

Title Minutes SCCTC October 2016
Version 1
Author(s) LS
Date 2016-12-08
Status Final
Distribution Centre Committee
ID CE-2016-0883



Participants

Davor Ostojic (AT), Matej Durco (AT), Jan Hajic (CZ), Dirk Goldhahn (DE), Lene Offersgaard (DK), Krista Liin (EE), Martin Mathiessen (FI), Dimitrios Galanis (GR), Miltiadis Deligiannis (GR), Andrius Utka (LT), Roberts Dargis (LV), Marc Kemps-Snijders (NL), Marcin Pol (PL), João Silva (PT), Leif-Jöran Olsson (SE), Caspar Jordan (SE), Cyprian (SI), Twan Goossen (CLARIN ERIC), Menzo Windhouwer (CLARIN ERIC) Dieter Van Uytvanck (CLARIN ERIC, Chair), Linda Stokman (CLARIN ERIC, minutes)

Excused: Kiril Simov (BG), Freddy Wetjen (NO), Tomaž Erjavec (SI)

0 Action points

Action	By whom	By when
Brainstorm about the future of AAI Taskforce	All	Before the next SCCTC meeting

1 Agenda

The order of the agenda is changed. Taskforce reports (originally point 5) were discussed before Approval of minutes & action points (originally point 2) and the Assessments (originally point 3). The agenda was accepted as follows:

1. Agenda
2. Reports from the task forces: (task force coordinators) -> (task force coordinators) -> oral reports (5 to 10 minutes)
 1. [CMDI](#)
[Metadata curation](#)
[Federated Content Search](#)
 2. [Persistent Identifiers](#)
 3. [AAI](#)
3. Approval minutes last meeting & action point status ([CE-2016-0854](#))
4. Assessments:
 1. Planning assessment round 8
 2. A-service assessment: feedback from assessment committee, next round
5. Document: technical perspective on interoperability
 1. Skeleton document: the CLARIN data fabric, including input from CLARIN-D on mimetypes ([CE-2016-0882](#))
6. Status update per country/member (all)
7. Any other business

2 Reports from the task forces

Short status update from other taskforces:

CMDI:

- CMDI 1.2 was introduced (July)
 - A new CMDI toolkit was completed and released
 - Required adaptations were made to the essential infrastructure components (Component Registry and Virtual Language Observatory)
 - A detailed technical specification was completed and published
 - Documentation for various user groups was written and published on clarin.eu
 - All centers can now
 - produce CMDI 1.2 metadata
 - make components and profiles using the new CMDI 1.2 features
 - adapt their software to (also) support CMDI 1.1
- Communication and collaboration with Concept Registry and CLAVAS developers at Meertens
 - CLAVAS
 - Available in production within the next month or so and supported by upcoming Component Registry release (version 2.2.0)
 - New vocabularies can be added but submitters need to commit themselves to the maintenance of their vocabularies
- Work for the coming months:
 - Complete support for all new features in the component registry
 - Describe best practices for metadata modelling and metadata creation
 - Work together with the Metadata Curation task force to make necessary adaptations to improve metadata quality

Metadata curation:

- Continuous work on curation module
- Curation instance of VLO – updated
- (- will present some first numbers on Friday 11:35)

Plan:

- Push work on controlled vocabularies and mapping/normalisation
- Planned for 4.0.2 version of VLO
- Establish workflow (involving CLAVAS, github & VLO)
- Facets to concentrate on: resource type, license/availability, format (mime-type)

Request:

- Automate update
- Inform the admins
(both are planned to be implemented soon)

Federated Content Search:

- New version (FCS 2.0)
- New aggregator by spraakbanken
- Endpoint implementations supporting FCS2.0: CORP@gu.se, ddc@BBAW
- Graphical query builder for CQP-flavour of FCS (react-components)
- Endpoint validator on its way
- Ask for more test cases
- Logging of the queries to aggregator in piwik - gives us statistics/insight about the usage

Persistent Identifiers:

- No updates

AAI:

- Interest in the AAI Taskforce meeting has been low, so far Martin and 2 others.
- Planned topics worth mentioning are the Lindat's AAggreg and the new EU Data Protection Regulation. Aaggreg has 10 services connected and works well. However there has not been an increase in connections since the last SCCTC meeting. More promotion outside of the taskforce mailing list seems to be needed (and will be achieved later on here at the CAC). The upcoming EU General Data Protection Regulation (GDPR) concerns not only AAI, but all CLARIN services which process personal data, also those with local accounts. Martin will take up the issue with the CLARIN Legal Committee.
- The AAI Taskforce might change shape and morph from a problem identifying and solving group to a knowledge sharing pool. This will be discussed in the AAI TF meeting today.

3 Approval minutes and action points

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

4 Assessments (report by Lene Offersgaard)

1. Planning assessment round 8

The B centre assessment was somewhat delayed because it is depending on the centre getting the Data Seal of Approval (DSA). DSA is being re-organized and the new DSA application will be available end of November, this postponed the second round in 2016.

Re-assessment of 4 B centres (Vienna, INT, Meertens, Copenhagen) are also postponed, the SCCTC has accepted to extend these seals to mid 2017.

Deadline for next B centre round: Jan. 31, 2017

If you have any B centre candidates, please report this in the SCCTC meeting.

B centre test should be more automatic and should be completed in 3 months

Action point for office: Monitor the expiration dates and warn centres of expiration 6 month before this date.

2. A-service assessment: feedback from assessment committee, next round

First A-service ("LINDAT repository") was approved by the SCCTC in September 2016:

Make a more clearer document of the A-service procedure adding:

- Info about funding support
- Listing dependencies/small "risk analysis"
- A way to register an A-service will be set up

A-services must be approved by SCCTC as relevant, before an assessment can take place, this must be added to the procedure.

Next deadline for A-service assessment: Q1 of 2017, perhaps end of January

5 Document: technical perspective on interoperability

1. Skeleton document: the CLARIN data fabric, including input from CLARIN-D on mimetypes ([CE-2016-0882](#))

DvU gives a short summary on the document.

As agreed in the previous meeting a small group was formed to create this skeleton. Dieter suggest to break up in small groups at the end of this SCCTC meeting to go over the document and do some editorial work. Not to complete the document during this meeting but to get things going.

The editorial team consists of:

- Pavel Straňák from CLARIN-CZ
- Martin Mathiessen or someone else from FIN-CLARIN
- Marc Kemps-Snijders or someone else from CLARIAH
- Davor Ostojic
- Dirk Goldhahn & Hanna from CLARIN-D

6 Status report of the countries

Austria

- Mapping CMDI -> CIDOC-PE (including 3 day practical/technical Meeting in Rome on mapping with the 3M tool) (Parthenos project)
- Started amc-network - a cocurrence Network Analysis of amc
- Continuous development of the Curation module
- Organisation of the TEI conference

Bulgaria

No report

Czech Republic

Denmark

- Done a Creating CMDI-metadata web-page to make it easier for users to create CMDI-metadata for the DATA resources
- Working on option for storing restricted data in CLARIN-DK, and planning how to make restricted data later aca or pub.
- Working on tutorials for the CLARIN-DK services
- Cleaning up some code, and adding more code to github and svn repositories
- Preparing corpora for Korp, currently only test corpora

Dutch Language Union

No report

Estonia

- Plan to participate in the assessment round 8 in January.
- User involvement at the annual Digital humanities conference (13-14. Oct) and more personally with people from Estonian digital community and people from the Literary Museum.
- Converting more corpora to KORP search interface (ongoing). <http://korp.keeleressursid.ee>
- Reviewing our existing CMDI metadata, preparing for the resources from Literary museum.

Finland

- Done
 - Centrally managed reference instructions: <https://www.kielipankki.fi/corpora/>
- In Progress
 - “Roadshow” to Finnish Universities
 - They show up in usage statistics.
 - Good way to get feedback and prioritise.
 - Moving development to Github, so far “private”:
 - Our Korp instance
 - Selected resources
 - Konversion scripts
 - Our resouce database
 - Defining usage of PIDs for updates, accruing corpora.
 - Metadata curation: Relations of resources.
 - Pending: providing ACA access to eligible persons that do not have the eduPersonAffiliation attribute set. Application via RES process is possible on a resource by resource basis

Germany

- New expansion-phase of CLARIN-D has started this October, lasting until September 2020. Kickoff meeting (CLARIN-D internal) on the 3rd and 4th of November in Leipzig
- Next developer meeting: November the 3rd in Leipzig
- Working group on usage of mime-types was formed, see: <https://office.clarin.eu/v/CE-2016-0882-the-CLARIN-data-fabric.docx>
- FCS 2.0: many centres are working on updating to version 2.0, e.g.:
 - Berlin, Hamburg in alpha version, ready to deploy. Munich work in progress
 - Tübingen will update their cqp-connector used by several centres (not yet started)
- CMDI 1.2: first centres support CMDI 1.2 (e.g. Leipzig), many others working on it
- PIDs: Mapping (forwarding) of EPIC v1 handles to EPIC v2 (including part identifiers) completed
EPIC v1 API shutdown not in the near future (careful planning at GWDG)
- Request by some German centres: Piwik training course

Greece

Hungary

No report

Italy

No report

Latvia

Lithuania

The Netherlands

- On the organizational side, the Meertens Institute, Huygens Institute and IISG have joined forces in the KNAW Humanities Cluster (HuC). The goal of this collaboration is to strengthen innovative, interdisciplinary, humanities research on important scientific and social topics.
- INL is currently known as the Instituut van de Nederlandse Taal (INT). They are, amongst others, responsible for implementing the FCS as part of the CLARIAH project. A project plan for this is currently under preparation and should be coordinated with the CLARIN EU activities.
- The latest versions of OpenSKOS, FLAT and MTAS will be demonstrated at the CLARIN Aix en Provence conference.

Norway

No report

Poland

Portugal

- Waiting for the answer from the national funding agency.

Slovenia

Sweden

- SND will be national metadata repository while the Language Bank will be drop-off point for deposited material. Harvesting functionality underway.
- Language Bank B-centre certification hopefully done by end of year.
- Swe-Clarín needs to apply for new funding in March 2017.
- Kick-off for toolbox some weeks ago. The toolbox is not yet online but will be shortly. It was demoed locally at the kick-off and has generated two researcher blog posts.

United Kingdom

No report

6 Any other business

Next meeting: A doodle will be send out in the course of November 2016