

Title Clarin Centre Types
Version 0.7
Author(s) Peter Wittenburg, Dieter Van Uytvanck, Thomas Zastrow, Lene Offersgaard
Date 2012-12-14
Status Approved 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:

1.1 Infrastructure Centres (Type A)

Task: Type A centres offer services that are relevant for the infrastructure as a whole and that need to be offered at a high level of commitment (stability, availability, persistence); in contrast to Type B they offer services that are used by other centres as well; in contrast to Type E they belong to a CLARIN ERIC member;

Examples: *joint metadata portal, data category registration service, schema registration services, web service orchestration engine, federated content search aggregator, ...*

Requirements: Type A 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.

1.2 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.

1.3 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.

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), give advice about CLARIN relevant matters such as standards (Type K) and also offer some infrastructure type of services (Type A). 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.

1.6 Recognized Centres (Type R)

Task: these centres offer resources and tools via standard web sites (or web services), but do not have funds (yet) to participate in the CLARIN infrastructure and cannot make commitments;

Requirements: Type R centres are not requested to fulfil the requirements mentioned in chapter 2.

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

¹ For CLARIN ERIC

2. Requirements for CLARIN Centres (A, B, E)

The following list of requirements only holds for centres of types A, 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 Data Seal of Approval or MOIMS-RAC approaches.
- (e) Centres need to offer component based metadata (CMDI) that make use of elements from accepted registries such as ISOcat 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 (A or 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. We foresee that it will take a while until all interested centres will achieve a fully CLARIN compliant state, therefore the evaluation process will associate a label (Gold, Silver, Bronze) with each centre: (1) **Gold** means that all requirements are functionally met. (2) **Silver** means that most essential criteria² are met, but that there is still work to be done. (3) **Bronze** means that the centre can participate, but that essential functions are missing.

² See <http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-78> for a list of the most essential criteria.

3. Centre Assessment Procedure

For all centres of types A, B and E 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 questionnaire³ 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 conference⁴ 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. A label “gold, silver, bronze” 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
- one of the labels “gold, silver, bronze”

The labels “gold, silver, bronze” need to be explained:

Gold means that all requirements are fulfilled with only minor missing points which can be improved easily by the centre without the need to change their basic architecture or including new components that may show erroneous behaviour and thus need to be assessed as part of a complex interacting set of components.

Silver means that the centre is pretty far on the way of meeting the requirements, but that one/two major aspects are missing.

³ The questionnaire mentioned is CE-2013-0095 – B centre checklist, available at <http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-78>. One question will be whether the centre has applied to get the Data Seal of Approval. Since this is a procedure taking some time, it probably means that only few centres will get the label “gold” at the beginning.

⁴ Only in special cases f2f meetings will be organized.

Bronze means that the centre is on its way of meeting the requirements, but that a few major aspects are missing.

For centres which are still working on some basics we will not give a label but just hand over the report. We expect centres, however, only approach the assessment committee if they see a chance to receive one of the three labels. The assessment committee is aware of the fact that we need practical experience to explain the labels in more detail.

4. Frequently Asked Questions

What are CLARIN technical centres?

Centres of the type A, B, C and E:

- An **A centre** is providing infrastructure services to other CLARIN centres. It could be combined with a B centre (= A+B centre).
- An **E centre** offers the same as an A centre (a service to other CLARIN centres) but is not a part of the national CLARIN consortium.
- **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 system⁵, WSDL/WADL descriptions for web services (next to the web service core model CMDI files⁶).
- 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 the K(nowledge) centres and the R(espected) centres:

- the K centre provides, rather than technical services or resources, consultancy.
- the R centre is the first step for becoming a C centre: offering a web-accessible resource or service, without the necessary metadata

Is an A centre automatically a B centre too?

Not necessarily. Consider e.g. a computing centre that participates in a national CLARIN consortium by providing certain services to other centres, like hosting a PID service or a Virtual Collection Registry. In practice however most A centres will also qualify as B centre. For the sake of clarity we will try to refer to them with the label **A+B centre**.

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 See <http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-62> for details on the PUB/ACA/RES classification.

⁶ See <http://www.isocat.org/clarin/ws/cmd-core/>

5. Checklist for B centre requirements

The following guidelines are meant as practical checks for the requirements mentioned earlier on in section 2. This document does not describe procedural aspects of checking in detail, it just describes the functionality which will be checked.

- **Centres need to offer useful services to the CLARIN community.** This needs to be indicated by:
 - short lists of services the existence and operation can be assessed via the web.
- **Centres need to adhere to the security guidelines, i.e. the servers need to have accepted certificates.** This needs to be indicated by:
 - proof of existence of an SSL certificate for all servers involved that provide a full trust chain
- **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.** This needs to be indicated by:
 - signed agreement with a national IDF or a trusted statement from the IDF
 - signed agreement with the CLARIN federation or a trusted statement from the SPF
 - demonstrate the working of a Shibboleth based login and authorization to access your service
 - demonstrate the working of a Shibboleth based login and authorization to access other services of the CLARIN federation
 - have the debug script `shib_test.pl` operational
- **Centres need to have a proper and clearly specified repository system and participate in a quality assessment procedure as proposed by the Data Seal of Approval or MOIMS-RAC approaches.** This needs to be indicated by⁷:
 - A short description of the major tasks and policies of the repository, references to these statements on the web and an indication of how they can be tested.
 - A short description of the architecture of the repository and its main procedures making it obvious that data and metadata are managed according to widely accepted norms (OAIS) and separately accessible in a stable and persistent manner.
 - Services that allow verifying the architecture claims and that allow accessing the metadata by searching/browsing and to access the data via the metadata records.
 - Indication what kind of quality assessment and improvement policies (and scripts) are in place.
 - A document that indicates the successful assessment through DSA or RAC or a document that indicates that a formal request for assessment through DSA or RAC has been submitted
- **Centres need to offer component based metadata (CMDI) that make use of elements from accepted registries such as ISOcat in accordance with the CLARIN agreements, i.e. metadata needs to be harvestable via OAI PMH.** This should be indicated by:
 - An indication how to look at the full metadata records (search/browse/inspection).
 - An indication of how metadata is generated and how CMDI compliance is being guaranteed.

⁷ What has been described to get the DSA or RAC assessment done, does not need to be provided twice.

- An indication indicating the existence and proper functioning of the OAI-PMH service provider function and how this can be assessed.
- ***Centres need to associate PIDs records according to the CLARIN agreements with their objects and add them to the metadata record.*** This should be indicated by:
 - An indication of the Handle assignment policies and procedures and ways to check that they are systematically applied.
 - An indication where to find Handle information in the metadata records and whether clicking on them will bring you to the data.
- ***Centres can choose to participate in the Federated Content Search with their collections by providing an SRU/CQL Endpoint.*** This should be indicated by:
 - An indication of the API implemented and how the functionality can be tested.
 - An indication which data can be contacted via this remote search function
- ***Each centre needs to make clear statements about their policy of offering data and services and their treatment of IPR issues.*** This should be indicated by:
 - official statements and agreement types that are being used for depositing, accessing and managing data

Advice: use the License metadata component to indicate the PUB/ACA/RES status of the resources on offer⁸.
- ***Each centre needs to make explicit statements to the CLARIN boards about its technological and funding support state and its perspectives in these respects.*** This should be indicated by:
 - official statements about funding type and duration
 - an indication about possible future funding support
- ***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.*** This should be indicated by:
 - indications about how users can or are embedding the services into their daily work

⁸ See http://catalog.clarin.eu/ds/ComponentRegistry?item=clarin.eu:cr1:c_1345561703649 or the license type data category: <http://www.isocat.org/rest/dc/5439>