The Knowledge Sharing Infrastructure
KSI

Steven Krauwer

Why a KSI?

- Building or using a complex installation requires specialized skills and expertise. CLARIN is no exception.
- CLARIN is populated with and providing services to a broad variety of people, with different types of knowledge, different knowledge level, and different knowledge needs
- Contrary to big physical installations where all players are present in one single place and can easily communicate, CLARIN is widely distributed, and groups have little knowledge of what the others do, know or need, without organised channels for knowledge transfer
- Negative effects: sub-optimal use of the infrastructure, under-exploitation of accumulated knowledge, risk to lag behind, duplication of efforts, reinvention of (square) wheels

Different dimensions

Different types of people:

- Those who (want to) use the infrastructure to do research
- Those who populate it with data and services
- Those who build and operate (parts of) the infrastructure
- Those who manage (parts of) the infrastructure

Different types of knowledge sharing:

- Knowledge sharing can be
  - Vertical: teacher to pupil
  - Horizontal: peer to peer
- Knowledge sharing can be
  - Proactive (e.g. tutorials, workshops)
  - Reactive (e.g. helpdesks)

Building bricks of the KSI

- Knowledge centres
- Knowledge sharing instruments
- Knowledge providers
- Knowledge consumers

Knowledge Centres

- Can be physical: group of people in one physical location
- Or virtual: people spread over various locations

- Can be local: services restricted to own national consortium
- Or CLARIN ERIC: services to all

- Like data and service centres knowledge centres come in types:
  - K-centres, officially recognized as CLARIN ERIC centres, satisfying specific criteria
  - Other centres: L-centres

Criteria for K-centres

First suggestions

- Accessibility: Provide services in English
- Responsiveness: Respond within 2 working days throughout the year
- Clear statement of scope of services offered (e.g helpdesk, hosting,
- Clear statement of thematic areas covered
- Critical mass and sustainability: at least chair or 3 fte permanent staff
- Quality: list 5 major publications over last 3 years

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

Instruments

- Wealth of possibilities, but can't do them all
- For central KSI operations: focus on a few, and do them well
- For national operations: free to choose, but let's keep each other informed of what works and what doesn't

Knowledge providers + consumers

Every single one of us is sitting on knowledge others may not have

Important that we are prepared to share it!

Topics mentioned in CAs

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

Next steps

- What we have:
  - Collect information on what is already there (via CLARIN Agreements)
  - Bring together in searchable archive
- What we need:
  - Collect information on what is needed
  - Identify gaps
- Propose solutions to fill the gaps, e.g.
  - Invite national consortia to contribute
  - Take central initiatives
  - Find funding from other sources, e.g. EC projects

Next steps

- Identify candidates for K-centres
- Formulate criteria for K-centres
- Go through K-centre assessment

Instruments & topics

- See http://www-sk.let.uu.nl/v/CE-2013-0149-KSI-v5.0.pdf
- Question to you:
  - What are the best instruments to use: what works and what doesn't
  - What are the priority topics for K-centres
  - What is the best strategy, e.g. start from an instrument (say summer school) and find appropriate topics & audiences
  - Or start from a topic and see which instruments we can use to support this topic for specific audiences
  - Or start from an audience and find appropriate topics and instruments

Concluding remarks

- We want to build the KSI, using existing resources wherever possible
- We want to find ways to fill any gaps we identify
- We have a core committee consisting of Bente Maegaard, Koenraad De Smedt, Martin Wynne, Kiril Simov, Gerrit Bloothooft, Steven Krauwer
- We invite others to join the committee to contribute ideas