

EOSC-hub task 7.1 roadmap

Authors: Dieter van Uytvanck, Twan Goosen, Willem Elbers

Date: 31st of January 2018

Document ID: CE-2018-1175

Table of contents

Table of contents	1
Introduction	2
Software Management Processes	2
Security Policies	2
Training and outreach	3
User Satisfaction	3
Uptime and service level agreements	3
Thematic Services	4
Subtask 1: Integration of the Virtual Language Observatory into EOSC-hub	4
Responsibilities	4
Development	4
EOSC-hub deployment hardware requirements	5
Subtask Milestones	5
Metrics	5
Training and outreach	6
Subtask 2: Integration of the Virtual Collection Registry into EOSC-hub	6
Responsibilities	6
Development	6
EOSC-hub deployment hardware requirements	7
Subtask Milestones	7
Metrics	7
Training and outreach	7
Subtask 3: Integration of the Language Resource Switchboard into EOSC-hub	8
Responsibilities	8
Development	8
EOSC-hub deployment hardware requirements	8
Subtask Milestone	8
Metrics	9
Training and outreach	9
Summary of subtasks	10

Introduction

This document will provide the work plan for integration of the CLARIN thematic services into the EOSC-hub. This is the initial version of the work plan (P1), aiming at providing a detailed plan for the first year of the project. A revision of this document will be provided in M12 (P2) and M24 (P3), each defining the following project year in more detail.

The document will start with a number of sections describing best practices, processes and policies used in the CLARIN infrastructure, applicable to all of the thematic services. The remainder of the document will describe the work plan for each of the thematic services in more detail, especially for the first project year.

Software Management Processes

For each of the thematic services described in this work plan, the same software management process is followed. All code is managed under the CLARIN-ERIC GitHub organization¹. Releases are developed based on a predefined set of milestones.

When all milestones for a release are completed an **alpha version** is released and deployed. After testing this release, any issues should get resolved before making and deploying a beta release. In the alpha phase it is still allowed to change the feature set.

Afterwards, there is a **beta testing** phase and time to resolve any issues before making a **production release**. During the beta stage feature changes are not allowed anymore.

If any serious issues are found in a production release, patches will be made available as soon as possible and deployed in production.

Within the CLARIN infrastructure it is considered best practice to follow a Docker-based deployment workflow². This applies to all thematic services described in this work plan, however due to technical restrictions the production VLO service is currently not deployed via the docker workflow.

Security Policies

In line with EUDAT, CLARIN ERIC is following the CSIRT³ security policies. Main responsible persons for taking up security incidents are the CLARIN sysop team, reachable at sypsops@clarin.eu. More information is available in CLARIN-PLUS deliverable 4.1: Report on risk management for e-Infrastructures⁴.

¹ <https://github.com/clarin-eric>

² <https://gitlab.com/CLARIN-ERIC/build-script>

³ <http://www.csirt.org/>

⁴ https://office.clarin.eu/v/CE-2016-0743-CLARINPLUS-D4_1.pdf

Training and outreach

CLARIN organises an annual conference, which is the main annual event for those working on the construction and operation of CLARIN across Europe, as well as for representatives of the target research communities in the humanities and social sciences. This conference is targeting a wide audience and provides the ideal platform to report on the involvement of CLARIN in the EOSC-hub project.

In addition to the annual conference there is also an annual centre meeting, which is aimed at a more technical audience, mainly the operators from the different CLARIN centres. These meetings can be used to provide technical training and information to the CLARIN centres on the integration of the CLARIN thematic services into EOSC-hub.

User Satisfaction

For all thematic services, user satisfaction is considered a very useful metric. Currently this is not measured but we envision an approach where a small portion of the user's will be asked to fill in a short questionnaire. Participation is on voluntary basis and the questionnaire should be kept as short and concise as possible. As a first approach just asking the user to rate the service could already yield interesting results. This process should be automated by making use of existing libraries.

Uptime and service level agreements

CLARIN ERIC is providing its central services on a best effort basis with a minimum target uptime of 97.5%. In practice it has been able to provide a 99.95% uptime in 2017 (downtime includes pre-announced maintenance windows). With respect to support it is intended to take up any incoming request⁵ within two business days.

Part of our services are running with academic providers:

- MPCDF (no formal SLA, best effort)
- Cessnet (no formal SLA, best effort)

Others are running at CLARIN centres:

- Virtual Collection Registry at the Institute for the German Language (no formal SLA)
- Language Resource Switchboard at the University of Tübingen (no formal SLA)

And some of the services are running on commercial infrastructure:

- TransIP⁶: 99.99% server uptime
- HostEurope⁷: 99.95% server uptime

⁵ <https://www.clarin.eu/content/support>

⁶ <https://www.transip.eu/legal-and-security/availability/>

⁷ https://www.hosteurope.de/download/Host_Europe_SLA_2016-07-12_3.1_english.pdf

Thematic Services

Subtask 1: Integration of the Virtual Language Observatory into EOSC-hub

The Virtual Language Observatory (VLO) is a metadata-based portal for language resources. It was developed within CLARIN as a means to explore linguistic resources, services and tools available within CLARIN and related communities, aims to provide an easy to use interface, allowing for a uniform search and discovery process for a large number of resources from a wide variety of domains and providers and is completely based on the Component Metadata (CMDI) standard⁸ and semantic mapping through the CLARIN Concept Registry⁹. This approach allows for the use of heterogeneous metadata schemas while maintaining semantic compatibility. Metadata is harvested from multiple repositories and is based on OAI-PMH¹⁰. Where needed a conversion is applied from the input format (e.g. OLAC, IMDI and the CLARIN LRT inventory) to their CMDI counterpart profiles. The VLO search interface presents a number of facets, for each of which one or more values can be selected in order to narrow down the selection of displayed records. More details are available in the paper: Semantic metadata mapping in practice: The Virtual Language Observatory¹¹.

Responsibilities

Name	Contact	Role
Twan Goosen	twan@clarin.eu	Lead developer
Thomas Eckart	teckart@informatik.uni-leipzig.de	Developer

Development

New EOSC-hub-specific features in VLO 4.5 include:

- Quality/sustainability:
 - Fully containerised production ready setup
 - Solr testing ([#125](#)) and performance
 - Solr authorisation ([#126](#))
- Functionality:
 - Shareable URLs ([#122](#))
 - Facet flexibility/extensibility (e.g. [#36](#))
 - Cross facet mapping
- Documentation

⁸ <https://www.clarin.eu/cmdi>

⁹ <https://www.clarin.eu/ccr>

¹⁰ <https://www.openarchives.org/pmh/>

¹¹ <http://hdl.handle.net/11858/00-001M-0000-000F-85EE-8>

- Guided tour ([#7](#))

M24: Will be further defined in the second iteration of this work plan, scheduled for project month 12

- More advanced VCR integration ([#80](#))
- Features for other communities and their data M30: Will be further defined in the third iteration of this work plan, scheduled for project month 24

M30: A final release is planned to take into account any feedback.

EOSC-hub deployment hardware requirements

Single VPS with:

- 8 vCPU cores
- 8 GB memory
- 1 TB storage, high IOPS (preferably SSD)

Subtask Milestones

Subtask	Month	Description	Inputs	Outputs
VLO.1	M6	Release of VLO 4.5	-	* Tagged GitHub release
VLO..2	M6	EOSC-hub deployment	* [CLARIN] Hardware requirements	* EOSC-hub deployment
VLO.3	M13	Integration with EOSC core services	* [EOSC] Definitions of services to integrate (aai, monitoring, ...) * [EOSC] Integration guidelines * [EOSC] Service endpoints to integrate with	* Service integrated with core EOSC services
VLO.4	M17	Integration of accounting and reporting	* [EOSC] Integration guidelines * [EOSC] Service endpoints to integrate with	* Service integrated with accounting and reporting services
VLO.5	M24	Separate VLO instance for e-Infra user communities	* [EOSC, CLARIN] Define communities * [EOSC] Provide resources	* e-Infra user community deployments
VLO.6	M30	Final release	* [CLARIN] Requirements	* Tagged GitHub release * EOSC-hub deployment

Metrics

- Piwik statistics (number of visits, geographical distribution, ...)
- Number of harvested metadata records
- Number of accessible resources
- User satisfaction (See section on User Satisfaction)

Training and outreach

- A guided tour providing an introduction to the VLO is planned for the VLO 4.5 release.
- An instructional video / webinar will be provided.

Subtask 2: Integration of the Virtual Collection Registry into EOSC-hub

A virtual collection is a coherent set of links and/or resolvable identifiers to digital objects (e.g. annotated text, video) that can be easily created, accessed and cited. The links can originate from different archives, hence the term virtual. A virtual collection is suitable for manual access (using a web-browser) as well as automated processing (e.g. by a webservice).

CLARIN provides a registry where scholars can create and publish their virtual collections. It is closely integrated with the infrastructure and provides persistent identifiers and federated login. The collection metadata is openly available and accessible via the Virtual Language Observatory.

Responsibilities

Name	Contact	Role
Willem Elbers	willem@clarin.eu	Lead developer
Twan Goosen	twan@clarin.eu	Developer

Development

The development on the Virtual Collection Registry will focus on the milestones defined for EOSC-hub. In M6 a new release is planned to prepare the deployment within EOSC-hub, in M21 a new release is planned to support the integration with B2SHARE and in M27 a release is planned to further integrate with B2STAGE. In M30 a final release is planned to accommodate any final feedback.

M6 release:

- Finish 1.2 release¹² by the end of M6, this includes:
- Improving integration with other tools such as the vlo ([#1](#) and [#80](#))
- Improving integration with LRS

The M21 release will be further defined in the second iteration of this work plan, scheduled for project month 12 and the M27 release will be further defined in the third iteration of this work plan, scheduled for project month 24..

¹² <https://github.com/clarin-eric/VirtualCollectionRegistry/milestone/4>

EOSC-hub deployment hardware requirements

Single VPS with:

- 2 vCPU cores
- 8 GB memory
- 50 GB storage

Subtask Milestones

Subtask	Month	Description	Inputs	Outputs
VCR.1	M6	1.2 release	-	* Tagged GitHub release
VCR.2	M12	[EOSC-hub] MC1 - Deployment	* [CLARIN] Hardware requirements	* EOSC-hub deployment
VCR.3	M13	Integration with EOSC core services	* [EOSC] Definitions of services to integrate (aai, monitoring, ...) * [EOSC] Integration guidelines * [EOSC] Service endpoints to integrate with	* Service integrated with core EOSC services
VCR.4	M17	Integration of accounting and reporting	* [EOSC] Integration guidelines * [EOSC] Service endpoints to integrate with	* Service integrated with accounting and reporting services
VCR.5	M21	[EOSC-hub] MC2 - B2SHARE integration	* [EOSC, CLARIN] Requirements	* Tagged GitHub release * EOSC-hub deployment
VCR.6	M27	<i>[EOSC-hub] MC3 - B2SHARE (and B2STAGE)</i>	* <i>[EOSC, CLARIN] Requirements</i>	* <i>Tagged GitHub release</i> * <i>EOSC-hub deployment</i>
VCR.7	M30	Final release	* [CLARIN] Requirements	* Tagged GitHub release * EOSC-hub deployment

Note: VCR.6 requires further investigation to decide on a meaningful way of integrating with B2STAGE. There is a risk that it turn out no such meaningful integration exists. This will be further specified in the third iteration of this work plan, scheduled in M24.

Metrics

- Piwik statistics (number of visits, geographical distribution, ...)
- Number of published virtual collections
- Number of linked resources
- User satisfaction (See section on User Satisfaction)

Training and outreach

No specific training events planned at this time This will be revised in M12.

Subtask 3: Integration of the Language Resource Switchboard into EOSC-hub

The Language Resource Switchboard (LRS) has been developed within the CLARIN-PLUS project as a means to link linguistic resources with the tools that can process them. Tool developers, please contact us in case you want your tool registered with the switchboard! More information available in the paper: The CLARIN Language Resource Switchboard¹³.

Responsibilities

Name	Contact	Role
Claus Zinn	claus.zinn@uni-tuebingen.de	Lead developer
Twan Goosen	twan@clarin.eu	Developer
André Moreira	andre@clarin.eu	Developer

Development

Features to be included at launch of version for EOSC-hub (M14):

- Fully containerised production ready setup

M30: Will be further defined in the third iteration of this work plan, scheduled for project month 24

EOSC-hub deployment hardware requirements

Single VPS with:

- 2 vCPU cores
- 8 GB memory
- 50 GB storage

Subtask Milestone

Subtask	Month	Description	Inputs	Outputs
LRS.1	M14	First release for EOSC-hub	* [CLARIN] Hardware requirements * [EOSC] Definitions of services to integrate (aai, monitoring, ...) * [EOSC] Integration guidelines	* Tagged GitHub release * EOSC-hub deployment * Service integrated with core EOSC services * Widened the scope of the LR Switchboard ¹⁴

¹³ https://www.clarin.eu/sites/default/files/zinn-CLARIN2016_paper_26.pdf

¹⁴ Widening the scope of the LR Switchboard to data processing web applications outside the language resource domain (available via EUDAT, EGI and Indigo Data Cloud).

			* [EOSC] Service endpoints to integrate with	
LRS.2	M15	Integration of B2DROP with the LR Switchboard	* [EOSC] B2DROP instance	* B2DROP integrated with LR switchboard
LRS.3	M17	Integration of accounting and reporting	* [EOSC] Integration guidelines * [EOSC] Service endpoints to integrate with	* Service integrated with accounting and reporting services
LRS.4	M18	Integration of B2SHARE with LR Switchboard.	* [EOSC] B2SHARE instance	* B2SHARE integrated with LR Switchboard
LRS.5	M30	Final release	* [CLARIN] Requirements	* Tagged GitHub release * EOSC-hub deployment

Metrics

- Piwik statistics (number of visits, geographical distribution, ...)
- Number of available processing tools
- (Optional) User satisfaction¹⁵ (See section on User Satisfaction)

Training and outreach

No specific training events planned at this time. This will be revised in M12.

¹⁵ For the LRS user satisfaction could be difficult to measure in a meaningful way, since for the end user the whole workflow will define the experience. The switchboard is just a small step in this workflow.

Summary of subtasks

Figure 1: Overview of the planned subtasks for each of the thematic services.

Project Month	2018												2019												2020											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Workplan	P1												P2												P3											
VLO	[Shaded]																																			
VLO.1	[Shaded]																																			
VLO.2	[Shaded]																																			
VLO.3	[Shaded]																																			
VLO.4	[Shaded]																																			
VLO.5	[Shaded]																																			
VLO.6	[Shaded]																																			
VCR	[Shaded]																																			
VCR.1	[Shaded]																																			
VCR.2	[Shaded]																																			
VCR.3	[Shaded]																																			
VCR.4	[Shaded]																																			
VCR.5	[Shaded]																																			
VCR.6	[Shaded]																																			
VCR.7	[Shaded]																																			
LRS	[Shaded]																																			
LRS.1	[Shaded]																																			
LRS.2	[Shaded]																																			
LRS.3	[Shaded]																																			
LRS.4	[Shaded]																																			
LRS.5	[Shaded]																																			