CLARIN Standards Committee meeting minutes 2019-09-30

Agenda

1. opening, roll call and introductions
2. activity reports 2018-2019
3. changes in the by-laws: presentation and discussion
4. towards statutory goals: spring f2f meetings
5. towards statutory goals: current suggestion for work towards a unified standards list
6. towards statutory goals: CSC homepage / presentation space
7. any other business

Members present

1. Piotr Bański (Germany)
2. Dieter Van Uytvanck (ERIC, BoD Liaison), until § 6
3. Jan Odijk (Netherlands)
4. Andrea Bellandi, sitting in for Fahad Khan (Italy)
5. Francesca Frontini (France)
6. Jussi Piitulainen (Finland)
7. Neeme Kahusk (Estonia)
8. Dorte Haltrup Hansen (Denmark), observer
9. Leif-Jöran Olsson (Sweden)
10. Bente Maegaard (ERIC), guest, present for § 3

Excused

1. Daan Broeder (Netherlands)
2. Maria Gavrilidou (Greece)
3. Hanna Hedeland (Germany)
4. Charly Mörth (Austria)
5. Andreas Witt (Germany)
6. Christian Thomas (Germany)

§ 1 Opening

The proposed work order is to do § 3 first, in order not to keep Bente from participating in another parallel meeting, and then to proceed through items 4-6 in order to manage to cover them while

Dieter is in the room (he has to leave at 15:00).

Role call:
- Piotr, chair
- Leif-Jöran, secretary
- Bente, guest for the discussion of the by-laws

§ 3 Changes to the by-laws

[Members have had an opportunity to look at the proposal prior to the meeting.]

No immediate questions to Bente. Bente notes that the bylaws are based on the assumption of reasonable behaviour.

Piotr points out the "can" in Article 2 "Membership" (in the sentence, "Every CLARIN ERIC member and observer consortium *can* select one or more experts for membership(…)") and mentions two views, one seeing the committee as relatively small but active, the other as large and with potentially many inactive members who nevertheless may constitute transmission channels for information from the CSC.

Someone : but there are other transmission channels in existence, e.g. according to the bylaws, the CSC meeting minutes are submitted to the NCF.

Piotr: in a presentation to the NCF given earlier today, I mentioned that the CSC would appreciate funding for its members to travel to the hands-on meeting (discussion upcoming in §4); the size of the committee may indirectly influence such funding decisions.

Piotr recalls Charly's opinion from the earlier discussion of bylaws (that took place between Bente, Charly and Piotr) that smaller consortia might not wish to appoint anyone as their representative on the CSC.

Jan: if no representative is selected by a national consortium, we would like to know the reason, to make sure that it was a conscious choice.

Piotr points out the two year terms and stresses that this is not meant as a strict formulation (we value our experts) but rather as a commonsense approach, possibly motivating for members. Earlier in the NCF meeting, Piotr suggested that the national representatives may ask for (e.g.) yearly reports, which would then be very handy for the CSC reporting tasks.

No further discussion, Bente leaves, with thanks from the other participants.

Jussi, afterthought: uneasy about secret voting. Why would we ever need that?

Francesca: members, for institutional reasons, may sometimes need to resort to that, it is good to keep that option.

Others: secret voting not first option so let's keep it as is, for fringe cases.

Someone (Jussi?): voting by email should be allowed.

Action: We request Bente to do the editorial changes and present the bylaws to the CSC for the last round of feedback before they are presented to the BoD.

§ 4 Towards statutory goals: spring f2f meetings

Piotr: the bylaws mention minimally 4 meetings per year. Suggestion (cleared with Charly): let us have a face-to-face meeting in the spring, designed solely as a hands-on session at the end of which we produce an updated list of CLARIN standards. This way, the minimal work cycle of the

committee will be: the mostly administrative meeting at the CAC, the hands-on meeting in March, and two virtual meetings inbetween. Dieter has already been consulted about this and declared that, if the meeting were to be organised in Utrecht, the CLARIN Office would provide organisational help and meals. Travel expenses would have to be paid by the members' institutions/consortia. Piotr mentioned this in the NCF meeting as one way in which NCF can provide assistance to the CSC.

Decision: suggestion accepted by the members present.

Piotr: started a doodle at <https://doodle.com/poll/k3v4zv892ksipuvc>, for individual days of December, for the first virtual meeting. At this meeting, we will lay the ground for the hands-on meeting (identify work packages, teams, action points to be completed in advance, etc.).

Members: better to approach the doodle already subdivided for parts of the day.

Action: Piotr will introduce 2-hour slices into the doodle in the evening so don't participate until tomorrow. Piotr will bother everyone until Wednesday to participate.

Later, as an afterthought, Jussi: what software are we going to use for the meeting, I wouldn't like to install anything extra for security reasons. Piotr: most probably zoom. Leif-Jöran: a different solution (Jitsi), Web-based, is also at our disposal if needed.

§ 5 Towards statutory goals: current suggestion for work towards a unified standards list

Dieter's suggestion:

- leverage the sources of information for one of the KPIs, defined as "percentage of centres that publish explicit information on what formats they accept data in"

- count the formats mentioned by the individual centre and sum up the counts for individual formats

- use a threshold to decide whether a format is "recommended" by CLARIN; with 7 centres (against the total of 22 B-centres) returning this information, the threshold can be placed at 2

This is a bottom-up approach that is synergistic with a BoD goal. Document:

https://www.dropbox.com/s/wovaiars1digj9n/formats-supported-mimetypes.xlsx?dl=1

[Update: the idea is elaborated on in the Bazaar poster presented on 1.10.2019: https://www.clarin.eu/sites/default/files/clarin2019_bazaar_csc.pdf]

Data from 8 centres is also present/linked here: https://www.clarin.eu/content/standards-and-formats#formats

Piotr: apart from the numerous "lists of recommended formats" floating around, there has been some research made by Hanna Hedeland, with results available at [1] and [2] (linked from the agenda googledoc and known to the members from last year's meeting and Bazaar), so we have a nice basis to move up from all this. And the fact that the reporting is tied to BoD research seems to be of paramount importance, because it hopefully ensures uniform and positive reactions from centres.

[1]: https://docs.google.com/spreadsheets/d/15wcsOzYZrhfQ1knFgpLQVstOMBxACvrlttgWY6cn8UQ/edit ("Centres and Standards")

[2]: https://docs.google.com/spreadsheets/d/1EvNA9FWMubVkPadeZHndj1FtRbg4iOQxFDVyO0p-BgE/edit

A small discussion ensues concerning the granularity of denotations of format names, with Jan reminding us that when it comes to providing format-related information by centres, the principle to be followed is "the most specific format wins" (if you require a dialect of TEI, don't say that you support "TEI", or, even worse, "XML".

Piotr: starting from Dieter's basis may be seen as a step back, but it's a start from a very firm ground.

Action: we create something more refined from this spreadsheet and the extracted information from centres which should be published at the end of the f2f meeting in March. Milestone: see § 6 for December.

§ 6 Towards statutory goals: CSC homepage / presentation space

Piotr: in a discussion with representatives of the BoD, Bente pointed out that the CSC might want to use presentation space at clarin.eu akin to how CLIC does that. Question: should we post about our intentions or rather when something is there?

Others: the latter. We can have a statement out after the virtual meeting in December.

Dieter: one standard way of handling it is to create a blog article.

Action: at the December telco, apart from preparations for the March f2f, let us prepare a statement to be published in the presentation space maybe in the form of a blog post.

[update: https://www.clarin.eu/content/standards ; note that this page is not a CSC page per se, because it seems to be virtually shared by the CSC, the Interoperability Committee, and two metadata-related taskforces; Update 2: decision from the December telco: go ahead and use this page, request more pages if needed]

Dieter leaves to join a parallel meeting, with thanks for his participation and input.

§ 2 Activity reports 2018-2019

Piotr: we used to be asked for reports every month but that didn't work out. Still, once a year is not a bad idea, both for us, and for the BoD, maybe even the GA. Piotr input his bullets before the meeting and asks other members to consider adding their short statements. Some of them appear during the meeting or soon after.

<https://docs.google.com/document/d/1XziJnL54BQMj7UwJ8KVi09sTCcRnMX8odgUY9gDhKQY/edit#>

§ 7 Any Other Business

a) Jan: Interoperability committee requests feedback by the end of October on a document.

Action: Jan sends link when ready. This is the link: https://docs.google.com/document/d/1mJiGQhReA_Km6DldiNmPpWbVlL2Yz6Fy0tXIQjy2ZZ8/edit?usp=sharing

Suggested action: We come with comments by end of October in the document.

b) Piotr informs the CSC about the ongoing initiative to create a common TEI-based standard for CLARIN and beyond to encode parliamentary data for linguistic research, Parla-CLARIN, led by a CSC member, Tomaž Erjavec, with his colleague Andrej Pančur. Jan has also played a key role on the organizational side.

Jan: talks about the initiative, mentions the first workshop in Amersfoort, mentions plans for a hands-on encoding workshop after Tomaž and Andrej have received feedback on the first version of the documentation, currently out and up for review at https://github.com/clarin-eric/parla-clarin

Small discussion ensues. Leif-Jöran: Ad-hoc tags but different semantics, so mappings should be enough to show that an effort is needed to make queries across the different datasets. Probably an action to ensure semantic interoperability is required as well.

c) Tomorrow's Bazaar

Piotr asks if the committee wants to be represented at the Bazaar and if so, he requests help in handling the potential visitors, Leif-Jöran volunteers, with Jussi offering support.

Action (in the longer perspective): Make sure your centre(s) is/are among those who offer explicit information on the formats they accept.

[Update: https://www.clarin.eu/sites/default/files/clarin2019_bazaar_csc.pdf]

d) Knowledge-Sharing Infrastructure Committee

This is an item that we talked about last year, and Jan has produced a slick spreadsheet for accumulating the information. It is at https://docs.google.com/spreadsheets/d/1tm9INEICSPvjWam_X8VC9bPrGNf2to6TV9HwG0Q4riY/edit

Action: Piotr will ask the CSC to check if their data are present in the document, and will share it with the KSIC.

[Update: shared with Steven Krauwer on Oct 2]

End of the meeting. Thanks to everyone for participating.