

Title Minutes SCCTC 28 April 2020
Version 2
Author(s) LS, MM, DVU
Date 2020-04-28
Status Approved
Distribution Centre Committee
ID CE-2020-1662



Participants

Martin Matthiesen (Chair, FI), Jan Hajič (CZ), Thomas Eckart (DE), Simon Gray (DK), Lene Offersgaard (DK), Krista Liin (EE), Vanja Štefanec (HR), Riccardo Del Gratta (IT), Roberts Dargis (LV), Daan Broeder (NL), Tomasz Naskręt (PL), Luís Gomes (PT), Martin Hennelly (SA), Leif-Jöran Olsson (SE), Dieter Van Uytvanck (CLARIN ERIC), Linda Stokman (CLARIN ERIC, minutes)

Excused: Maria Skordi (CY)

0 Action points

#	Action	By whom	By when
1	Inform the involved centres of the 14 th assessment round about the outcome of the evaluation	Lene/ CAC	April 2020
2	Distribute the B-centre validity spreadsheet, so a planning for upcoming rounds can be made	Dieter	Before next meeting (28 April)
3	Send minutes which shows the history of the 'b' label discussion to Luis	Dieter	April 2020
4	Prepare a broader discussion on the FAIR Data Maturity Model within the competence of the PID Task force this week and send out an invitation for this. (see questions Martin M. sent to SCCTC by email on 27 April 2020)	Daan & Martin M.	May 2020

1 Agenda

1. Agenda:
 - a. Request for changes? None
2. Approval minutes last meeting & action point status
 - a. Minutes of 31 March meeting ([CE-2020-1652](#))
3. Centre assessments
 - a. Update on the GAMS assessment
4. The "B-Centre" discussion from last meeting
The issue with "B" sounding second best came up again. Suggestion: Let us not reopen up the naming discussion and instead update <https://www.clarin.eu/content/clarin-centres> and explain better that "B" is the Best we have.
5. Review period (until 13 May) of [FAIR Data Maturity Model: specification and guidelines](#) (RDA)
We as CLARIN should be aware of this.
6. The 2020 [draft workplan](#)
Input is still welcome, we aim for final approval in the next meeting.

7. Status update per country/member (please provide a short bullet-wise summary in [the Google Doc](#))
8. Any other business
Organize a virtual meeting on Fedora v6 and migration paths? (suggestion from CLARIN-D developers)

2 Approval minutes and action points

- a) The minutes of 31 March ([CE-2020-1652](#)) are approved.

3 Centre assessments

- a) Update on the GAMS assessment

There are two comments:

1. The user interface is improved, and this point is now settled. Using the VLO as a user interface is acceptable as discussed in the previous meeting.
2. Not settled, according to the curation module over 99% of the metadata files had XML validation issues. not acceptable to move on with the assessment. DVU: Since the validity problems look very similar this can probably be solved in a trivial way. Dieter will contact the GAMS team about this.

Since GAMS has addressed the conditions set in the last meeting the certification can be approved.

4 The “B-Centre” discussion from last meeting

The issue with “B” sounding second best came up again. Suggestion: Let us not reopen up the naming discussion and instead update <https://www.clarin.eu/content/clarin-centres> and explain better that “B” is the **Best** we have.

5 Review period (until 13 May) of FAIR Data Maturity Model: specification and guidelines (RDA)

We as CLARIN should be aware of the FAIR Data Maturity Model: specification and guidelines (RDA): <https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines>

Currently the committee has no further comments to these guidelines.

Martin Matthiessen has sent around an email with the following questions/comments regarding the model:

- -RDA-F1-01D Data identified by a persistent identifier
 - -RDA-F1-02D Data is identified by a globally unique identifier
- I consider those "Useful" at most, not essential. (RDA-F1-01/02M are both Essential)

The reason is given here:

"RDA-F3-01M Metadata includes the identifier for the data"

I am of the opinion that only Metadata should be cited, and thus only Metadata needs a PID, the PID is not essential for data.

- RDA-A1-05D Data can be accessed automatically (i.e. by a computer program)
This is far too vague. Should data be accessible forever just to fulfil this criterion? The requirement is contradicted here:
"RDA-A2-01M Metadata is guaranteed to remain available after data is no longer available"
- If data is not persistent, what is the use of a PID for data (provided Metadata is persistent). If the data is gone, it cannot be accessed automatically. Scripts that rely on that will break.
- RDA-I3-01D Data includes references to other data
How can data reference other data? I understand how metadata can reference other metadata, but not data.
- RDA-I3-02M Metadata includes references to other data
Why would metadata point to anything else but other metadata?

These questions and a brief summary of a related workshop that Martin M. attended, will be discussed in a PID Task force telco. Daan and Martin M. will prepare this broader discussion within the competence of the PID Task force this week and send out an invitation for this. **AP4**

Anyone that likes to take part in the discussion can indicate this by sending Martin M. an email. Participation will be much appreciated.

6 The 2020 draft workplan

See: <https://docs.google.com/document/d/1z7gglw8-rm8HpgnccRI4I7wiRPK45RtTFWlraHX37P0/edit>

- Riccardo's comment to point 1/ Tasks pertaining to the central infrastructure components:

'For humanist is essential to be able to go from a context to the book/text/codex that contains it. This is especially true for historical languages. Do you think it would be available in the future? In FCS, I mean'

There are IDs in FCS, It is just a matter of what ID is needed, based on an example Ricardo and Leif-Jöran will discuss this further.
- Martin Hennelly's comment to point 1/ Tasks pertaining to the central infrastructure components:

'I believe improvement of search is one good candidate for repository related work. Currently (at least in SA) search is limited to text files and basic terms (look up). For

example, audio may have non-CDMI metadata file which is not searchable, or no transcription is available. So expansion of search functionality could be considered? E.g. expansion to audio files, conceptual search, synonym search, exposure to AI assistants... Federated content search is important to search multiple datasets.

This is something that is already available in the FCS. Leif-Jöran suggest having a one on one shot for this with Martin Hennelly, to make sure you are providing the right layers

- There is consensus that pushing the Discovery Service (activity under point 1) is a good idea.
- Martin M. will try to find out what is the situation with countries that are not actively participating in the centre committee meetings to see what can be done about this ((activity under point 1).

The centre committee approves the SCCTC Workplan 2020 as it is, including the topic, 'Outreach about current FCS2 capabilities and uptake of FCS2 in relation to legacy and older versions', added under point 3 by Leif-Jöran.

7 Status update per country/member

Austria

- No report

Bulgaria

- No report

Croatia

- Nothing to report

Czech Republic

- Working on modest equipment acquisition under a new project by the MEYS, with partners (E100k), preparation for the Humaine AI Net projects, writing another H2020
- Finishing TEITOK updates and deployment to Czech Parliament corpora
- Continued preparation of two Shared Tasks (data, tools), other projects (e.g. PDT)

Cyprus

- No report

Denmark

- Nothing to report, only minor upkeep on existing infrastructure.
- Took part in the "FAIRification of Nordic and Baltic data repositories" seminar on 22.04

Estonia

- Took part in the “FAIRification of Nordic and Baltic data repositories” seminar on 22.04.
- Organized a seminar on corpora and annotation for literary sciences on 27.04.
- Set up [an instance](#) of [Inception](#).
- Adding new corpora (language learning studybooks and correspondences) to KORP.

Finland

- Took part in Workshop “[FAIRification of Nordic and Baltic data repositories](#)”
- Joint “FIN-CLARIAH” application in preparation
- Korp update progressing
- Infrastructure consolidation in progress (virtual machines, collaboration tools)

France

- No report

Germany

- Further work on integration of Shibboleth in FCS aggregator (with partners from SWE-CLARIN and CLARIN-ERIC)
- First case of "resource takeover" at a CLARIN-D centre to ensure long-term availability; discussion with CTS board
- Discussions about future use of Fedora Commons

Greece

- No report

Hungary

- No report

Iceland

- No report

Italy

- ILC4CLARIN
 - Import of ALIM Archive into repository. Continuing...

Latvia

- Nothing to report

Lithuania

- CLARIN-LT [consortium](#) agreement renewed

The Netherlands

- No news

Norway

- Took part in the “FAIRification of Nordic and Baltic data repositories” seminar on 22.04.

Poland

- Nothing to report

Portugal

- Continued work on improving the PORTULAN CLARIN workbench tools
- First steps on implementing a Federated Content Search endpoint

Slovenia

- Finishing the revised application for CTS
- Published call for CLARIN.SI projects 2020 (<http://www.clarin.si/info/storitve/projekti/>, in Slovene)
- Published siParl 2.0, <http://hdl.handle.net/11356/1300>
- Organisation of the final RDA Node Slovenia event (virtual, June 2020)
- Presentation of CLARIN(.SI) for students of computer science

South Africa

- No report

Sweden

- Working on FCS-AAI e.g. integrating federated login and new restricted access resources functionality in Aggregator with partners from CLARIN-D and CLARIN-ERIC
- Held Swe-Clarín consortium meeting
- Preparing new NER resources for release
- Supporting grant application activities

United Kingdom

- National Coordinator (Martin Wynne) has been placed on furlough for 3 weeks from 11/5/20 by the University of Oxford, so no activity is expected to take place

8 AOB

- Organize a virtual meeting on Fedora v6 and migration paths? (suggestion from CLARIN-D developers)

Thomas explains that the general problem is that most centres in Germany use Fedora v3 but this version is now outdated. They changed the whole system including the data model and APIs. The recent version's problem is that some add-ons (e.g. OAI-PMH) that are really needed are no longer part of the system.

There will be an improved update so we can migrate from v3 to v5 or v6. Many centres have to migrate because running outdated software is not good. The question is if it is a good idea to migrate to v5 or v6 or to find another other solution like D-Space.

Dieter comments that this could be a workshop, where someone from Fedora is invited to present their future plans. The idea is to partially look inside our own centres at what the problem is but also to give Fedora the opportunity to present their future plans and to put the obvious question concerning OAPMH support on the table.

This virtual workshop can be initiated under the CLARIN umbrella (CLARIN workshop). The topic can be seen in a broader perspective. It is an interesting recurring topic, giving centres a sneak preview into the future of repositories and to make choices about what repositories to use.

Before sending a Doodle to the centre list we need to check the availability of the Fedora people, suggest 3 or 4 days where they can make it. Thomas will initiate an email to find out who to contact at Fedora (with DVU in cc).

Next meeting: Tuesday 26 May 2020 at 10:00 CEST (via Zoom).