**Topic:** A CLARIAH Virtual Research Environment for Language Research

IDs in Google Sheets: T137 Author: Daan Broeder

**What exists:** The CLARIAH infrastructure currently consists of a number of fragmented services and tools that not offer the user a single point of entry and makes the user responsible for many data management tasks that can be automated. We are looking to build an integrative platform that would alleviate these problems offering the user a one-stop-shop for as many of his needs as possible, offering support for the whole data life cycle.

Currently there are a number of standards, tools and services available that can possibly be partly reused or can be use as an example.

- different existing LT services and software libraries to create them e.g. FLAT
- The Nederlab research portal as a basis platform with default integrated MTAS search engine
- MTAS, BlackLab search engines
- Folia and CMDI as annotation and metadata standards
- TTNWW/PiCCL as examples for chained linguistic service recipes
- The VRE effort in CLARIAH-CORE
- Examples of VRE platforms outside CLARIAH e.g. PARTHENOS platform

## What must be adapted / extended / created anew:

The VRE would need to support (following the data life cycle)

- data generation (via import) and manual editing of metadata and content
- data discovery (using metadata, content, semantic<sup>1</sup> search)
- data selection and (virtual) collection management,
- private workspaces for individual users
- data processing via a registry of recipes of chained LT services
- visualization, statistical analysis
- export and data publication/archiving.

The VRE should offer a plug-in architecture allowing to link in external web-services, but also allow integration of local tools as Praat and ELAN or local packages as R. The Portefolio of available services can be adapted to the user profile. Users should be able to analyze their own uploaded data in the context of existing standard data-sets e.g. those available in Nederlab or OpenSonar.

## Why important for CLARIAH (scientific impact):

There is no new scientific impact, but it will greatly enhance the impact of the CLARIAH infrastructure. By bringing CLARIAH tools, services and collections to the scientific user's desktop more easily many more researchers can take advantage of CLARIAH to address their research questions.

**Targeted/Actual users:** all humanities scholars that need an integrated environment for the CLARIAH tools, services or collections

Actual use (quantify!): none as yet

Social Impact (concrete examples): more easy use of CLARIAH infrastructure, more users

<sup>1</sup> Semantic search is used here in the sense that users are able to add meaningful shareable labels to data objects in the VRE which can be use to group and search data

Estimate in PMs (try to justify): 18 PM: coordination, creating and running software as a

service, testing

Lead + PMs: MI: 12PM

**Participants + PMs:** 6PM for tool/service integration FLAT, ELAN, Praat, PiCCL...