

CLARIN Data Providers and Associated Centres

CLARIN Data Providers

There is a range of centres in the Netherlands (and elsewhere) that by their very nature and mission collect, maintain, and make accessible large amounts of data. Such centres can provide data that are very important for researchers in the humanities and social sciences. Examples of such centres in the Netherlands include the *Koninklijke Bibliotheek*, *Beeld en Geluid*, and het *Nationaal Archief*. CLARIN-NL wants these data to be visible, and accessible in a transparent manner in the CLARIN infrastructure, but the centres involved are often not prepared to become a full CLARIN centre with all obligations that that brings with it.

We therefore propose to define a new type of centre, called *CLARIN Data Provider*, to meet these apparently contradictory requirements. The basic idea is that we want the centres to operate as much as possible in accordance with their own mode of operation, but with minimal requirements to make them CLARIN-compatible

A centre can become a *CLARIN Data Provider* in this sense, when it meets the following requirements:

- It is a centre that has the acquisition, maintenance, long term preservation and distribution of data in its core mission.
- It makes available (a subset of) its data to researchers that have access to the CLARIN infrastructure
- It makes available metadata for these data
 - in CMDI format
 - with the metadata contents in accordance with its own conventions
 - with Dublin Core as a minimum
- It allows harvesting of these metadata by other CLARIN centres in accordance with the protocols required by CLARIN (OAI-PMH)
- It allows others (CLARIN centres, researchers) to make enriched versions of the metadata, which will be stored and maintained at a CLARIN-centre and linked to the original metadata
- The center guarantees stable URIs for the resources augmented with a system of unique identifiers that can be used for creating citations.
- If there are limitations on some of the data of the centre, it joins the national identity federation where available and joins the CLARIN service provider federation to support single identity and single sign-on operation based on SAML2.0 and trust declarations. If all data are made available publicly, this is not necessary.
- It has a proper and clearly specified repository system and participates in a quality assessment procedure as proposed by the Data Seal of Approval or MOIMS-RAC approaches.

CLARIN Data Providers and Associated Centres

- It makes clear statements about its policy of offering data and its treatment of IPR issues

Associated Centres

There is a different class of organizations, for example *Fryske Akademy*, and the *Tweede Kamer*, which have very interesting data that they also are prepared to make available in a CLARIN-compatible manner. However, storing, maintaining and distributing data is not their core task, and long term preservation is not in their mission at all. Such organizations can become an *Associated Centre*.

- It is a centre that is willing to make available (a subset of) its data to researchers that have access to the CLARIN infrastructure
- It makes available metadata for these data in accordance with CMDI formats, profiles and components
- It allows harvesting of these metadata by other CLARIN centres in accordance with the protocols required by CLARIN (OAI-PMH)
- The data and metadata are referred to by persistent identifiers (PIDs) supported in CLARIN.
- If there are limitations on some of the data of the centre, it joins the national identity federation where available and joins the CLARIN service provider federation to support single identity and single sign-on operation based on SAML2.0 and trust declarations. If all data are made available publicly, this is not necessary.
- It makes arrangements with a CLARIN centre (e.g. DANS) to ensure long term preservation of the data
- It has a proper and clearly specified repository system and participates in a quality assessment procedure as proposed by the Data Seal of Approval or MOIMS-RAC approaches.
- It makes clear statements about its policy of offering data and its treatment of IPR issues

Application Procedure

A centre can become a Candidate CLARIN Data Provider/Associated Centre when

- It makes a request to the CLARIN-NL board to become a CLARIN Data Provider/Associated Centre, and this request is honored.
- It presents a plan (incl. timeline and budget) to meet the requirements it must satisfy
- The center will appoint a CLARIN contact that will have the appropriate expertise to represent the center in relevant CLARIN discussions and that can be approached in case of problems with the center's agreed services
- The center will sign a memorandum stating that the agreed services will remain stable for a fixed period of time (that can be renewed) and that in case of 'force majeure' or if the center does not want to renew the agreement, it will cooperate to transfer these services to another (associated) CLARIN center.

CLARIN Data Providers and Associated Centres

If the plan is approved, the CLARIN-NL board may decide to provide support (financial or otherwise) to the centre so that it is aided in making the transition from candidate CLARIN Data Provider to recognized CLARIN Data Provider. Funding will only be provided for tasks that are needed to make this transition. For organizations such as the KB, B&G and NA, it can be expected, because of the very nature of these organizations, that meeting the requirements mentioned will be relatively easy. For organizations such as Fryske Akademy and the Tweede Kamer more effort and especially more guidance will be required to obtain the desired status.