Switchboard metadata as described in the next use-case.
[6] See https://beta-switchboard.clarin.eu/tools

Benefits: After the ingestion of this collection - which is planned anyway - the Switchboard tool inventory is visible on the tool level in the MP and also as a collection on the source level. It will also be possible to keep the collection character of the tools. The possibility to browse through the whole Switchboard tool inventory collection at the MP would most likely improve the user experience, while currently the Switchboard comes without any further documentation or tutorial regarding the tools (which both are usually available at the tools homepage).

Feasibility: As described above, the technical effort to include the Switchboard tool inventory in the MP is overseeable. The challenge lies in the fact that the tools need descriptions - which is currently not delivered by the Switchboard directly.

Implementation status

> IMPLEMENTED
>
> The Switchboard's tool inventory is already included as source in the MP. However the wish remains to enhance the tool descriptions on side of the Switchboard. As the Switchboard is an authoritative source for the MP, any enhancements (better illustrations, more description) would become visible in the MP.

F - MP desktop tool info for the Switchboard

This use case revolves around desktop based tools, which is a common occurrence in the humanities, but which fall out of the scope of the Switchboard as it is aimed at making resources actionable using readily available web services for processing, analysing or visualizing research data. Currently the Switchboard only provides access to web services and allows the user to invoke these. Desktop based tools are not included since they cannot be invoked from a web scenario.

Scenario: This use case is inspired by earlier discussions on the possibility that the MP provides information (sheets) on concepts and entities that are encountered by a user working with the Switchboard. Until now the best example of such an entity was the data type of a data resource, in which case the user could invoke a link to an information page in the MP. For instance: Explaining the TEI format. In discussions with the German CLARIAH project, it became clear that in that project it is a requirement that the Switchboard would also provide information on available desktop installed tools. Even if the user cannot invoke a desktop tool via the Switchboard, it is considered already useful that the Switchboard provides information on which desktop tool would be applicable and where to find installation information.

Requirements and challenges: Including desktop installed tools in the MP is possible and a common appearance. For the Switchboard the case is different as its "making resources actionable" usage concept relies on using web services: The user uploads a research data set and receives a list of recommended tools, which are instantly available to him or her by clicking through. Although this is a very convenient approach for the user it is technically not necessary to leave out desktop installed tools if the aim is to only provide information about applicable services. Viewed from this end, the inclusion of desktop installed tools in the Switchboard may be a valuable contribution for the user.

Benefits: At least two goals are met by this use case: Firstly, the Switchboard widens its current approach to include information about applicable desktop installed tools as well. Secondly, the user has access to a broader range of tools and might receive more recommendations. This scenario seems very reasonable with regard to the DARIAH offered range of tools - some of them falling into the desktop installed category, but clearly with potential use for users of the Switchboard.

Feasibility: Currently the Switchboard does not offer any detailed information on the listed tools in the inventory. As long as the listed tools are web services coming with proper documentation on the tool site itself, this is not a problem (although maybe a problem with regard to usability). With desktop tools the question of standardised accompanying metadata arises. Most of these desktop installed tools come with a homepage anyway, but this scenario may offer a good opportunity to rethink the Switchboard approach to go without standardised metadata.

First it was considered that such desktop tools info would be served from a Switchboard environment, however it would seem beneficial to both Switchboard and MP if such information can be provided by the MP and curated in MP and have a clear separation of responsibilities wrt such information. It would also align SSHOC with CLARIAH-DE plans.

Implementation status

SOLVED

In the meantime the Switchboard's tool inventory includes desktop tools. One example is provided below. For this reason a dedicated implementation of this use case is not necessary anymore.

Figure 4: The DARIAH-DE TopicsExplorer in the tool inventory of the BETA Switchboard. The TopicsExplorer is a desktop tool, a tool category not considered by the Switchboard's tool inventory in the past.

SSHOC

🌐 **www.sshopencloud.eu**

🐦 **@SSHOpenCloud**

in **in/company/sshoc**

✉ **info@sshopencloud.eu**