

# Task: New CLARIN Centre CLST Nijmegen

ID: T018

**Author:** Henk van den Heuvel, Maarten van Gompel

**What exists:** CLST holds the status of CLARIN C and T centre. As a CLARIN C Centre CLST already offers component based metadata (CMDI) that make use of elements from accepted registries such as the CCR1 in accordance with the CLARIN agreements (harvestable via our OAI-PMH bridge). As T Centre CLST already has the proper CLARIN FIM in place.

Furthermore we have the successful [CLST Webservice Portal](#) hosting a variety of CLAM-based webservices for various NLP tools (Frog, ucto, PICCL, T-scan, Colibri Core, Valkuil, etc..). These services are currently protected by a simple authentication layer with a stand-alone user database. People can register for a free account.

**What must be adapted / extended / created anew:** CLST has the ambition to host language resources for atypological speech (e.g. of language learners and pathological language speakers), and strives to become a B centre for such resources. In order to become a CLARIN B Centre CLST should fulfil the CLARIN Assessment criteria laid down in <https://www.clarin.eu/node/3767> and specified in check list presented at <https://www.clarin.eu/node/3577>.

We will prepare the certification in 2019, request long term commitment and matching from the Faculty of Arts by the end of 2019 (go/nogo), and after a positive outcome build the remaining required infrastructure for the qualification as B centre which should then follow in 2020.

A strong CLARIN-proof authentication procedure is relevant for a B Centre, especially for the resources that we will host. Therefore further integration in the CLARIAH infrastructure would be welcome by moving to the federated single-sign on framework. First of all, however, the feasibility of this has to be studied: CLAM has already been extended with OAuth2 support on request of CLARIN years ago, but nobody still uses this and we're not sure if the necessary OAuth2 infrastructure is already in place. The single-sign on solution needs to be compatible with RESTful webservices, which is not trivial. Proper metadata for the webservices has to be created and published as well, a large part of which is taken care of in the scope of T098 (LaMachine).

**Why important for CLARIAH (scientific impact):** The Netherlands is in need of a CLARIN B centre for atypical speech resources from which such data can be safely shared with other CLARIN partners.

Also in this context, the work on authentication improves accessibility by enabling researchers to access various services using their institutional credentials. This would contribute to turning Nijmegen into a CLARIN centre.

Although not yet a CLARIN B centre, we de-facto already provide a variety of CLARIN/CLARIAH based services to the outside world through our Webservice Portal for a long time.

**Targeted/Actual users:** Researchers

**Actual use (quantify!):** Current actual use of the webservices is quantified exactly in our [statistical report](#).

**Social Impact (concrete examples):** better accessibility/findability

**Lead + PMs:** Henk van den Heuvel

**Proposed PM estimation (try to justify):** 6PM (+4PM conditionally)

- Maarten van Gompel: 3 PM (FIM),
- Wessel Stoop & Erwin Komen & ??: 2 PM (2019) and, conditionally, after obtaining go-decision: 4 PM
- C&CZ system administrator: 1PM;

**Assigned PMs:** 5PM (-3)

**Remarks:**

- (This task is mostly system administration rather than development)
- See also T065
- [CLAM](#)

## Deliverables

1. Prepare the certification for B Centre in 2019
2. Request long term commitment and matching from the Faculty of Arts by the end of 2019 (go/nogo)

3. (After a positive outcome) build the remaining required infrastructure for the qualification as B centre which should then follow in 2020.
4. Access to CLAM-based webservices at CLST via federated single-sign on
5. Access to FLAT at CLST via federated single-sign on