

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 account 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 $\langle \text{Dataset}, \text{License}, \text{User} \rangle +$ extended attributes required by the license
- ✦ A submitter can define a new (custom) license and set its attributes
 - ✦ Available to all future submitters

Distribution agreement

I Grant the License

In order for Repository to reproduce, translate and distribute your submission worldwide, you must agree to the following terms.
Grant the standard distribution license by selecting 'I Grant the License'.

[Show/Hide the whole agreement](#)

- Czech National Corpus (Shuffled Corpus Data)
- CC-BY-NC-SA + LDC99T42
- BSD 3-Clause "New" or "Revised" license
- BSD 2-Clause "Simplified" or "FreeBSD" license
- Artistic License (Perl) 1.0
- Artistic License 2.0
- GNU General Public Licence, version 3
- HamleDT 2.0 Licence Agreement
- HamleDT 1.0 Licence Agreement
- PDTSL
- CC0-No Rights Reserved
- Creative Commons - Attribution 3.0 Unported (CC BY 3.0)
- Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)
- Attribution-NoDerivs 3.0 Unported (CC BY-ND 3.0)
- Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0)
- Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0)
- Apache License 2.0
- Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)
- PDT 2.0 License
- The MIT License (MIT)
- GNU General Public License, version 2

My Account

- Logout
- Profile
- Help Desk
- Submissions

Context

- > Edit Collection
- > Item Mapper
- > Export Collection
- > Export Metadata

Administrative

- Access Control
- Registries
- Items
- Collections & Communities
- Withdrawn Items
- Control Panel
- Statistics

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

Grant the License

In order for Repository to reproduce, translate and distribute your submission worldwide, you must agree to the following terms.
Grant the standard distribution license by selecting 'I Grant the License'.

[Show/Hide the whole agreement](#)

⚠ If you have questions regarding this licence please contact the [Help Desk](#).

License

HamleDT 2.0 Licence Agreement

i Specify the license under which we will distribute your work.

[See more details for the licenses](#)

None of these licenses suit my needs

If you need to use a license we currently do not offer, proceed as follows:

- Locate the license.
- Send an email to [Help Desk](#) and either attach the license text, or provide a link to a license that is available on the web.
- Save the unfinished submission and wait. We will add the license to the selection list and contact you.
- Finish the submission.

Logout

Profile

Help Desk

Submissions

Context

- > Edit Collection
- > Item Mapper
- > Export Collection
- > Export Metadata


Administrative














- Access Control
- Registries
- Items
- Collections & Communities
- Withdrawn Items
- Control Panel
- Statistics
- Import Metadata

Defining a license

Users need to ask admins

Define new licence	
License name	<input type="text"/>
License definition URL	<input type="text"/>
License requires confirmation	Not required
License Labels	<input checked="" type="radio"/> Publicly Available (PUB) <input type="radio"/> Academic Use (ACA) <input type="radio"/> Restricted Use (RES)
License Labels	<input type="checkbox"/> Attribution Required (BY) <input type="checkbox"/> Share Alike (SA) <input type="checkbox"/> Noncommercial (NC) <input type="checkbox"/> Redeposit Modified (ReD) <input type="checkbox"/> No Derivative Works (ND) <input type="checkbox"/> Inform Before Use (Inf) <input type="checkbox"/> Distributed under Creative Commons (CC) <input type="checkbox"/> No Copyright (ZERO) <input type="checkbox"/> GNU General Public License, version 3.0 (GPLv3) <input type="checkbox"/> GNU General Public License, version 2.0 (GPLv2) <input type="checkbox"/> BSD (BSD) <input type="checkbox"/> The MIT License (MIT) <input type="checkbox"/> The Open Source Initiative (OSI)
Additional required user info	<input type="checkbox"/> The user will receive an email with download instructions.



-  Browse
- > All of the Repository
-  My Account
-  Logout
-  Profile
-  Help Desk
-  Submissions
-  Administrative
-  Access Control
-  Registries
-  Items
-  Collections & Commu
-  Withdrawn Items
-  Control Panel

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 (hopefully)
- ✦ Mostly technical, DSA one point of one point (repository)