

Title Minutes SCCTC September 2016
Version 1
Author(s) LS
Date 2016-10-26
Status Final
Distribution Centre Committee
ID CE-2016-0854



Participants

Davor Ostojic (AT), Pavel Straňák (CZ), Dirk Goldhahn (DE), Krista Liin (EE), Martin Mathiessen (FI), Andrius Utkas (LT), Marc Kemps-Snijders (NL), Marcin Pol (PL), Dana Dannélls (SE), Cyprian (SI), Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes)

Excused: Kiril Simov (BG), Lene Offersgaard (DK), Xaris Papageorgiou (GR), Freddy Wetjen (NO), Tomaž Erjavec (SI)

0 Action points

Action	By whom	By when
Listing FCS 2.0 compliance end point implementations: https://trac.clarin.eu/wiki/Taskforces/FCS/Endpoint_s#ExperimentalFCS2.0compliantendpoints	DvU	Done
Following up on citation issues put on agenda for Aix with background document	DvU	In Aix 26 October
Prepare short summary of PID situation for Aix put on agenda for Aix	Pavel	In Aix 26 October

1 Agenda

The following agenda was accepted:

1. Agenda
2. Approval minutes last meeting & action point status ([CE-2016-0821](#))
3. Status report from the assessment committee (Lene)
 1. A-service assessment pilot ([CE-2016-0850](#))
4. Request from NCF: report about the technical perspective on interoperability (introduction by Dieter, background: [CE-2016-0845](#), [mimetype inventory](#), TEI mimetype discussion [part 1](#) and [part 2](#))
5. Reports from the task forces: (task force coordinators) -> see [reporting Google Doc](#) (will be updated)
 1. [CMDI](#)
[Metadata curation](#)
[Federated Content Search](#)
 2. [Persistent Identifiers](#)
 3. [AAI](#)
6. Status update per country/member (all)
7. Any other business

2 Approval minutes and action points

The minutes of the June 2016 SCCTC meeting have been approved.

3 Status of the A-service assessment pilot

DVU:

- Willem Elbers (CLARIN ERIC) has been able to replicate and Dockerize the LINDAT Dspace repository so it is even more convenient to set it up at another server.

This operation took 4 to 8 hours, this is something to take in mind when thinking of testing the replicability of A-services. It comes with a cost when setting this up, but it is a worthwhile investment.

- Concluding from the assessment pilot and the assessment committee report the A-service status is granted to the LINDAT Dspace repository since all requirements have been met.

During Aix we will ask the people who did the evaluation to report on any loose ends or improvements for the general service.

- The SCCTC has received the report from the assessment committee, there are no objections on approving the A-service status for LINDAT.

The A-service status for LINDAT is hereby approved.

- The B-centre assessment for FIN-CLARIN has been outstanding and overdue because of the summer holidays. The B-centre status is achieved and approved. FIN-CLARIN will receive the necessary official material (logo, certificate, news item) this week.

4 Request from NCF: report about the technical perspective on interoperability

With respect to the Interoperability discussion ([CE-2016-0845](#) of Jan Odijk), there is a need for more clarification on this both from the side of the NCF as well as on the side of the Standards Committee.

It is good timing to discuss what interoperability is and it is a non-trivial exercise to match language resources with tools to process the language resources, for the purpose of making analyses and/or visualize the outcome.

There is an interest for these kind of analyses from 3 different perspectives:

1. NCF: They would like us to come up with overview of the current situation, largely data based, and really looking at the tools and resources that are available through CLARIN today
2. Europeana: has a similar request about the usability and the possibility of using CLARIN tools to analyze the resources which they are providing. CLARIN ERIC is participating in the Europeana project and the aim of our task is to connect the CLARIN tools to the Europeana datasets. We need it for internal purposes as well.

3. RDA: The RDA working group on data fabric is also about matching data and tools and how the data are exactly described.

Explicit input has been asked from our side.

TEI mimetype discussion

At the same time a discussion/ exchange of idea on how to type TEI documents has started. Since TEI is a type of skeleton but does not explicitly define its subtypes (it can, but it is often not the case) there has been a discussion by some of the people who are working around processing TEI in weblicht about how to type and how to make that match between TEI files and tools that can process that.

It is worthwhile to read the discussion thread in the email. (discussion [part 1](#) and [part 2](#))

To bring the discussion further along DVU started a Mime Type Inventory list. The idea is to analyse in how far for the resources for which we have metadata descriptions we can use the mime to have an exact specification of the datatype of the Language Resources.

It will be important to come up with an up-to-date description of how we are dealing to describe those data types within CLARIN and how we see future steps with regards to matching files and tools which are provided to CLARIN to process those types.

DVU wants to write up a document indicating the context for this discussion and try to give an overview of the kind of frequently occurring file types and data types within CLARIN and possibilities of describing that.

This is a task for the SCCTC. DVU wants to set up a core-author team for writing the document. This document will eventually be reviewed by the centre committee.

The following members would like to participate:

- Pavel Straňák (CZ)
- Martin Mathiessen or one of his colleagues (FI)
- Marc Kemps-Snijders (NL), can contribute from CLARIAH and can probably provide input, but will have to discuss internally who to assign to this task.
- Davor Ostojic (AT)

DVU will coordinate the writing of the document.

5 Reports from the task forces

Short status update from other taskforces:

- **CMDI:**
Roll-out of CMDI 1.2 was successful.
- **Metadata curation:**
Curation module service changes:
 - Normalisation
 - Aligned with CMDI 1.2
- **Federated Content Search:**
Test version of the FCS 2.0 aggregator should be ready this week <https://spraakbanken.gu.se/ws/fcs/2.0/aggregator/?queryType=fcs>, together with the first endpoint (using Korp, see below). The new aggregator will still be backwards compatible with the older endpoints.

The following FCS reference implementation libraries are available:

- FCS Simple Endpoint (Version 1.3.0)
 - Source code: <https://trac.clarin.eu/browser/FCSSimpleEndpoint>)
 - Binaries available from CLARIN Maven repository.
- FCS Annotation Translator(for use in endpoints)
 - Source code: <https://trac.clarin.eu/browser/FCSannotrans>)
 - Not yet released to CLARIN Maven repository.
- FCS Simple Client (Version 1.3.0)
 - Source code: <https://trac.clarin.eu/browser/FCSSimpleClient>
 - Binaries available from CLARIN Maven repository.
- Support libraries, i.e. software libraries that are used by the aforementioned components:
 - FCS QL
 - Source code: <https://trac.clarin.eu/browser/FCS-QL>
 - Binaries available from CLARIN Maven repository.
 - SRU Server (Version 1.8.0)
 - Source code: <https://trac.clarin.eu/browser/SRUserver>
 - Binaries available from CLARIN Maven repository.
 - SRU Client (Version 1.8.0)
 - Source code: <https://trac.clarin.eu/browser/SRUclient>
 - Binaries available from CLARIN Maven repository.

A simple demo implementation with Lucene search engine can be found at <http://clarin.ids-mannheim.de/downloads/clarin/DigiBibSRU-source-2016-02-08.zip>, however this project does not use advanced FCS (e.g. searching on POS).

A korp reference implementation using the FCS 2.0 advanced features is available both online since June at Språkbanken <https://spraakbanken.gu.se/ws/fcs/2.0/endpoint/korp/sru> and now the source is moved from internal subversion into the CLARIN ERIC github <https://github.com/clarin-eric/fcs-korp-endpoint>. It is also the base implementation for examples in the Clarin FCS 2.0 endpoint developers' tutorial [link] which will continue to be promoted during the autumn together with support efforts to get further centres to add/move to fcs 2.0 endpoints.

The work on FCS 2.0 was funded by CLARIN-PLUS.

- **Persistent Identifiers:**
All earlier reported issues have been addressed.
- **AAI:**
No meeting held. [Lindat's Aaggreg](#) has now 10 participating parties, combining it with [Geant's Shibboleth attribute checker](#) is done.

6 Status report of the countries

Austria

- Organisation of the CLARIN-PLUS [Workshop on Tools & Services](#) (2-3 June 2016)
- Substantial presence at DH Krakow
- Preparation of the [TEI conference](#) (26 - 30 September)
- Preparation of a number of DH project proposals (mainly as technological partner) for a national call for digitisation initiatives [go!digital 2016](#) (deadline: 15 July, volume: 2 Mio. EUR, ~ 15 projects to be awarded)

- After evaluation phase, started reimplementation of the repository ([CCV](#)) based on Fedora 4 (as Fedora 3 reached End of Life by the end of 2015) Aiming for going to production (and CLARIN-B + DSA re-assesment) beginning 2017
- Continuous improvements on the [Metadata Curation service](#) especially support for CMDI 1.2
- Testing MTAS (corpus search engine developed by @Meertens) on our big newspaper corpus (amc - almost reaching 10 bio tokens)
- Heavy involvement in PARTHENOS project, especially
 - Explaining CMDI to the rest of the Parthenos world (again & again - it is really difficult to make it comprehensible to outsiders) E.g. at the F2F meeting of WP5/6 in Heraklion 6. - 8. June)
 - Working on the mapping from CMDI to the common model "Parthenos entities" (an extension of CIDOC-CRM)
 - Preparing/testing the integration of CLARIN services into the envisaged provisioning infrastructure ([gCube](#) / [D4Science](#))

Bulgaria

No report

Czech Republic

- Providing support to Clarin-Lithuania, Italy and Spain to setup DSpace repository and AAI
- Finishing UI issues in CLARIN Plus
- Preparing for CAC 2016
- Contributed to RDA meeting on PIDs: <http://hdl.handle.net/11346/slides-7QDP>

Denmark

No report

Dutch Language Union

No report

Estonia

Slow work with improving user interfaces & integrating our resources into them, checking our SP and IdP metadata and connections.

Finland

- Done
 - ACA resources require now eduPersonAffiliation=member (was: faculty)
 - Implemented in korp.csc.fi and our download service.
- In Progress
 - Centrally managed reference instructions: A relatively complicated issue.
 - Pending: providing ACA access to eligible persons that do not have the eduPersonAffiliation attribute set.

Germany

- Next developer meeting will be held on 20.09.2016 in Mannheim
- Kick-Off of CLARIN-D "expansion" phase in November in Leipzig
- Currently: Negotiations with computing centres for ongoing cooperation
- Ongoing meetings with DARIAH-DE on deeper cooperation. Good progress on integrating DARIAH-Metadata into VLO

Greece

- A substantial expansion of our cloud infrastructure in terms of hardware

- A new Big Data text processing architecture in place
- Exploiting our infrastructure and its processing capacity in two projects in collaboration with the National Centre for Social Research and the Panteion University of Social and Political Sciences
- Preparation for B-center application according to the B-Centre checklist

Italy

No report

Latvia

No report

Lithuania

- In summer on CLARIN-LT website we have launched a search engine for the Corpus of Lithuanian Language as a new service. The search engine has new features: lemma search, context search, sorting, tag-based search etc. It is fully operational, but still in testing mode.
- Finally, we have set up a CLARIN-LT repository. After successful exchange of SAML data, we have initiated the signing of a new agreement with CLARIN SPF.

The Netherlands

- Open SKOS developments aligned with NDE, agreement to merge code branches and continue collaborative development. Version OpenSKOS 2.1 and 2.2 planned and approved as part of NDE activities. This includes a.o. SKOS-XL support. Further talks will take place on a joint governance structure for the consortium, CLARIN EU is expected to be represented in this.
- New production version of Nederlab to be released in early October. This includes MTAS (Multi Tier Annotation Search), delivered as part of the CLARIAH project. Several initiatives started to test usability with other interested parties using other formats (ELAN, TEI, 'SketchEngine like').
- LOD activities within CLARIAH are progressing and slowly moving into a phase where concrete use cases are proposed and planned for implementation

Norway

No report

Poland

- Continuing web services development
- Some web services for english language: <http://ws.clarin-pl.eu/demo2/parserEn.shtml?en>

Plans:

- Extend cluster server
- Cooperation with Czech Republic on d-space developing

Portugal

No report

Slovenia

- Mostly continuing development of merge script for TEI and WebAnno TSV formats

Sweden

- Our B-centre application has not been approved yet, we received a few questions we need to answer and some minor remarks we need to approach.

- We will have an inauguration ceremony in Gothenburg on the 7th of October. There will be several presenters from Gothenburg, Stockholm and Uppsala. We will have a demonstration session where we will present our tools such as for example name and place identification, text classification, and geographic visualisation of place names over time. More information about the ceremony will be made available on our website: <https://sweclarin.se/eng/Inauguration_of_the_Swe-Clarin_toolbox_webform>.
- The Swe-Clarin web is getting a new appearance and we will launch new functionalities during the end of this year. So far there are 36 members subscribed to our newsletter.
- Dana Dannélls will be on parental leave until Maj 2017, Caspar Jordan will be taking over during this time.

United Kingdom

No report

6 Any other business

None

Next meeting: Wednesday 26 October 2016, Aix-en-Provence