

# Legal issues in LRT development and use in LINDAT/Clarin (Czechia)

Pavel Straňák

# Czech IP Law

- ✦ No generic research exception
- ✦ Citation exception
- ✦ Exception for use in education,
  - ✦ but not research; even by the same people on the same topic

# 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- Withdrawn Items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

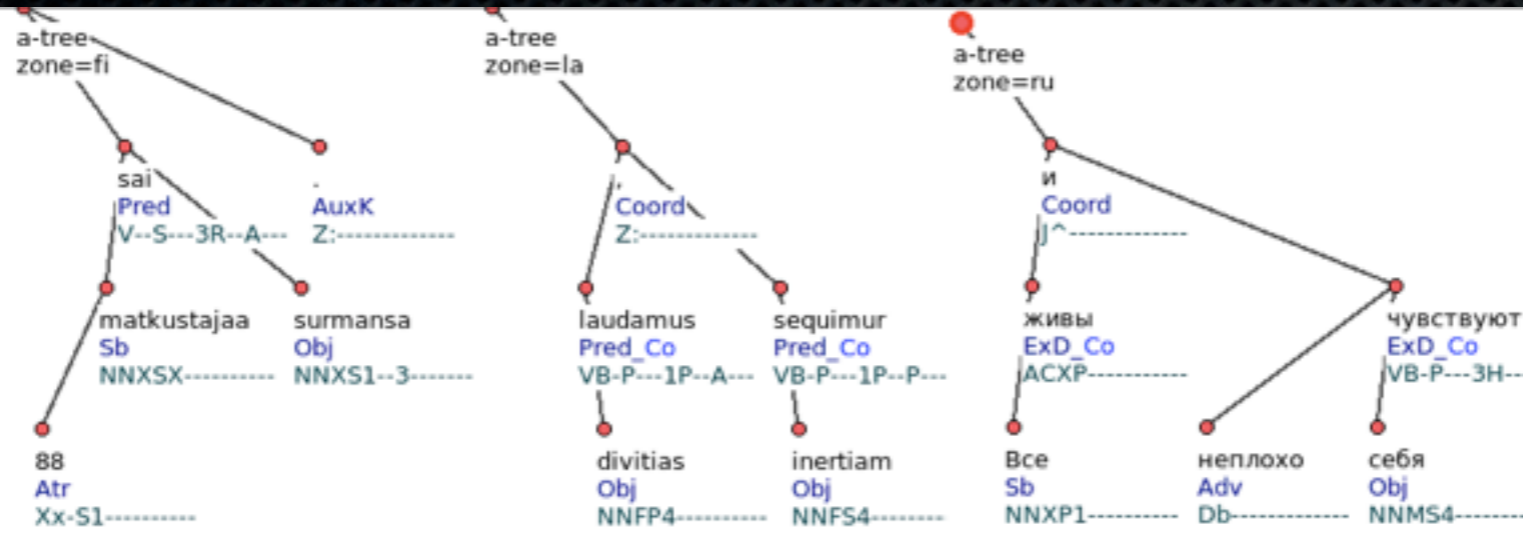
# Choosing a License – Problems

- ✦ Submitters don't know, what they are allowed to do with their data
  - ✦ They are often even not sure they obtained it legally
- ✦ Licenses are not trivial to read and understand (GPLs)
  - ✦ There are so many
  - ✦ Data vs. Software
- ✦ Ownership: the individual? university? institute?
  - ✦ Individual as long as at the university or institute?



# Choosing a License – TODO

- ✦ Related efforts:
  - ✦ EUDAT – B2SHARE work
  - ✦ “Ville’s license chooser”
  - ✦ Clarin Common Knowledge Base
  - ✦ META-SHARE Licensing Wizard
- ✦ It might be useful to join efforts



(wiki)



# HamleDT

HArmonized Multi-LanguagE Dependency Treebank  
 (30 treebanks, harmonised annotation)

Dear colleagues,

we are happy to announce that we just released HamleDT (harmonized multilanguage dependency treebank) version 2.0, now with 30 languages and two central annotation styles (Prague and Stanford). If you do not know HamleDT yet, you should visit

<http://ufal.mff.cuni.cz/hamledt>

If you want to access the entire dataset (including the data that we cannot make publicly available), you will find it at our network at

```
/net/projects/tectomt_shared/data/archive/hamledt/  
2.0_2014-05-24_treex-r12700
```

Enjoy!

Dan



Dear colleagues,

we are happy to announce that we just released HamleDT (harmonized multilanguage dependency treebank) version 2.0, now with 30 languages and two central annotation styles (Prague and Stanford). If you do not know HamleDT yet, you should visit

<http://ufal.mff.cuni.cz/hamledt>

If you want to access the entire dataset (**including the data that we cannot make publicly available**), you will find it at our network at

`/net/projects/tectomt_shared/data/archive/hamledt/  
2.0_2014-05-24_treex-r12700`

Enjoy!

Dan



Dear colleagues,

we are happy to announce that we just released HamleDT (harmonized multilanguage dependency treebank) version 2.0, now with 30 languages and two central annotation styles (Prague and Stanford). If you do not know HamleDT yet, you should visit

<http://ufal.mff.cuni.cz/hamledt>

**If you want to access the entire dataset** (including the data that we cannot make publicly available), you will find it at our network at

`/net/projects/tectomt_shared/data/archive/hamledt/  
2.0_2014-05-24_treex-r12700`

Enjoy!

Dan



Dear colleagues,

we are happy to announce that we just released HamleDT (harmonized multilanguage dependency treebank) version 2.0, now with 30 languages and two central annotation styles (Prague and Stanford). If you do not know HamleDT yet, you should visit

<http://ufal.mff.cuni.cz/hamledt>

**If you want to access the entire dataset** (including the data that we cannot make publicly available), **you will find it at our network** at

`/net/projects/tectomt_shared/data/archive/hamledt/  
2.0_2014-05-24_treex-r12700`

Enjoy!

Dan



# Web Apps and Services

- ✦ **PML-TQ**: Treebank queries for approx. 30 treebanks
- ✦ **Treex::Web**: dozens of tools to process your data
  - ✦ Uses a lot of stat. models trained on restricted data
- ✦ Corpus query service for all LINDAT/Clarin corpora
- ✦ ...
- ✦ Can we make these all public, for anybody to use?

# LINDAT/Clarin – Conclusions

- ✦ License signing implemented:
  - ✦ Allows to download many more datasets
  - ✦ Flexible custom licenses (PCEDT, HamleDT, etc.)
- ✦ Severe problems remain:
  - ✦ Core uncertainty about research data legal status
  - ✦ Choosing a license is not trivial, neither is observing
  - ✦ What services can we offer over what data?
- ✦ We need a broad research exception or Fair Use