Trusted data services in LINDAT/CLARIN

Pavel Straňák



Open Access

- Open Access to data just starting
 - H2020 OA Pilot, many things unclear
- License requirement
 - Without an explicit license, nothing is allowed
- Great idea, but many obstacles



Open Access²

- Challenges:
 - IP Law. Especially for NLP (language data, i.e. texts, audio, video) a big challenge.
 - Publishers want to strenghten the copyright and not allow research exceptions or Fair Use
 - Financing: OA data and tools are free. We love it.
 - But how do we atract commercial partners now?



LINDAT/CLARIN Data Services and Challenges¹

- Require as little as possible.
 - But require all that is needed.
- Maximum Open Access.
- All services requiring AAI use Shibboleth (national identity federations, Clarin SPF, EduGAIN)
 - Use your own accout you trust
 - Lets make it an account we trust too



LINDAT/CLARIN Data Services and Challenges²

- Repository
- Other Applications and Web Services (REST)
 - Kontext (search texts)
 - Tree Query (search annotation of texts)
 - various analysers
- Link all services (get, search, process, store data)
- Trust all of these. Repository is just one part.



LINDAT/CLARIN Data Repository

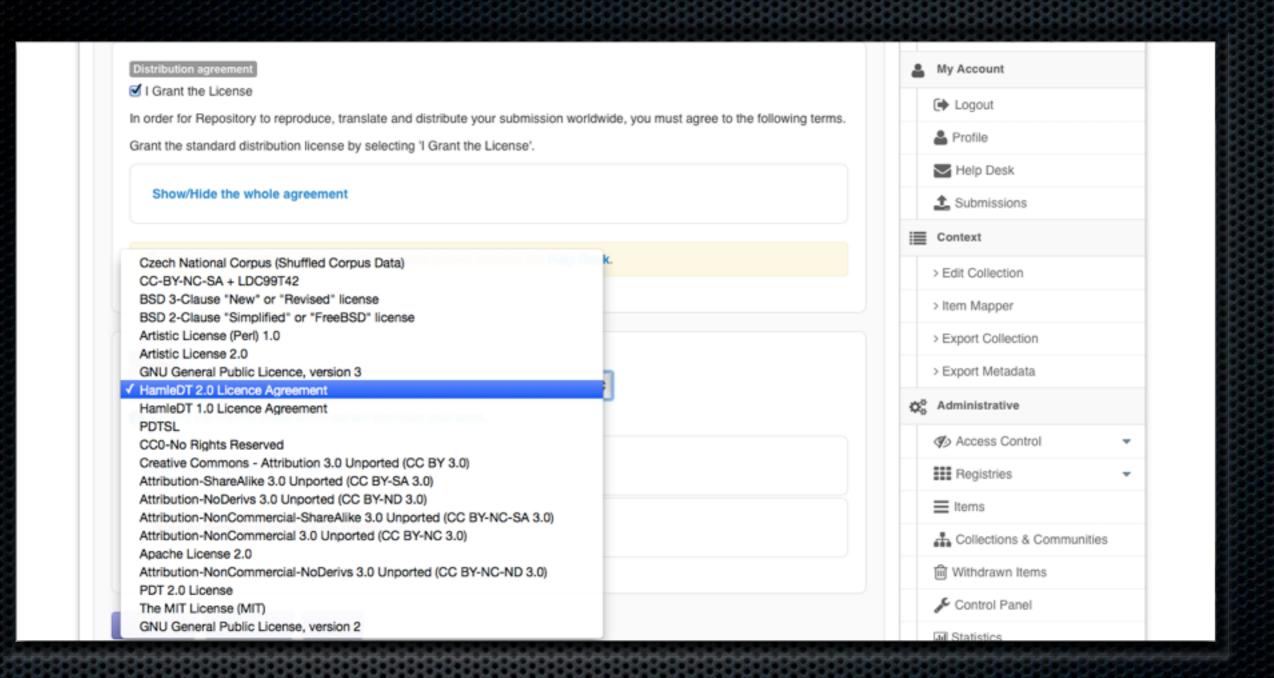
- Standard, robust system
 - If something is missing, add it. Licensing
- Nice to use (simple, fast, FAQ, ...)
- Make it actually useful
 - badges, statistics, reports, citations, ...
- Underlying robustness needed, but not the goal



LINDAT/Clarin Repository Licensing Framework

- License definition
 - The license text
 - Attributes
- Signing a license: a triple (Dataset, License, User) +
 extended attributes required by the license
- A submitter can define a new (custom) license and set its attributes
 - Available to all future submitters



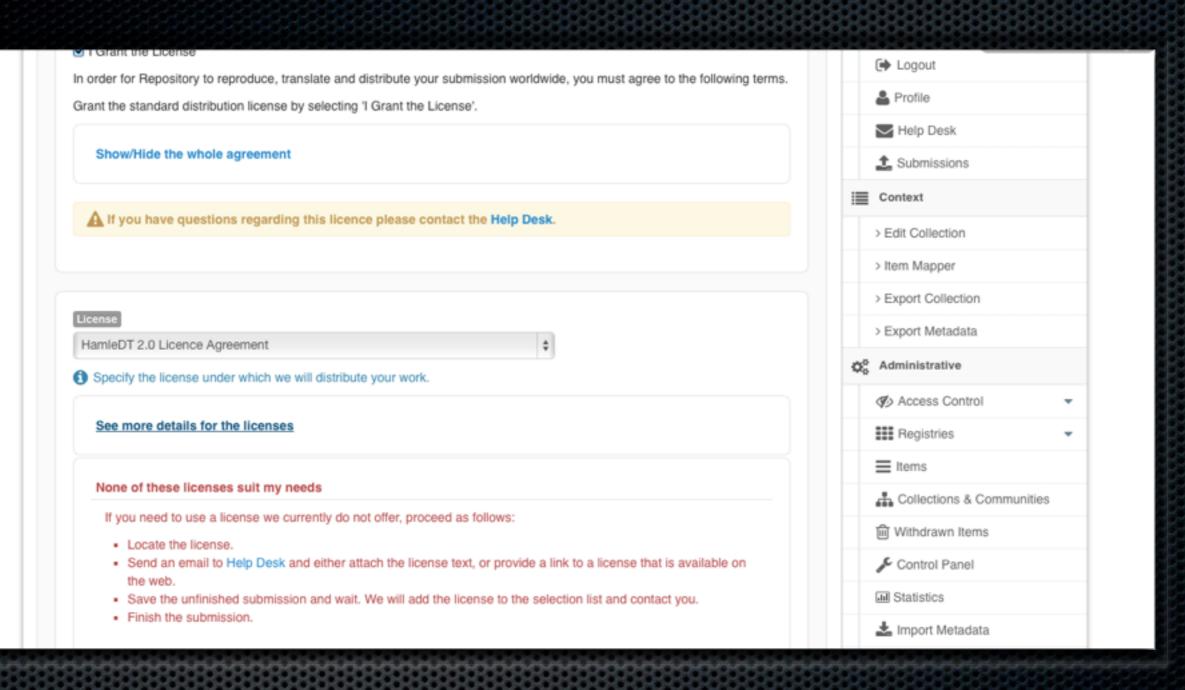


2 agreements Distribution Agreement + License

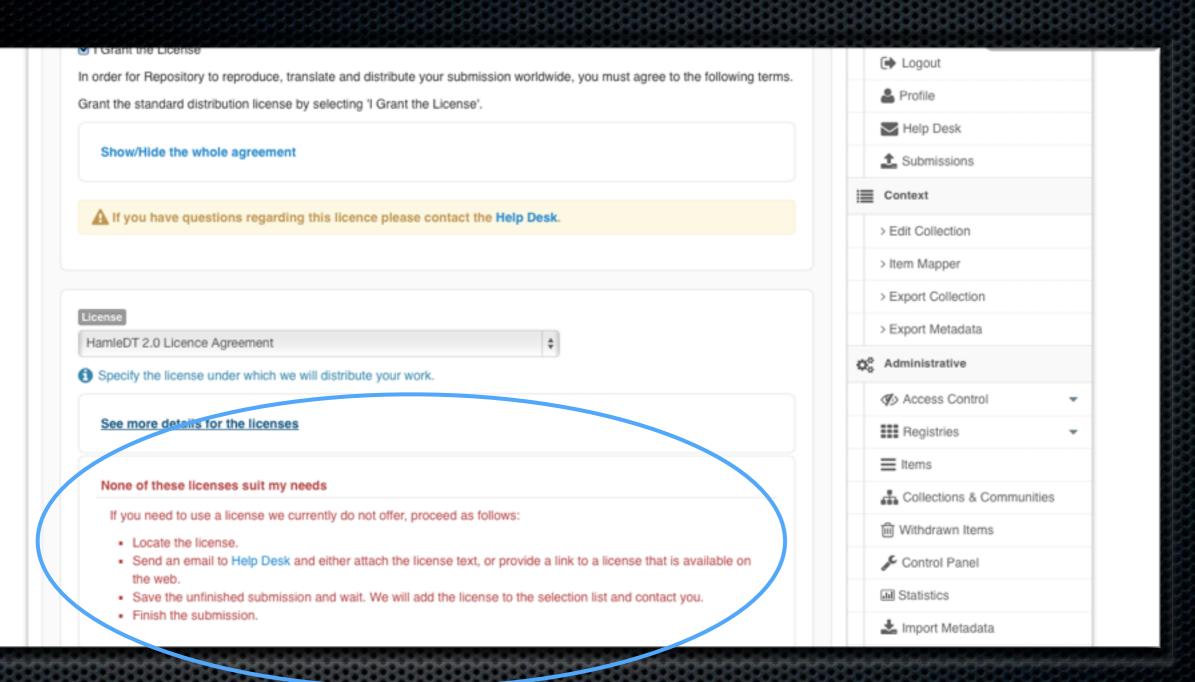
License attributes

- Requires signing? (Yes / No)
 - Each time data is accessed?
- Requires additional user attributes? (form)
- Basic category: Free, Academic, Restricted
- Other standard conditions:
 - BY, NC, ND, SA, ReD ... logos

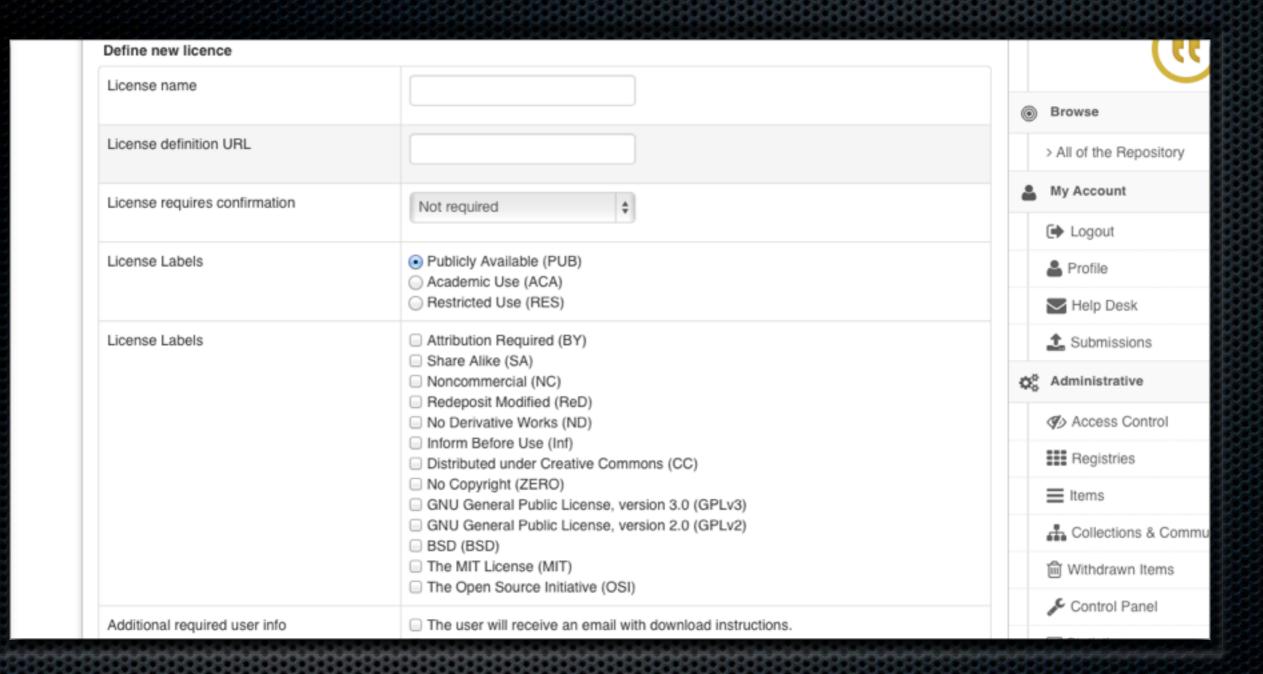




Defining a license Users need to ask admins



Defining a license Users need to ask admins



Defining a license

Administrative interface

Certification¹

- DSA certification: Repository only
 - Not very hard
 - Still a bit problematic
 - Very useful in the end
- Clarin B-Centre certification
 - Clarin internal, technical requirements. Hard requirements, but simple and short forms.



Certification² – Clarin

- Whole centre, not just one service
 - Makes great sense
- Serves to build trust in the whole Clarin:
 - All centres are good and robust, so
 - Clarin becomes a mark of quality and trust (hopefuly)
- Mostly technical, DSA one point of one point (repository)

