

CLARIN

Common Language Resources and Technology Infrastructure



Persistent Identifier Service for CLARIN and others?

some general statements



- necessity of indirection steps is an issue for ongoing debate
 - also ongoing debate about suggestions for solving this
 - as we have seen there are open ends, which CLARIN may need to discuss broadly
 - what is a persistent object and what is its identity
 - what is the appropriate granularity (lexicon or lexical attribute, etc)
 - do we share all the same opinions in CLARIN
 - we need community guidelines
 - but one step after the next please
 - we are happy wrt PIDs that there is a service provider based on a stable machinery
 - if there is another service provider meeting the essential requirements fine for us
-

CLARIN requirements



- PID framework must support all contexts of references - actionable
 - compliant with standards (but what does that mean)
 - no forced semantics in PIDs
 - support for fragment addressing
 - support of multiple copies
 - support for additional data such as metadata
 - performance, robustness, availability of the service
 - secure and sufficiently easy management
 - suitable API
 - independence from any single service provider
 - no single point of failure
 - feasible cost model for millions of PIDs (\ll 1 cent)
-

Ideas for additional services



- want to check authenticity before copying etc (MD5 field)
- want to add citation data info extracted from metadata records
- want to solve the problem of having several centers safely sharing PID information
- want a hook to access permission information
- proper monitoring services