

CTS@ILC

A long story
riccardo.delgratta@[ilc.]cnr.it



Center: ILC4CLARIN¹

Consortium: CLARIN-IT²

- ILC4CLARIN certified³ DSA (Data Seal of Approval) under 2017-2019 guidelines
 - The process lasted ~ 1 y.
- ILC4CLARIN started the new certification with CTS under 2020-2022 guidelines⁴ in March 2021
 - The process is still ongoing;
 - 4 review rounds (so far...).

¹<https://ilc4clarin.ilc.cnr.it/>

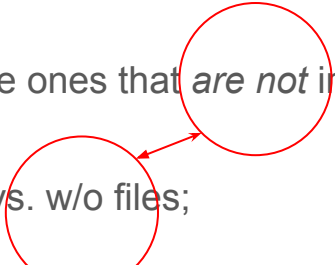
²<http://www.clarin-it.it/>

³<https://www.coretrustseal.org/wp-content/uploads/2018/04/The-ILC4CLARIN-Centre-at-the-Institute-for-Computational-Linguistics-.pdf>

⁴<https://zenodo.org/record/3638211#.ZEliaOxByCc>

Issues pointed out by reviewers (but solved....

.... either addressing the points of the reviewers or changing the level of the requirement)

- About how the repository works
 - Format migration;
 - Standards in LRT community;
 - Quality of data (especially for the ones that *are not* in the repository).
 - About our implementation
 - Number of submissions w files vs. w/o files;
 - Staff involved.
- 

R15, first certification process

R15 was accepted at level 4, fully implemented, in the first certification.

- R15 Technical infrastructure R15.
 - The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
- R15 Guidelines.
 - Guidance: Repositories need to operate on reliable and stable core infrastructures that maximizes service availability. Furthermore, hardware and software used must be relevant and appropriate to the Designated Community and to the functions that a repository fulfils. Standards such as the OAIS reference model specify the functions of a repository in meeting user needs. About our implementation
- Answer (main points)
 - Based on DSPACE by (formerly) LINDAT/CLARIN;
 - A clear GitHub repository, constantly maintained;
 - LRT standards (from the CLARIN community) naively adopted;
 - ILC4CLARIN has a redundant architecture;
 - ILC4CLARIN is connected to the GARR and GEANT network.

R15 second certification process

R15 was accepted at level 4, fully implemented, in the first certification.

- R15 Technical infrastructure R15.
 - The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
- R15 Guidelines.
 - Guidance: Repositories need to operate on reliable and stable core infrastructures that maximizes service availability. Furthermore, hardware and software used must be relevant and appropriate to the Designated Community and to the functions that a repository fulfils. The OAIS reference model specifies the functions of a repository in meeting user needs.
- R15 addendum on disaster plan
 - Does the repository have a disaster plan and a business continuity plan? In particular, are procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage? What are they?

R15 first review loop

R15 was submitted at level 4 as in the first certification, with minor changes wrt the previous submission in March 2021. Namely:

- Few information on the network, since we do not offer real time services
- Disaster plan not addressed since R16 was accepted at level 4.

In July 2021 we received the first review loop:

- Reviewer 1 (R1)
 - The guidance document asks for information on: "Are availability, bandwidth, and connectivity sufficient to meet the needs of the Designated Community?"

"Does the repository have a disaster plan and a business continuity plan? In particular, are procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage? What are they?"

There is no information apparent on these two considerations. In particular, disaster recovery and backup information.
- R1 noticed an incongruence between the number of of servers in R15 and R16. We fixed it.

R15 second review loop

R15 was re-submitted in October 2021, again at level 4, but with additional information to meet R1 comments:

- The local infrastructure is completed with a Synology NAS hosted at a different CNR institute.
- Both the location of the servers and of the backups are hosted within the CNR Research Area of Pisa which provides
- network security, monitoring, and protection (such as firewalls).
- The disaster and recovery plans take into account the following aspects:
 - 1) Software repository GitHub
 - 2) Customization GitHub
 - 3) Periodic Backups
 - 4) Distinct locations for server and backups

In October 2021 we received the second review loop

- R1
 - This does not appear to be fully implemented - see comments below.
REVISION: Accepted
 - The interface reported the previous comments:
 - "Does the repository have a disaster plan and a business continuity plan? In particular, are procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage? What are they?"
There is no information apparent on these two considerations. In particular, disaster recovery and backup information.

R15 third review loop

We worked on other requirements and in December 2021, R15 was re-submitted as it was. This because we interpreted the information

- This does not appear to be fully implemented - see comments below. with the usual comments The interface reported the previous comments: "Does the repository have a disaster plan and a business continuity plan? In particular, are procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage? What are they?" There is no information apparent on these two considerations. In particular, disaster recovery and backup information as a sort-of-history of the CTS interface
- We took the sentence REVISION: Accepted as the fact that R15 was already accepted by R1.

In September 2022 we received the third review loop

- R1 (again)
 - This does not appear to be fully implemented - see comments below.
REVISION: Accepted
 - The interface reported the previous comments:
 - "Does the repository have a disaster plan and a business continuity plan? In particular, are procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage? What are they?"
There is no information apparent on these two considerations. In particular, disaster recovery and backup information.

R15 fourth review loop

We re-submitted the application in October 2022, fixing some typos.

- This does not appear to be fully implemented - see comments below. with the usual comments The interface reported the previous comments: "Does the repository have a disaster plan and a business continuity plan? In particular, are procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage? What are they?" There is no information apparent on these two considerations. In particular, disaster recovery and backup information as a sort-of-history of the CTS interface.
- We took the sentence REVISION: Accepted as the fact that R15 was already accepted by R1.

In April 2023 we received the fourth review loop

- R1
 - The disaster recovery plan is mentioned at the end of the section, but is not available for review. As such this is still not really addressed. If the applicants can point to a hyperlink to the disaster plan and business continuity plan, this would address these concerns.

And here we are, Your Honor (1/3)

Some comments

- First interaction 4 months, second interaction 4, third interaction 9, fourth interaction 7:
 - On average 6 months
- Fifth?
- I understand the revision process takes a long time. I also understand CTS problems, but a 6-month-average between submission and response is too long;
 - If the process requires 2 or 3 review loops, part of the certification validity is spent for the process;
- As ILC4CLARIN we have been particularly unlucky and we spent the complete period of certification validity (2021-2023) to complete the certification process which, in theory, expires on April 2024.

And here we are, Your Honor (2/3)

Some comments

- ILC4CLARIN is part of CLARIN ERIC and, like many other centers, uses CLARIN-DSPACE and "copies" its strategies and technicalities, such as (for example) the use of standard and format migrations, from the CLARIN-DSPACE developers;
- We spent a lot of time, in the process, to tune the texts of CTS requirements about format migration to meet the reviewers' comments.
- We decided to use a third-party software even to be "freed" from such technicalities: we know that if CLARIN-DSPACE meets the requirement, so we do the same;
- To be part of a community means also to access shared knowledge, which is used in case of problems;
- Many of the points of both reviewers make sense if the software is not widely used or if it is in-house.

And here we are, Your Honor (3/3)

My personal impression on R1

- The impression on R1 is that the reviewer was not clear when asked for additional requirements and the text revision ACCEPTED was misleading;
- Probably the interface did not help, now it seems to be clearer;
- What R1 asks is, however, outside the ILC4CLARIN center. Since:
 - The hosting institute (ILC) is part of a Research Area which is in charge of
 - managing network, its topology, bandwidth, firewall rules and connection between LAN and WAN;
 - offering smart isles to host the HW. When we write that our data is backed-up at a different institute, we demand all the security to the Research Area;
 - The Research Area of Pisa does not have a plain disaster plan, not in a modality I can verify. The area hosts many CNR institutes with different needs.

So, what can we do?

As ILC4CLARIN responsible I ask how to proceed, taking into account the following aspects

- We'd like not to go back to level 3, since we were at level 4 and we are continuing working on these aspects especially to the big NRRP H2IOSC project;
- We can not provide a detailed Disaster Plan
- Are the considerations enough to re-submit the application as it is and to be certified?

- And for CLARIN:
 - what about our period of certification validity?