Title	The Knowledge Sharing Infrastructure [KSI]
Version	8.0
Author(s)	SK
Date	11-11-2014
Status	For information
Distribution	GA
ID	CE-2013-0149

Introduction

In CLARIN we see at least four distinct groups of people:

1. Those who (want to or should be encouraged to) use the infrastructure (Users)
2. Those who populate it with data and services (Providers)
3. Those who build and operate the technical infrastructure (Operators)
4. Those who manage the infrastructure (Managers)

Each of these groups need knowledge and expertise to do their jobs, and they generate new knowledge and expertise as they go along. Contrary to big installations, where the knowledge and expertise can be brought together in one physical location known to all, and can easily be shared within and across the various groups, a distributed infrastructure such as CLARIN exists in dozens or maybe even hundreds of locations, distributed all over Europe, and the same applies to specific CLARIN knowledge and expertise.

The Knowledge Sharing Infrastructure [KSI]

The KSI will be our instrument to ensure that such knowledge and expertise will not exist as a fragmented collection of unconnected bits and pieces, but will be made accessible to the whole CLARIN community and to the Humanities and Social Sciences [HSS] research community at large.

Parallel to what we have for the data and services infrastructure the main building bricks of the KSI will be knowledge centres, which can be physical (i.e. concentrated in one physical place) or virtual (spread over various institutions or centres in CLARIN participant countries).

Knowledge centres may operate at the CLARIN ERIC level, in which case we call them K-centres, or at the local national level, in which we will call them L-centres. Both play a role in the distribution of knowledge and expertise, but whereas the activities of K-centres will be coordinated at the CLARIN ERIC level, and will have to meet specific quality and service level criteria, L-centres will be coordinated at the national level, and are free to organize their own services, as long as they commit themselves to adhering to CLARIN standards and principles.

Knowledge centres may target specific groups (e.g. the ones listed in the introduction), or cross- or sub-sections. They can be based on language ("the K-centre for the Danish language"), modality ("the K-centre for sign language"), domain ("the L-centre for Dutch historical texts"), technology ("the virtual K-centre for machine learning"), etc etc.

Knowledge centres may use specific instruments to make their knowledge and expertise available, both proactively (e.g. organizing courses, workshops, web courses, social media, mobility schemes) and reactively (e.g. help desks, FAQs, technical support on demand, portals).

K-centres vs L-centres

The distinction between K-centres and L-centres has been introduced to avoid confusion about the status of such centres and the expectations users may reasonably have. A typical characteristic of K-

centres is that they have to provide their services to the CLARIN community at large, which normally implies that services should be provided at least in English and that they should have a high degree of availability.

Not every centre will be in a position to do this, not only because of the potential additional workload, but also because of the language. Apart from e.g. in linguistics or computational linguistics, where English seems to be the normal vehicle for communication, local languages may dominate in many subareas of the humanities and social sciences, and it would not always be obvious how to open up a course in e.g. processing of mediaeval Polish poetry, taught in Polish, to people who don't speak the Polish language.

Sustainability and critical mass are other important factors for K-services: activities that rely on the expertise of a single expert may not be as sustainable as those that are supported by a whole team of experts. Rather than to exclude centres below critical mass from being part of the CLARIN Knowledge Sharing Infrastructure they can be listed as L-centres, which means that they are recognized as contributions to CLARIN, but that the scope of their services may be limited in a number of ways, to be determined by the centres themselves or their national consortia.

K- or L-centres can also be contributed by countries or regions not covered by CLARIN ERIC, or by individual institutions.

Step 1: A first inventory

At this moment we don't know yet how many centres in CLARIN ERIC are in a position to act as K- or L-centres, and what types and volumes of knowledge sharing activities they offer. This may differ from country to country, depending on their funding conditions and their ambitions. In a first round we have made an inventory of what is being offered, both in terms of centres and in terms of activities; this inventory will be put on-line in a searchable registry and will be continuously updated.

From the CLARIN Agreements and earlier CLARIN documents we have extracted a whole battery of possible instruments for knowledge sharing. They constitute the main vehicles to share knowledge across the CLARIN provider and user communities. Most of them are already in use in CLARIN, but do not necessarily exist for all languages and for all topics.

The instruments come in types: some are passive (e.g. websites, reference manuals, FAQs), some are reactive (helpdesks, consultancy), some are proactive (tutorials, workshops), and some are a mix (mailing lists, forums). The choice of instruments for the CLARIN KSI depends on many factors, such as topic, audience, language and available human and financial resources.

See Annex 1 and 2 for an unstructured list of instruments and topics.

The task of the KSI committee is to identify an optimal constellation of quadruples [topic, audience, instrument, language] that will ensure the most effective, efficient and sustainable flow of knowledge and expertise between and within the participants in CLARIN (providers and users).
An analysis of what exists will help to identify gaps as well as potential synergies between what exists in various countries.

This optimal constellation should also form an important basis for our planning for HORIZON2020, as this will allow us to apply for special EC funding to strengthen the CLARIN Knowledge Sharing Infrastructure in areas where we feel it is lacking or not sufficiently developed.

Step 2: Identifying K- and L-centres

Background

One of the building bricks of the CLARIN Knowledge Sharing Infrastructure are knowledge centres. Their role is to provide a place (physical or virtual) where people can get access to knowledge and expertise in specific areas.

They can be

- physical: a group of people in one physical location or institute
  or virtual: people spread over various locations in one or more countries
- local: services restricted to own national consortium or sub-communities
  or global to CLARIN ERIC: services to all
- thematic: specialized in one topic
  or broad: covering a variety of topics
- proactive: e.g. by offering face-to-face or web-based training courses
  or reactive: e.g. helpdesks, advice, consultancy), or both.

The instruments they use may vary. Knowledge centres may or may not offer technical infrastructure or data services as well, and may be located in or outside CLARIN ERIC countries.

Knowledge centres can apply for recognition as CLARIN K- or L-centres if they live up to a number of service and quality criteria.

K-centres

Knowledge centres serving the CLARIN community with knowledge services can get formal recognition **as CLARIN K-centres** if they satisfy the following criteria:

1. They serve the CLARIN community at large (as opposed to operating just at the national level), and offer their services at least in English. [confirm international and English]
2. At least part of their knowledge services are reactive (i.e. they react to requests from customers, as opposed to e.g. just offering a website with useful information). [confirm reactive services]
3. They respond to queries and requests within 2 working days (at least acknowledgement and info about further procedure). [confirm response time]
4. In their application and on their website they provide a clear description of the scope of their services (e.g. helpdesk, hosting, technical assistance, courses). [max 1 para text for scope]
5. In their application and on their website they provide a clear description of the thematic areas they cover. [max 1 para text for areas]
6. In their application and on their website they provide a clear statement of the audiences they serve (e.g. humanities scholars, infrastructure operators, …) [max 1 para text for audience]
7. The knowledge sharing activities they offer (proactive and reactive) should be registered in the CLARIN Knowledge Sharing Registry (see comment 3) and kept up-to-date. [confirm willingness to register]
8. They should have sufficient critical mass consisting of at least 1 chair or 3 fte permanent staff (see comment 1 below), which may be distributed over different physical locations, not necessarily in the same country. [confirm critical mass] , [indicate mono-/multi-site]
9. In their application for recognition they should list 5 major publications over the last 3 years in the thematic areas covered. [list publications]
10. If the centre is virtual (i.e. spread over different institutions) the application has to specify the participating institutions and to provide a description of the way the collaboration between the centres is organized to ensure that they can act as one knowledge centre. [list participants], [1 para text for organization]

L-centres
Other knowledge centres can apply for recognition as a **CLARIN L-centre**. For recognition as an L-centre only criteria 4, 5, 6 and 7 should be satisfied.

Application process
Applications (see comment 2) to be recognized as a CLARIN K-centre or L-centre should be submitted by the head of the institution(s) to the CLARIN ERIC Office. They will be reviewed by the CLARIN Knowledge Sharing Infrastructure Committee, which will send its recommendations to the Board of Directors. The recognition will be valid for two years.

Comments
1. Critical mass: In some countries (and Denmark seems to be one of them) the existence of a chair in an area has no predictive value for the critical mass, and that fte permanent staff would count as a better measure. On the other hand we know that in German universities permanent positions are only available for professors (and maybe also cleaning staff), and that the average professor comes with a cloud of PhD students and Postdocs on temporary contracts. Rather than to embark on a bureaucratic exercise to develop objective and universal criteria for critical mass we hope that the two criteria mentioned will lead to a common interpretation of what is meant by critical mass: properly anchored in the organization, and not based on individual hobby horses.
2. Application forms: At this moment there is no template for the application yet, but we will provide one as soon as the first applications have been processed and we know which problems have been encountered when completing them (which will inevitably happen). You can just use the 10 points above as a check list and provide a numbered list of responses.
3. KS registry: At this moment the registry already contains the KSI contributions that you have provided through the CLARIN Agreement. We will soon make it editable so that you can easily add, delete and update your own items, but we don't expect this to be up and running before the Annual CLARIN Conference 2014 in Soesterberg .

Concluding remarks
A research infrastructure cannot exist without a closely connected infrastructure that ensures a proper flow of knowledge and expertise between all categories participants, so that new research can build on work done by others. In a single-site physical infrastructure this is relatively easy to accomplish, and people will know where to find the information, and if necessary the training they need to optimally use the infrastructure.

In a widely distributed data infrastructure, where users have access to the infrastructure from all parts of Europe, across national and language barriers a special effort is required in order to ensure access to information and expertise needed to use the infrastructure. In the case of CLARIN we can make use of the fact that many CLARIN centres and institutes are based in universities, where there is a strong tradition of knowledge transfer. The challenge for CLARIN is to make optimal use of the knowledge and the knowledge transfer mechanisms that already exist, and to try to make sure that a much broader community can benefit from them, so that all thresholds for using the infrastructure in order to enhance and innovate European research in the humanities and social sciences will disappear.

ANNEX 1: Instruments

From the CLARIN Agreements and earlier CLARIN documents we have extracted a whole battery of possible instruments for knowledge sharing.

Below we list the types encountered, in alphabetic order.

instruments

automated helpdesk
awareness seminars
background documents
best practice documents
books
conferences
consultation
cooperation with associations etc
course modules
curation service
curriculum development
customisation of resources
directory of expertise
directory of players
documentation
ec projects (e.g. marie curie)
education
faq
forums
grants
help desk
hosting
interviews
journals
ksi network
mailing lists
manuals
movies
newsletters
panel sessions
personal advice
physical centres of expertise
portal
questionnaires
scientific papers
short guides
showcases
social media
summer schools
training
tutorials
virtual centres of expertise
web presentation of knowledge
web courses and lectures
websites
wiki
workshops

ANNEX 2: Topics

In the CAs we found a large number of different topics on which expertise is offered, using various instruments, and for a variety of languages. We list them here in alphabetic order.

topics

aai
annotation tools
archiving
audio data
chunking
cmdi
collecting data from web
concept registries
conversion
corpora in education
crowdsourcing
curation
digital humanities
extraction of lexical semantic networks
field work
formal semantics
gis
human-human comms
human-machine comms
individual languages
information extraction
information retrieval
ipr
isocat
knowledge management
language documentation
large corpora
legal issues
lexicography
licenses
linking resources
linking web services
localisation to own language
longitudinal studies
machine learning
metadata infrastructure
multilingual corpora
multimodal
multi-word expressions
named entity recognition
nlp analysis tools
ontologies
orthography
parsing
pid
porting between languages
registers
repositories
segmentation
semantic networks
shallow parsing
speech
standards
statistics
tei integrator
text
text mining
topic extraction
translation
treebanks
validation
video processing
visualization
web services
wordnets
workflows
wsd