

# CLARIN

Short description	CLARIN
Type of community	Thematic Services
Community contact	Dieter Van Uytvanck, Willem Elbers
Interviewer	Sara Ramezani
Date of interview	2018. 5. 17
Meetings	
Supporters	

## User stories



### Instruction

Requirements are based on a user story, which is an informal, natural language description of one or more features of a software system. User stories are often written from the perspective of an end user or user of a system. Depending on the community, user stories may be written by various stakeholders including clients, users, managers or development team members. They facilitate sensemaking and communication, that is, they help software teams organize their understanding of the system and its context. Please do not confuse user story with system requirements. A user story is an informal description of a feature; a requirement is a formal description of need (See section later).

User stories may follow one of several formats or templates. The most common would be:

"As a <role>, I want <capability> so that <receive benefit>"

"In order to <receive benefit> as a <role>, I want <goal/desire>"

"As <persona>, I want <what?> so that <why?>" where a persona is a fictional stakeholder (e.g. user). A persona may include a name, picture; characteristics, behaviours, attitudes, and a goal which the product should help them achieve.

Example:

"As provider of the Climate gateway I want to empower researchers from academia to interact with datasets stored in the Climate Catalogue, and bring their own applications to analyse this data on remote cloud servers offered via EGI."

Note with respect to roles:

- 'Researcher' is a non-specific researcher
- 'CLARIN user' is a user working in the CLARIN infrastructure environment. Typical CLARIN users are humanities researchers, most of them are linguists or historians.
- 'CLARIN data' is data specific for the CLARIN community and stored and available from CLARIN centers
- 'Community manager' is responsible for some part of community services e.g. a repository or a community specific service contributed to EOSC

No.	User stories
US1	A researcher wants to find relevant resources by the available metadata, using keywords or other search dimensions (facets) such as date, location, language, format, etc. to use in their work. Many of such resources are available through the CLARIN infrastructure.
US2	A researcher wants to be able to find (software) tools that can be used to process the data that they have found. For instance, they want to find a tokenizer for the Dutch language.
US3	A repository manager wants to make a repository and its resources findable for researchers. There may be various forms of resources which may have anywhere from no metadata to well-defined elaborate metadata based on specific schema.
US4	A community manager wants to make some language technology tools findable for researchers. The tools have minimal metadata.
US5	A user wants to be able to discover & access the content associated with a (virtual) collection. The collection can be discovered via a search engine or other means.
US6	A researcher wants to manage a group of resources (not limited to a single existing collection or site) that are relevant for her in a way that they are easily findable, accessible, and citable.
US7	A community manager wants to group related resources from their repository in citable collections.

<b>US8</b>	A researcher wants to know what tools can be used to process a given resource. The resource could have been found through an EOSC compatible repository or discovery service or it may have been produced by the researcher. The resource itself may also be a virtual collection. The researcher would like to have an overview quickly showing a selection of tools that are relevant and useful.
<b>US9</b>	A researcher or software engineer has developed a tool for processing resources. They want to make this tool available, findable, and accessible to as many researchers and users as possible. They prefer if they can make the tool available and maintain it themselves without having to ask help from a middle layer
<b>US10</b>	A user of an EOSC-hub compatible data discovery tool wants to be able to find and access virtual collections from within the service they are using.
<b>US11</b>	A linguist using one of the EOSC-hub compatible discovery or repository services wants to be able to see what linguistic tools they can use to process a given data object, without leaving the environment of the service that they are using.
<b>US12</b>	A user of the VLO (CLARIN or other instance) wants to create and exploit semantic annotation added to VLO metadata. Semantic annotations are meaningful texts that describe or comment on the tagged resource.
<b>US13</b>	CLARIN centers need to backup & archive their data for persistency.

## Use cases



### Instruction

A use case is a list of actions or event steps typically defining the interactions between a role (known in the Unified Modeling Language as an actor) and a system to achieve a goal.

Include in this section any diagrams that could facilitate the understanding of the use cases and their relationships.

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
<b>UC1</b>	<ul style="list-style-type: none"> <li>The VLO is a search tool that can be used to find relevant data by searching through metadata</li> <li>Searching can be done various facets, e.g. date, location, genre, collection</li> <li>The primary users are researchers working in a specific domain for which the set of facets is optimised.</li> <li>Different instances for specific domains can be provided</li> <li>Community data managers can use the VLO to make their data available to a larger audience</li> <li>B2FIND is a general metadata catalogue for non-specific researchers an easier to use, but less specific tool to find resources</li> <li>B2FIND to be used for CLARIN collection level metadata and not for individual resource metadata</li> </ul>	VLO, B2FIND (US1-4)
<b>UC2</b>	<ul style="list-style-type: none"> <li>The VCR is a repository of collection metadata</li> <li>In the VCR virtual collections are defined that contain links to resources stored in other repositories</li> <li>Researchers can use the VCR to group data into virtual collections of their choosing and make these collections findable and citable</li> <li>Researchers can seamlessly access virtual collection content</li> </ul>	VCR, B2ACCESS (US5-7)
<b>UC3</b>	<ul style="list-style-type: none"> <li>The LR switchboard can provide a set of relevant tools for specific data types</li> <li>It is not meant as a stand-alone service, but as a feature that can be integrated with other services.</li> <li>Users can use this feature to find relevant tools for processing data objects directly from within the data discovery or hosting service.</li> <li>Tool and service providers can make their tools available via the LR switchboard by specifying its features (tool metadata).</li> </ul>	LR Switchboard (US8-9)

<b>UC4</b>	Reverse integration of VCR with B2SHARE (B2DROP/B2STAGE?).  Users can access and use the VCR from within relevant EOSC services to download or copy the data to a new destination.	VCR,  B2SHARE, B2DROP  (US10)
<b>UC5</b>	Reverse integration of LR Switchboard with B2FIND/B2SHARE/B2DROP  The LR Switchboard can be accessed by users from within relevant EOSC services to show what tools can be applied on a given resource	LR Switchboard,  B2SHARE, B2DROP  (US11)
<b>UC6</b>	Integration of B2NOTE with the VLO.  B2NOTE can be used from the VLO as an added discovery mechanism using semantic annotations in addition to metadata. Users can add semantic annotations and search for them.	B2NOTE, VLO  (US12)
<b>UC7</b>	B2SAFE/B2STAGE use for safe data replication. This already occurs, but CLARIN needs clear costs and SLA statements. The B2SAFE HTTP API is used for connecting Space with B2SAFE.	B2STAGE, B2SAFE, B2SAFE HTTP API  (US13)

## Requirements

### Technical Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case	Related ticket
Example	EOSC-hub AAI	Yes: EOSC-hub AAI doesn't support the Marine IdP	EOSC-hub AAI should accept Marine IDs	UC1	
<b>RQ1</b>	<b>VLO, B2FIND</b>	Yes, not all CLARIN metadata is harvested (scaling and mapping problem on the B2FIND side)	Make VLO metadata also findable in B2FIND. At least the available collection metadata should be harvested. B2FIND to harvest CLARIN collection level metadata only. This is an adapted technical req (orig. harvest collection metadata through a managed VLO OAI-endpoint.)	UC1	<div>  </div> <div> <a href="#">EO SC WP 10-48 - Jira</a> </div>

RQ2	VCR, B2ACCESS	Yes.	AAI integration between CLARIN and EOSC-hub. different levels of integration are conceivable: (a) CLARIN thematic services part of EOSChub ID federation (b) CLARIN users able to access EOSChub services. CLARIN can indicate which extra users (IdPs) will be enabled using EOSC services in this way.  Note that this (SAML federation interoperability) is not a technological but a legal & administrative problem.	UC2	
RQ3	VCR, B2FIND, B2SHARE	Yes.  Link from VCR to EOSC-hub services	Make metadata resources available in relevant EOSC-hub services, e. g. B2SHARE, B2FIND, etc., accessible from the VCR and in suitable format for building virtual collections	UC2	 EO SC WP 10- 17 - Jira
RQ4	VCR, B2SHARE, B2FIND, EGI DataHub, ...	Yes.  Link from EOSC-hub services to VCR	Make the VCR available when using EOSC-hub repository and discovery services, such that this data can be downloaded e.g. used to make new collections	UC4	 EO SC WP 10- 17 - Jira
RQ5	LR Switchboard , B2DROP, B2SHARE, B2FIND, ...	Yes.  Invoking the LR Switchboard from within EOSC-hub services	Adding a capability to EOSC-hub services to invoke the LR switchboard for a data item. This requires information about the data items to be sent to the LR switchboard via a call to its API	UC3, UC5	 EOSCWP10-16 - Jira  EOSCWP10-78 - Jira
RQ6	B2NOTE	Yes  B2NOTE not integrated with CLARIN VLO nor connected to CLARIN SPF	Integrate B2NOTE in the VLO, add B2NOTE to CLARIN SPF (note any legal requirements)	UC6	 EOSCWP10-14 - Jira

RQ7	B2SAFE /B2STAGE	Yes  No single costs and condition listing available	Provide information on all relevant aspects including costs and conditions	UC7	
-----	--------------------	--	---	-----	--

Capacity Requirements

EOSC-hub services	Amount of requested resources	Time period
hosting	hosting services for VLO, VCR and LR-Switchboard	indefinite if suitable SLA and costs