Fedora Commons in the CLARIN-D infrastructure

Thomas Eckart & Felix Helfer
NLP Group
Institut of Computer Science, Leipzig University, Germany
{teckart, helper}@informatik.uni-leipzig.de
• CLARIN – Common Language Resources and Technology Infrastructure

• A web and centres-based research infrastructure for the social sciences and humanities; Part of ESFRI (European Strategy Forum on Research Infrastructures)

• Linguistic data, tools and services are provided in an integrated, interoperable and scalable infrastructure

• CLARIN-D – National consortium in Germany

• Website: https://clarin.eu & https://de.clarin.eu
Distributed infrastructure with

- standardized interfaces (for metadata, data and tools)
- processing layers and
- communication protocols

Certification every 2 years, including:

- Data Seal of Approval / Core Trust Seal
- OAI-PMH endpoint as central interface
• Requirements for a repository solution:
  • "Technical" platform
  • Long-term support
  • OAI-PMH interface for non-DC metadata
• CLARIN-D (10 centres) started around 2011
  • Majority: Fedora Common 3.x + ProAI
• Different strategies to deal with major 4.x changes:
  • Migration to 4.x → massive effort and problems with OAI-PMH
  • Migration to DSpace
  • Creation of home-brew solutions
  • Majority: still stuck with 3.x, (attempts to use migration tool)

• Current state:
  • Contact with Lyrasis → virtual meeting about the issue
  • Creation of an interest group in CLARIN-EU
  • Creation of a requirements document