



Clarin.dk - DIGHUMLAB

Lene Offersgaard
University of Copenhagen
leneo@hum.ku.dk





Clarin.dk:

- Part of DIGHUMLAB:
 - National infrastructure initiative, launched 2012
 - Partners now: 4 Danish Universities, will be extended
- Repository for **resources** and **tools** for the humanities
- All metadata are publicly accessible
- Single sign-on through WAYF (Where Are You From) using shibboleth
 - All research institutions in Denmark can use WAYF
- Access: Resources are public, academic or restricted
 - Licenses are currently clarin.dk-specific, based on CLARIN work in preparatory phase
 - Data provider accepts license when depositing
- PID in metadata (locally defined)
- Version control of resources
- Repository implemented using eSciDoc, based on Fedora repository system



Resource type	Count	Metadata formats
Texts, period: 1500-2010 some with facsimile	46200	TEIP5(DKCLARIN), DC, OLAC, CMDI
Text annotations	143096	TEIP5(DKCLARIN), DC, OLAC, CMDI
Audio	37	IMDI, DC, OLAC
Video	80	IMDI, DC, OLAC
Annotations of audio or video	45	IMDI, DC, OLAC
Lexica	3	TEIP5(DKCLARIN-LEX)
Tools, downloadable	5	DC, OLAC, CMDI

- Metadata formats determined in collaboration with content providers
- OAI-PMH harvesting: currently oai_dc and OLAC, to be extended with CMDI
- Resources are being deposited, more texts...
- More resource types can be added



Services

- Metadata search, all metadata are public
- Download of resources, if user accepts license!
- Integrated tools which can process resources in repository, currently text tools
- Workflow service: user selects resources and goal, workflow assistant plans which tools to use to fulfil the users goal
- Deposit of resources
- Content search facilities, to be extended
- OAI-PMH
- to be extended ...
 - Search in corpora
 - Combined search in metadata, text and annotations

More repositories than clarin.dk might be included from DK

