

Title	Clarín Centre Types
Version	0.8
Author(s)	Peter Wittenburg, Dieter Van Uytvanck, Thomas Zastrow, Lene Offersgaard
Date	2020-02-18
Status	For approval by SCCTC
Distribution	SCCTC, CAC, BoD
ID	CE-2012-0037



The purpose of this document is to establish the procedure for selecting the centres that will participate in the Standing Committee for CLARIN Centres.

CLARIN distinguishes a number of different centre types that have different impact for the emerging language resources and tools landscape, and for completeness we list them all here.

1. Centre Types

At the core of the CLARIN infrastructure backbone are:

Service Providing Centres (Type B)

Task: Type B centres offer services that include the access to the resources stored by them and tools deployed at the centre via specified and CLARIN compliant interfaces in a stable and persistent way;

Examples: *the corpora stored at the centre, the language tools being developed by that centre, etc.*

Requirements: Type B centres need to fulfil the requirements mentioned in chapter 2 where they do apply.

Agreement: CLARIN ERIC will sign a Service Level Agreement to specify type and characteristics of the offered services.

Knowledge Centres (Type K)

Task: Type K centres offer expertise and advice about various matters that are relevant for the researchers to easily make use of the CLARIN services and that are not covered by the other centres;

Examples: *how to do the digitization, OCR and integration of book material, how to find taggers and parsers for medieval documents, etc.*

Requirements: Type K centres need to fulfil requirements which need to be specified in an agreement.

Agreement: CLARIN ERIC will sign a Service Level Agreement to specify type and characteristics of the offered expertise.

1.4 External Centres (Type E)

Task: Type E centres offer CLARIN relevant services, but these services are not offered by members of CLARIN; in general these will be common infrastructure services, i.e. external centres will often be type A centres¹;

Examples: *persistent identifier service, a long-term preservation service, etc.;*

Requirements: Type E centres need to fulfil the requirements mentioned in chapter 2 where they do apply.

Agreement: CLARIN ERIC will sign a Service Level Agreement to specify type and characteristics of the offered services.

¹ For CLARIN ERIC

Most of the services offered by these centres are crucial so that CLARIN ERIC will sign Service Level Agreements with the corresponding centres, that specify the characteristics of the offered services, and will take measures to monitor the accessibility of them. In case of knowledge centres CLARIN ERIC will want to assess the quality of the advice that is given etc. In general Service Level Agreements with centres offering infrastructure type of services which are crucial for the whole will be formulated with a high expectation on availability.

Several CLARIN centres may give a mixture of service types, i.e. it is possible that very strong centres offer the resources stored by them (Type B) and give advice about CLARIN relevant matters such as standards (Type K). This simply means that such centres take over more responsibilities.

There will be many more institutions that have interesting language resources and tools to offer, but who are not able or do not want to fulfil the CLARIN requirements and thus cannot offer core services. These can roughly be classified in two types:

1.5 Metadata Providing Centres (Type C)

Task: Offer machine readable metadata in a stable and persistent way allowing service providers to harvest their metadata and making them browsable, searchable and combinable;

Requirements: Type C centres are not requested to fulfil the requirements mentioned in chapter 2; however they are expected to serve metadata via the OAI-PMH protocol.

Agreement: there will be no Service Level Agreement being signed, i.e. researchers cannot rely on the availability of any service.

2. Requirements for CLARIN Centres (B, E)

The following list of requirements only holds for centres of types B and E

- (a) Centres need to offer useful services to the CLARIN community and to agree with the basic CLARIN principles (own architecture choice, explicit statement about quality of service, usage of persistent identifiers, adherence to agreed formats, protocols and APIs).
- (b) Centres need to adhere to the security guidelines, i.e. the servers need to have accepted certificates.
- (c) Centres need to join the national identity federation where available and join the CLARIN service provider federation to support single identity and single sign-on operation based on SAML2.0 and trust declarations. In case all resources at a centre are open, setting up a Service Provider is optional.
- (d) Centres need to have a proper and clearly specified repository system and participate in a quality assessment procedure as proposed by the CoreTrustSeal.
- (e) Centres need to offer component based metadata (CMDI) that make use of elements from accepted registries such as the CCR in accordance with the CLARIN agreements, i.e. metadata needs to be harvestable via OAI PMH.
- (f) Centres need to associate PIDs records according to the CLARIN agreements with their objects and add them to the metadata record.
- (g) Each centre needs to make clear statements about their policy of offering data and services and their treatment of IPR issues.
- (h) Each centre needs to make explicit statements to the CLARIN boards about its technological and funding support state and its perspectives in these respects.
- (i) Centres need to employ activities to relate their role in CLARIN to the research community in order to guarantee a research based status of the infrastructure and allow researchers to embed their services in their daily research work.
- (j) Centres that are offering infrastructure type of services (E) need to specify their services for CLARIN and the terms of giving service.
- (k) Centres are advised to participate in the Federated Content Search with their collections by providing an SRU/CQL Endpoint. This content search is especially suitable for textual transcriptions and resources.

Service Level Agreements will help to make all offerings explicit and describe the availability conditions.

3. Centre Assessment Procedure

For all centres of type B the CLARIN ERIC shall have an assessment procedure that will check what the value of the services for CLARIN is, what the state of the services is, how the quality of the service can be evaluated over time and what kind of agreement will be required. To carry out this process the Board of Directors will set up an assessment committee, including CLARIN and external experts.

The procedure shall be as follows:

1. A negotiation phase will either be started by an interested centre or by CLARIN ERIC.
2. The centre and CLARIN ERIC will discuss the services to be offered and classify them.
3. The CLARIN ERIC will ask the assessment committee to check the state of the centre and the services.
4. The assessment committee will send a questionnaire² where the centres need to make specifications. The assessment committee will appoint evaluators which will go through all answers and evaluate them. It will also do practical checks where possible to verify whether the answers given match reality.
5. The evaluators will write an internal report which will be sent to the corresponding centres to get their feedback. A video conference³ could be organized to address the points of concern. Based on the feedback and possibly some additional checks a report will be formulated and sent to the CLARIN ERIC. An assessment report will summarize the results.
6. The quality of the services will be assessed regularly (in general once per year).

The role of CAC will be to ask questions and it is the role of the centres to answer them and to indicate how the answers can be verified. The CAC will not invent procedures, but use the provided material to come to statements.

The report to the ERIC is thus a description of a state at a certain moment and nothing that stays for ever. The centre can ask for a new assessment when it has improved on the points being mentioned as missing. The report will thus include references to missing points.

4. Frequently Asked Questions

What are CLARIN technical centres?

Centres of the type B, C and E:

- **B** is probably what most CLARIN centres will try to become. It is integrated into the infrastructure with all necessary building blocks: a stable repository with CMDI metadata, PIDs, AAI-access for protected resources, a classification of its licenses based on the PUB/ACA/RES system⁴, WSDL/WADL descriptions for web services (next to the web service core model CMDI files⁵).

² The questionnaire mentioned is CE-2013-0095 – B centre checklist, available at <http://hdl.handle.net/11372/DOC-78>. One question will be whether the centre has applied to get the CoreTrustSeal.

³ Only in special cases f2f meetings will be organized.

⁴ See <http://hdl.handle.net/11372/DOC-62> for details on the PUB/ACA/RES classification.

⁵ See <https://www.clarin.eu/cmd-core>

- An **E centre** offers a service to other CLARIN centres but is not a part of the national CLARIN consortium.
- The **C centre** is the bare minimum: a centre that has web-accessible resources that provides OLAC or CMDI metadata about these over OAI-PMH.

And the other types, what about them?

There are also K(nowledge) centres: these provide, rather than technical services or resources, consultancy.

If all my resources are publicly available, do I need to establish a Shibboleth Service Provider?

Strictly spoken: no. It is however advisable to build up some know-how on this in your centre, as it might prove useful at a later stage, by e.g. establishing a test Service Provider.

5. Checklist for B-centre requirements

The B-centre checklist is available at <http://hdl.handle.net/11372/DOC-78>