

# Data Repositories

From the Perspective of The National Repository Platform

PAVEL STRAŇÁK

National Czech  
Programme



# What is a Repository

- Library (or publishing) model
  - a book you give to a library is taken care of
  - not yours to modify as you please, any more
- Software systems for preservation, sharing and tracking changes
  - different thing than VCS systems like Git (and their extensions like Github) – but some similarities
- Preservation
- Metadata (data description) + Search
- Persistence + tracking changes („provenance record“)
- View + citations
- Linking with the outside world

# An Ideal Repository?

- General principles and ideas
  - Demonstrated (mostly) on [lindat.cz/repository](https://lindat.cz/repository)
- Not all repositories have this all
- But some have more (not mentioned here):
  - good visualisations
  - CRIS features („research graph“)
- Some have less, in various degrees:
  - user uploads?
  - limited permissible formats?
  - licensing for non-open data?
- For NRP we want more, not less

# Safe Preservation

## Upload and don't worry

### How to Deposit

Only authenticated users can deposit items. If you cannot find your home organisation in the Login dialog list of organisations then register at [clarin.eu](http://clarin.eu) and authenticate using "clarin.eu website account". In case you cannot use any authentication method above or if you encounter a problem, do not hesitate to contact our [Help Desk](#) and we can create a local account for you.

### Step 1: Login

To start a new submission you have to login first. Click Login under My Account in the right menu panel.

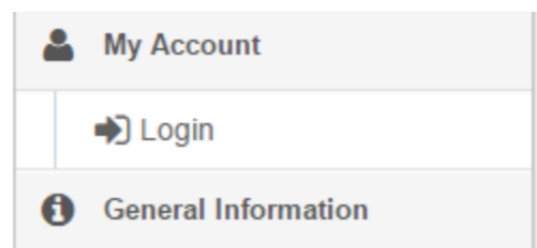


Fig1. Menu Login

### Step 2: Starting a new submission

Now you have a new menu item 'Submissions' under My Account. Click on Submissions to go to the Submissions screen.

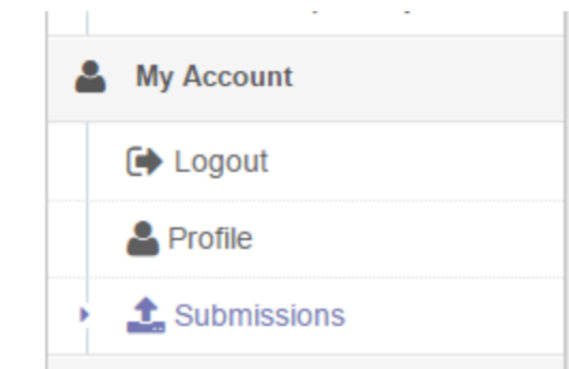
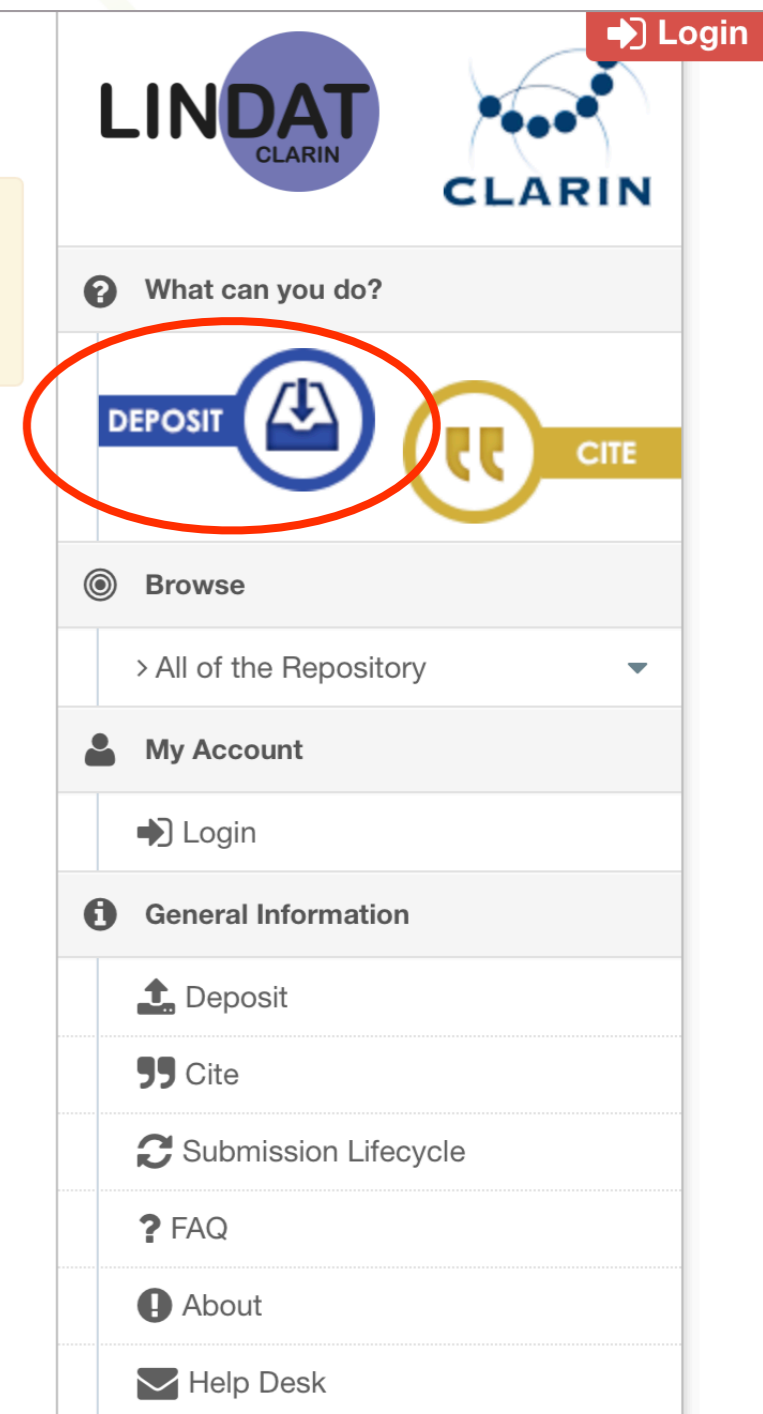


Fig2. Menu Submissions





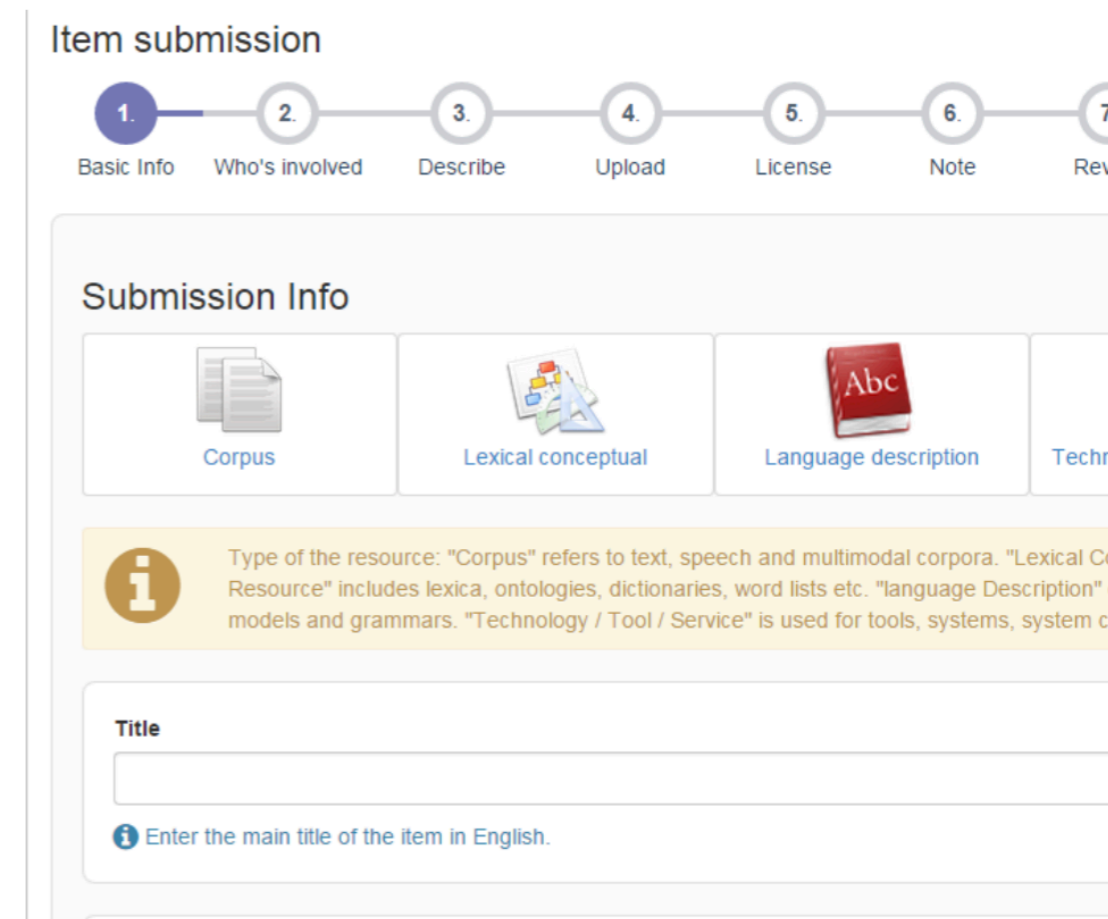
# Deposition Guide

## step-by-step description

1. login
2. fill-in metadata
3. drag&drop data
4. select a license
5. submit

## Step 3: Select type of your submission

You have initiated a new workflow item. In the next few steps you will provide the details about the item, select the type of the resource you are about to submit.



Item submission

1. Basic Info 2. Who's involved 3. Describe 4. Upload 5. License 6. Note 7. Review

Submission Info

Corpus Lexical conceptual Language description Technology / Tool / Service

Type of the resource: "Corpus" refers to text, speech and multimodal corpora. "Lexical Conceptual Resource" includes lexica, ontologies, dictionaries, word lists etc. "language Description" includes models and grammars. "Technology / Tool / Service" is used for tools, systems, system components.

Title

Enter the main title of the item in English.

Fig4. Submission info

Click on one of the type buttons e.g. Corpus. Proceed with filling the basic information such as the title in the following step.

## Step 4: Describe your item

In the following two steps you will provide more details for your item. First describe the people, organization, and the item.



Item submission

1. Basic Info 2. Who's involved 3. Describe 4. Upload 5. License 6. Note 7. Review

# Login to Deposit


- No local accounts:
- institutional logins  
(via EduID-cz, EduGAIN)
- academic institutions all  
over the world
- academic identity by  
default, to submit data
- other options available too


Sign in to **LINDAT/CLARIN Repository**  
Login via Your home institution (e.g. university)

- Univerzita Karlova v Praze  
Czech Republic 6 km
- Institute of Biotechnology CAS, v.v.i.  
Czech Republic 4 km
- Czech University of Life Sciences Prague  
Czech Republic 4 km
- Institute of Art History of the Academy of Sciences of the Czech Republic  
Czech Republic 4 km
- College of Polytechnics Jihlava  
Czech Republic 4 km
- Global Change Research Institute CAS  
Czech Republic 4 km
- Czech Language Institute of the Czech Academy of Sciences  
Czech Republic 4 km
- Institute of Chemical Process Fundamentals of the AS CR  
Czech Republic 4 km
- Institute of Theoretical and Applied Mechanics AS CR



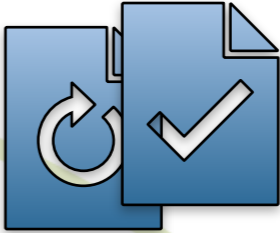


Q: or search for a provider, such as Example University

▶ **Please help, I cannot find my provider**

 Locate me and show nearby providers

Show providers in Czech Republic  show all countries

# Safe Preservation

- No Deletes (persistence is crucial)
- Really? What if I published a nonsense by accident?
- Really.
  - Editors, Reviewers (configurable)  
  - Versioning (replace and explain)  
  - Retractions + Tombstones (worst case) 

# Metadata Search

simple ...

Search

[Advanced Search](#)

## Author

Hajič, Jan (47)  
Žabokrtský, Zdeněk (32)  
Straka, Milan (29)  
Zeman, Daniel (29)  
Bojar, Ondřej (28)  
... View More

## Subject

Germanistik (47)  
machine translation (39)  
corpus (34)  
treebank (30)  
morphology (26)  
... View More

## Language (ISO)

English (222)  
Czech (192)  
German (159)  
Dutch (92)  
Spanish (83)  
... View More

What's New

? What can you do?

ToolService

LINDAT / CLARIN

DEPOSIT



# ... or with facets

Advanced Search

Search


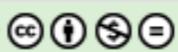
Limit your search

- Author
- Subject
- Rights
- Language (ISO)
- Type
- Contain Files
- Community


Showing 1 through 10 out of 1038 results

1 2 3 > 104

Corpus LINDAT / CLARIN

**AKCES 2 ver. 2**   
(Charles University in Prague, ÚČJTK / 2013-12-18)  
**Author(s):** Šebesta, Karel ; Goláňová, Hana  
This item contains 1 file (3.85 MB). Publicly Available 

LexicalConceptualResource LRT + Open Submissions

**A Gold Standard Word Alignment for English-Swedish (2015-10-12)**   
(Linköping University / 2015-10-12)  
**Author(s):** Ahrenberg, Lars ; Holmqvist, Maria  
This item contains 1 file (590 KB).



What can you do?



Browse

> All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

Help Desk

# Faceted Search

  [Search](#)

[Advanced Search](#)



? What can you do?



🎯 Browse

> All of the Repository

👤 My Account

➔ Login

📘 General Information

📄 Deposit

🗉 Cite

🔄 Submission Lifecycle

? FAQ

📄 About

✉ Help Desk

## Limit your search

Author

Subject

Rights

Language (ISO)

Type

Contain Files

Community

Showing 1 through 10 out of 1038 results

1 2 3 > 104



Corpus

LINDAT / CLARIN

AKCES 2 ver. 2

(Charles University in Prague, ÚČJTK / 2013-12-18)



Author(s):

Šebesta, Karel ; Goláňová, Hana

This item contains 1 file (3.85 MB).

Publicly Available

LexicalConceptualResource

LRT + Open Submissions

A Gold Standard Word Alignment for English-Swedish (2015-10-12)

(Linköping University / 2015-10-12)



Author(s):

Ahrenberg, Lars ; Holmqvist, Maria

This item contains 1 file (590 KB).



# Discovery

Google



prague dependency treebank 3.0



Vše

Obrázky

Nákupy

Mapy

Videa

Více

Nastavení

Nástroje

Přibližný počet výsledků: 13 900 (0,44 s)

## Vědecké články o prague dependency treebank 3.0

**Prague dependency treebank 3.0** - [Bejček](#) - Počet citací tohoto článku: 47

**Prague Dependency Treebank** - [Hajič](#) - Počet citací tohoto článku: 385

**The Prague dependency treebank** - [Böhmová](#) - Počet citací tohoto článku: 423

## Prague Dependency Treebank 3.0 | ÚFAL

<https://ufal.mff.cuni.cz/pdt3.0> ▼ Přeložit tuto stránku

Introduction. The **Prague Dependency Treebank 3.0** (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDiT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed and improved in various aspects. Moreover ...

## The Prague Dependency Treebank 2.0.

<https://ufal.mff.cuni.cz/pdt2.0/> ▼ Přeložit tuto stránku

The **Prague Dependency Treebank 2.0** (PDT 2.0) contains a large amount of Czech texts with complex and interlinked morphological (2 million words), syntactic (1.5 MW) and complex semantic annotation ... Please note that new versions of this corpus have been published: PDT 3.0 (2013), PDiT 1.0 (2012), PDT 2.5 (2012).

## Prague Dependency Treebank 3.0 (PDT 3.0)

[https://lindat.mff.cuni.cz/repository/xmlui/bitstream/.../PDT30\\_index\\_lindat.html?...](https://lindat.mff.cuni.cz/repository/xmlui/bitstream/.../PDT30_index_lindat.html?...)

**Prague Dependency Treebank 3.0** (PDT 3.0). Overview. The **Prague Dependency Treebank 3.0** (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDiT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed ...

# Direct Data Citations

## Credit for Data

enTenTen

Please use the following text to cite this item or export to a predefined format:

BIBTEX CMDI

Masaryk University, NLP Centre, 2011, *enTenTen*, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8>.

This resource is also integrated in following services:

KonText

Share:

Item Identifier	<a href="http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8">http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8</a>
Date issued	2011-12-16
Type	corpus
Language(s)	English
Description	Very large English web corpus enTenTen, comprising 3,268,798,627 tokens.
Publisher	Masaryk University, NLP Centre
Acknowledgement	Lexical Computing Ltd.
Subject(s)	English large corpus



What can you do?

DEPOSIT

Browse

> All of the Reposit

My Account

Logout

Profile

Submissions

Context

> Edit this item

> Export Item

> Export Metadata



# Direct Data Citations

## Credit for Data

### Etalon 1.0

for word processors

“ Please use the following text to cite this item or export to a predefined format:

Skoumalová, Hana, 2021, *Etalon 1.0*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-3698>.

BIBTEX

CMDI

bibtex

for LaTeX users

```
@misc{11234/1-3698,  
title = {Etalon 1.0},  
author = {Skoumalov{'a}, Hana},  
url = {http://hdl.handle.net/11234/1-3698},  
note = {{LINDAT}/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University},  
copyright = {Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)},  
year = {2021} }
```

Share:  

LINDAT / CLARIAH-CZ

Authors

Skoumalová, Hana

Item identifier

<http://hdl.handle.net/11234/1-3698>



Browse

> All of the Repository

My Account

Login

Statistics

Statistics

General Information

Deposit

Cite

Submission Life Cycle

FAQ

About

Please use the following text to cite this item or export to a predefined format:

Straka, Milan; Náplava, Jakub; Straková, Jana and Samuel, David, 2021, *RobeCzech Base*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-3691>.

Share:  

LINDAT / CLARIAH-CZ

**Authors** Straka, Milan ; Náplava, Jakub ; Straková, Jana and Samuel, David

**Item identifier** <http://hdl.handle.net/11234/1-3691>

**Date issued** 2021-05-25

authors =  
DC.creator + DC.creator ...

- Browse
  - All of the Repository
- My Account
  - Login
- Statistics
  - Statistics **BETA**
- General Information
- Deposit

Elements Console Sources Network Timelines Storage Graphics Layers Audit

1-3691 Response

```
<title xmlns:i18n="http://apache.org/cocoon/i18n/2.1">RobeCzech Base</title>
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/">
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
<meta name="DC.creator" content="Straka, Milan"/>
<meta name="DC.creator" content="Náplava, Jakub"/>
<meta name="DC.creator" content="Straková, Jana"/>
<meta name="DC.creator" content="Samuel, David"/>
<meta name="DCTERMS.dateAccepted" content="2021-05-25T08:15:54Z" scheme="DCTERMS.W3CDTF"/>
<meta name="DCTERMS.available" content="2021-05-25T08:15:54Z" scheme="DCTERMS.W3CDTF"/>
<meta name="DCTERMS.issued" content="2021-05-25" scheme="DCTERMS.W3CDTF"/>
<meta name="DC.identifier" content="http://hdl.handle.net/11234/1-3691" xml:lang="*" scheme="DCTERMS.URI"/>
<meta name="DC.description" content="RobeCzech is a monolingual RoBERTa language representation model trained on Czech data. RoBERTa is a robustly optimized Transformer-based pretraining approach. We show that RobeCzech considerably outperforms equally-sized multilingual and Czech-trained contextualized language representation models, surpasses current state of the art in all five evaluated NLP tasks and reaches state-of-the-art results in four of them. The RobeCzech model is released publicly at https://hdl.handle.net/11234/1-3691 and https://huggingface.co/ufal/robeczech-base, both for PyTorch and TensorFlow." xml:lang="en_US"/>
<meta name="DC.language" content="ces" xml:lang="en_US" scheme="DCTERMS.RFC1766"/>
<meta name="DC.publisher" content="Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL)" xml:lang="en_US"/>
<meta name="DC.rights" content="Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)" xml:lang="*">
<meta name="DC.rights" content="http://creativecommons.org/licenses/by-nc-sa/4.0/" xml:lang="*" scheme="DCTERMS.URI"/>
<meta name="DC.subject" content="Czech" xml:lang="en_US"/>
<meta name="DC.subject" content="BERT" xml:lang="en_US"/>
<meta name="DC.subject" content="RoBERTa" xml:lang="en_US"/>
<meta name="DC.title" content="RobeCzech Base" xml:lang="en_US"/>
<meta name="DC.type" content="languageDescription" xml:lang="en_US"/>
<meta name="DC.rights" content="PUB" xml:lang="*">
<meta content="Czech; BERT; RoBERTa; languageDescription" name="citation_keywords"/>
<meta content="RobeCzech Base" name="citation_title"/>
<meta content="Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL)" name="citation_publisher"/>
<meta content="ces" name="citation_language"/>
<meta content="Straka, Milan" name="citation_author"/>
<meta content="Náplava, Jakub" name="citation_author"/>
<meta content="Straková, Jana" name="citation_author"/>
<meta content="Samuel, David" name="citation_author"/>
<meta content="https://lindat.mff.cuni.cz/repository/xmlui/bitstream/11234/1-3691/2/robeczech-base-tf.zip" name="citation_pdf_url"/>
<meta content="2021-05-25" name="citation_date"/>
<meta content="https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-3691" name="citation_abstract_html_url"/>
<link rel="stylesheet" type="text/css" href="//lindat.mff.cuni.cz/aai/discojuice/discojuice.css"/>
<script type="application/ld+json">
{"@context":"https://schema.org","@type":"Dataset","identifier":"http://hdl.handle.net/11234/1-3691","license":"http://creativecommons.org/licenses/by-nc-sa/4.0/","creator":
[{"@type":"Person","name":"Straka, Milan","familyName":"Straka","givenName":"Milan"}, {"@type":"Person","name":"Náplava, Jakub","familyName":"Náplava","givenName":"Jakub"},
{"@type":"Person","name":"Straková, Jana","familyName":"Straková","givenName":"Jana"}, {"@type":"Person","name":"Samuel, David","familyName":"Samuel","givenName":"David"}], "keywords":
```

Google metadata – SEO



Please use the following text to cite this item or export to a predefined format:

BIBTEX CMDI

Straka, Milan; Náplava, Jakub; Straková, Jana and Samuel, David, 2021, *RobeCzech Base*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-3691>.



Share:

LINDAT / CLARIAH-CZ

Authors Straka, Milan ; Náplava, Jakub ; Straková, Jana and Samuel, David

Item identifier <http://hdl.handle.net/11234/1-3691>

Date issued 2021-05-25

Browse

> All of the Repository

My Account

Login

Statistics

Statistics

BETA

General Information

Deposit

Elements Console Sources Network Timelines Storage Graphics Layers Audit

1-3691 Response

```

97 <meta content="Straka, Milan" name="citation_author"/>
98 <meta content="Náplava, Jakub" name="citation_author"/>
99 <meta content="Straková, Jana" name="citation_author"/>
100 <meta content="Samuel, David" name="citation_author"/>
101 <meta content="https://lindat.mff.cuni.cz/repository/xmlui/bitstream/11234/1-3691/2/robeczech-base-tf.zip" name="citation_pdf_url"/>
102 <meta content="2021-05-25" name="citation_date"/>
103 <meta content="https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-3691" name="citation_abstract_html_url"/>
104 <link rel="stylesheet" type="text/css" href="//lindat.mff.cuni.cz/aa1/discojuice/discojuice.css"/>
105 <script type="application/ld+json">
106 {
  "@context": "https://schema.org",
  "@type": "Dataset",
  "identifier": "http://hdl.handle.net/11234/1-3691",
  "license": "http://creativecommons.org/licenses/by-nc-sa/4.0/",
  "creator": [
    {"@type": "Person", "name": "Straka, Milan", "familyName": "Straka", "givenName": "Milan"},
    {"@type": "Person", "name": "Náplava, Jakub", "familyName": "Náplava", "givenName": "Jakub"},
    {"@type": "Person", "name": "Straková, Jana", "familyName": "Straková", "givenName": "Jana"},
    {"@type": "Person", "name": "Samuel, David", "familyName": "Samuel", "givenName": "David"}
  ],
  "keywords": ["Czech", "BERT", "RoBERTa"],
  "name": "RobeCzech Base",
  "description": "RobeCzech is a monolingual RoBERTa language representation model trained on Czech data. RoBERTa is a robustly optimized Transformer-based pretraining approach. We show that RobeCzech considerably outperforms equally-sized multilingual and Czech-trained contextualized language representation models, surpasses current state of the art in all five evaluated NLP tasks and reaches state-of-the-art results in four of them. The RobeCzech model is released publicly at https://hdl.handle.net/11234/1-3691 and https://huggingface.co/ufal/robeczech-base, both for PyTorch and TensorFlow.",
  "url": "http://hdl.handle.net/11234/1-3691"
}
107 </script>
108 </head>
109 <!--[if lt IE 7 ]> <body class="ie6"> <![endif]-->
110 <!--[if IE 7 ]> <body id="lindat-repository" class="ie7"> <![endif]-->
111 <!--[if IE 8 ]> <body id="lindat-repository" class="ie8"> <![endif]-->
112 <!--[if IE 9 ]> <body id="lindat-repository" class="ie9"> <![endif]-->
113 <!--[if (gt IE 9)!(IE)]><!-->
114 <body id="lindat-repository">
115 <!--<![endif]-->
116 <nav class="lindat-header lindat-common" role="navigation" data-version="2.4.0" data-build="4804a6ee9b428697c4040cbebbbed83884739b89e">
117 <button type="button" class="lindat-menu-btn" onclick="document.querySelector('.lindat-menu').classList.toggle('lindat-open')">
118 <span class="lindat-menu-icon">
119 <span class="lindat-icon-bar"></span>
120 <span class="lindat-icon-bar"></span>
121 <span class="lindat-icon-bar"></span>
122 </span>
123 </button>
124 <ul class="lindat-menu">
125 <li class="lindat-home-item">
126 <a href="http://lindat.mff.cuni.cz/en/" class="lindat-logo">
127 <span>Home</span>
128 </a>
129 </li>
130 <li class="lindat-repository-item">
131 <a href="https://lindat.mff.cuni.cz/repository/xmlui/?locale-attribute=en">
132 <span>Repository</span>
133 </a>

```

Google Dataset Search metadata



# Cite data properly and easily

- Force-11 Data Citation Principles
  - Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 <https://doi.org/10.25490/a97f-egykyk>
- Formated citation in simple text: copy&paste for MS Word
- BibTeX record for those who typeset papers in LaTeX
- DataCite DOI – open door for more citation formatting and export...

# Licensing

## As Open as Possible

### Choose a License

Answer the questions or use the search to find the license you want

↻ Start again ← | →

What do you want to deposit?


Software Data

Search for a license...

---

**Public Domain Mark (PD)**



The work identified as being free of known restrictions under copyright law, including all related and neighboring rights.

Publicly Available 

---

**Public Domain Dedication (CC Zero)**



CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Publicly Available   **OPEN DATA**

---

**Creative Commons Attribution (CC-BY)**

This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

Publicly Available   **OPEN DATA**

# Licensing

## As Open as Possible (not more)

 **Publisher** Faculty of Arts, Institute of the Czech National Corpus, Charles University in Prague

 **Acknowledgement** Ministerstvo školství, mládeže a tělovýchovy

Project code: LM2011023

Project name: Český národní korpus

 **Subject(s)** representative corpus written language

 **Collection(s)** LINDAT / CLARIN Data & Tools

[Show full item record](#)

### Files in this item



Download instructions for command line

This item is **Academic Use** and licensed under:  
Czech National Corpus (Shuffled Corpus Data)



<b>Name</b>	syn2015.gz
<b>Size</b>	1.48 GB
<b>Format</b>	application/x-gzip
<b>Description</b>	corpus
<b>MD5</b>	e0242cc77e999794af6cfaf57f843c12



[Download file](#)

CLEAR RULES  
CUSTOM LICENSES  
LICENSE SIGNING

# Licensing Framework:

Any License (Open Source / Open Data preferred)

## Manage Licenses

All Licenses

Define License

Define License Label

	License Name	Definition (URL)	Confirmation	Required user info	License Label	Extended Labels	Used by Bitstreams
<input type="checkbox"/>	Universal Derivations v0.5 License Agreement	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UDer-0.5">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UDer-0.5</a>	Not required		PUB	CC	1
<input type="checkbox"/>	Licence Universal Dependencies v2.4	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.4">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.4</a>	Not required		PUB	CC GPLv3	4
<input type="checkbox"/>	License agreement for The Multilingual corpus of literal occurrences of multiword expressions	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-mwe-literal">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-mwe-literal</a>	Not required		PUB	CC GPLv3	5
<input type="checkbox"/>	Licence Universal Dependencies v2.3	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.3">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.3</a>	Not required		PUB	CC GPLv3 GPLv2	4
<input type="checkbox"/>	PARSEME Shared Task Data (v. 1.1) Agreement	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-mwe-1.1">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-mwe-1.1</a>	Not required		PUB	CC GPLv3	22
<input type="checkbox"/>	Licence Universal Dependencies v2.2	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.2">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.2</a>	Not required		PUB	CC GPLv3 GPLv2	7
<input type="checkbox"/>	Licence Universal Dependencies v2.1	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.1">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.1</a>	Not required		PUB	CC GPLv3 GPLv2	4
<input type="checkbox"/>	PARSEME Shared Task Data (v. 1.0) Agreement	<a href="https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.1">https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.1</a>	Not required		PUB	CC GPLv3	21



? What can you do?



🎯 Browse

> All of the Repository

👤 My Account

🚪 Logout

👤 Profile

📤 Submissions

⚙️ Administrative

🔧 Control Panel

🔑 Access Control

> Content Administration

📊 Registries

🏠 Collections & Communities

# Licensing Framework

## Defining a New License

### Manage Licenses

[All Licenses](#)[Define License](#)[Define License Label](#)

#### Define new licence

License name	<input type="text"/>
License definition URL	<input type="text"/>
License requires confirmation	<input type="text" value="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)



? What can you do?



🎯 Browse

> All of the Repository

👤 My Account

🚪 Logout

👤 Profile

📤 Submissions

⚙️ Administrative

🔧 Control Panel

🔑 Access Control

> Content Administration

📊 Registries

👤 Collections & Communities



# Licensing Framework

## Defining a licensing label / attribute

### Manage Licenses

[All Licenses](#)[Define License](#)[Define License Label](#)

#### Define New Label

Short Label	<input type="text"/>
Label Title	<input type="text"/>
Is extended	<input type="text" value="Yes"/>
Icon image	<input type="button" value="Choose File"/> no file selected
	<input type="button" value="Save"/> <input type="button" value="Return"/>



? What can you do?



🎯 Browse

> All of the Repository

👤 My Account

🚪 Logout

👤 Profile

📄 Submissions

⚙️ Administrative

🔧 Control Panel

👁️ Access Control

> Content Administration

🗄️ Registries

👥 Collections & Communities

# Licensing Framework

## License with labels in use

 **Publisher** [Faculty of Arts, Institute of the Czech National Corpus, Charles University in Prague](#)

 **Acknowledgement** [Ministerstvo školství, mládeže a tělovýchovy](#)

Project code: LM2011023

Project name: Český národní korpus

 **Subject(s)** [representative corpus](#) [written language](#)

 **Collection(s)** [LINDAT / CLARIN Data & Tools](#)

[Show full item record](#)

 **Files in this item**



[Download instructions for command line](#)

This item is **Academic Use** and licensed under:  
Czech National Corpus (Shuffled Corpus Data)



<b>Name</b>	syn2015.gz
<b>Size</b>	1.48 GB
<b>Format</b>	application/x-gzip
<b>Description</b>	corpus
<b>MD5</b>	e0242cc77e999794af6cfaf57f843c12



[Download file](#)

LICENSE TYPE, NAME,  
LINK TO FULL TEXT,  
LABELS

# Licensing Framework

## License with labels in use

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: MSM 0021620838

Project name: Moderní metody, struktury a systémy informatiky

### Subject(s)

MorphoDiTa Czech morphological analysis morphological generation PoS tagging

### Collection(s)

LINDAT / CLARIN Data & Tools

### Other versions

List all versions ▾

- ▶ Czech Models (Morfflex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
- Czech Models (Morfflex CZ 160310 + PDT 3.0) for MorphoDiTa 160310
- Czech Models (Morfflex CZ + PDT) for MorphoDiTa

[Show full item record](#)

### Files in this item



Download instructions for command line

This item is **Publicly Available** and licensed under:  
Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)



<b>Name</b>	czech-morfflex-pdt-161115.zip
<b>Size</b>	69.18 MB
<b>Format</b>	application/zip
<b>Description</b>	Czech Models (Morfflex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
<b>MD5</b>	adde38cd363219759e19165b06baa4ce



[Download file](#) [Preview](#)

LICENSE TYPE, NAME,  
LINK TO FULL TEXT,  
LABELS

# Versioning

prefer latest, preserve all

Project name: Internet jako jazykový korpus

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: LN00A063

Project name: Centrum počítačnické lingvistiky

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: MSM 0021620838

Project name: Moderní metody, struktury a systémy informatiky

## Subject(s)

MorphoDiTa

Czech

morphological analysis

morphological generation

PoS tagging

## Collection(s)

LINDAT / CLARIN Data & Tools



This item is replaced by a newer submission:

<http://hdl.handle.net/11234/1-1836>

Please refer to the submission above for the latest available data. If you nevertheless need the original data, please click [here](#).

List all versions ▼

# Versioning

prefer latest, preserve all

## Other versions

List all versions ▾

- ▶ Czech Models (Morflex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
- Czech Models (Morflex CZ 160310 + PDT 3.0) for MorphoDiTa 160310
- Czech Models (Morflex CZ + PDT) for MorphoDiTa

[Show full item record](#)

## Files in this item



Download instructions for command line

This item is **Publicly Available** and licensed under:

Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)



<b>Name</b>	czech-morfflex-pdt-161115.zip
<b>Size</b>	69.18 MB
<b>Format</b>	application/zip
<b>Description</b>	Czech Models (Morflex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
<b>MD5</b>	adde38cd363219759e19165b06baa4ce



[Download file](#)

[Preview](#)

# Versioning

prefer latest, preserve all

The submission process includes describing the item and uploading the file(s) comprising it. Each community or collection may set its own submission policy.

## Archived Submissions

These are your completed submissions which have been accepted into DSpace.

	Date accepted	Title	Collection
<input checked="" type="checkbox"/>	2015-11-15	<a href="#">Universal Dependencies 1.2</a>	<a href="#">LINDAT / CLARIN Data &amp; Tools</a>
<input type="checkbox"/>	2015-08-21	<a href="#">HamleDT 3.0</a>	<a href="#">LINDAT / CLARIN Data &amp; Tools</a>
<input type="checkbox"/>	2015-01-28	<a href="#">Lingua::Interaset 2.026</a>	<a href="#">LINDAT / CLARIN Data &amp; Tools</a>
<input type="checkbox"/>	2015-01-19	<a href="#">Universal Dependencies 1.0</a>	<a href="#">LINDAT / CLARIN Data &amp; Tools</a>
<input type="checkbox"/>	2014-05-24	<a href="#">HamleDT 2.0</a>	<a href="#">LINDAT / CLARIN Data &amp; Tools</a>

Add new version


DEPOSIT



CITE

 Browse


> All of the Repository

 My Account

 Logout

 Profile

 Submissions

 General Information

 Deposit

 Cite

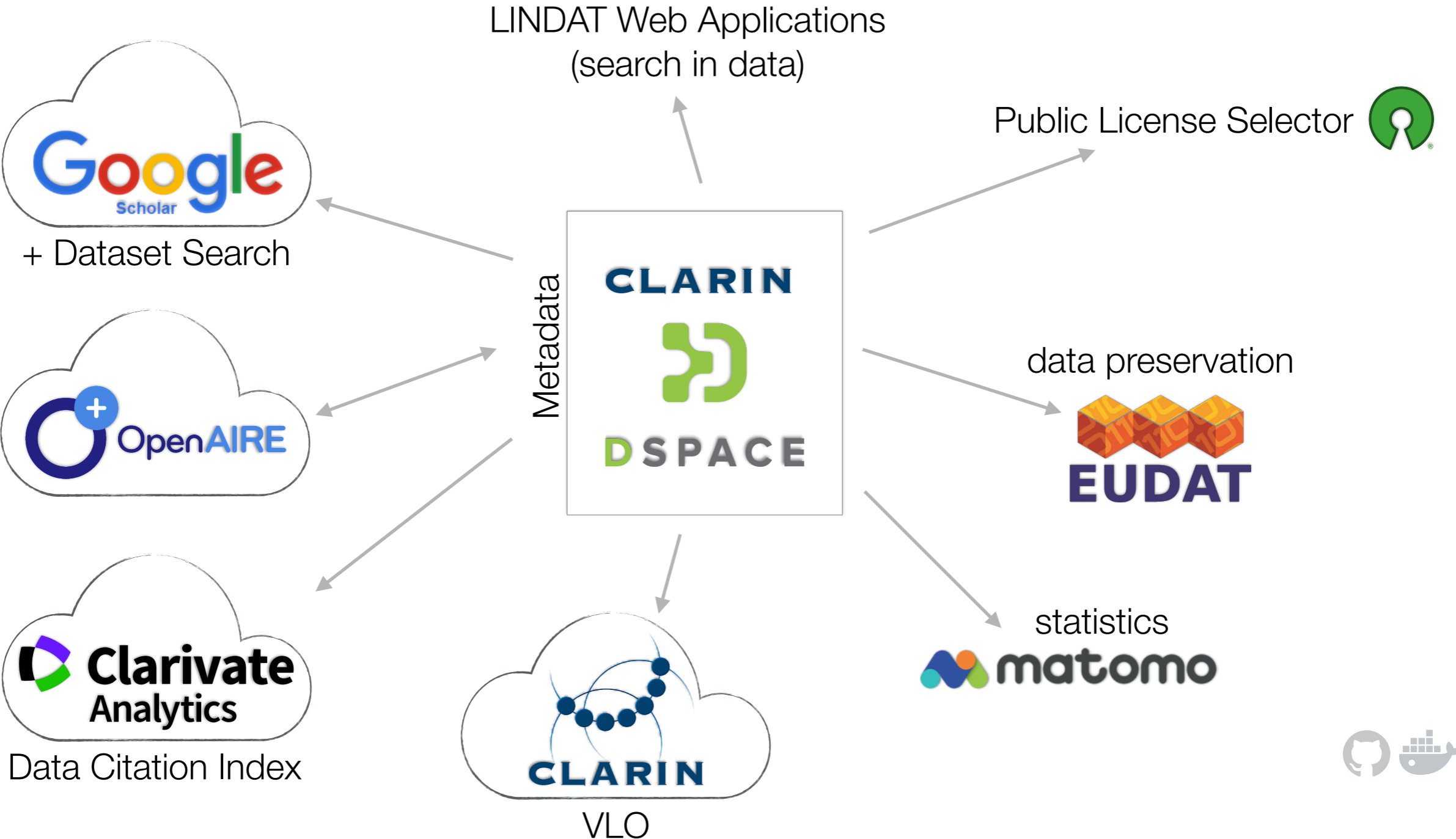
 Submission Lifecycle

? FAQ

 About

 Help Desk

# Repository Integrations (example)



# FAIR summary

- **F***indable*: Google, Google Scholar, Data Citation Index, CLARIN VLO, OLAC... and the repository itself
- **A***ccessible*: records with data (even when restricted), complete licensing, Open Access (where possible), login only when needed, persistence
- **I***nteroperable*: common data formats, full documentation (enhanced metadata, documentation bitstreams), machine readability
- **R***eusable*: records with data, known formats, complete licensing, full versioning, direct data citations, maximal OA



# Statistics and Reporting

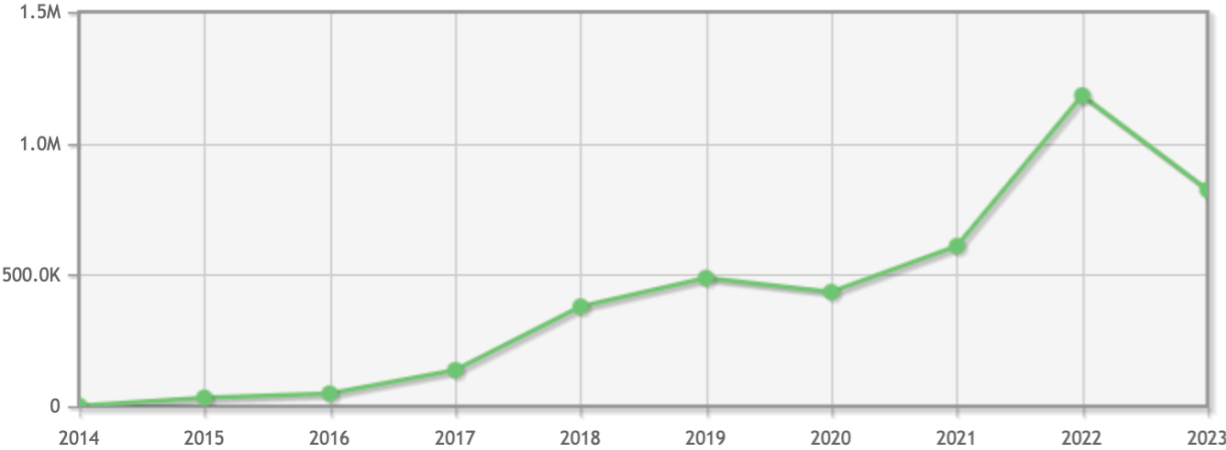
Overall Repository (Views) **Repository (Downloads)** LRT Services Others Notes

Showing statistics for 2014 - 2023

The download statistics are based on the tracking request PIWIK received from the JAVA API directly from DSpace. The download counts also contain hits from bots.

### Downloads Over Time

2014 - 2023



Click on a data item to summarize by year / month.

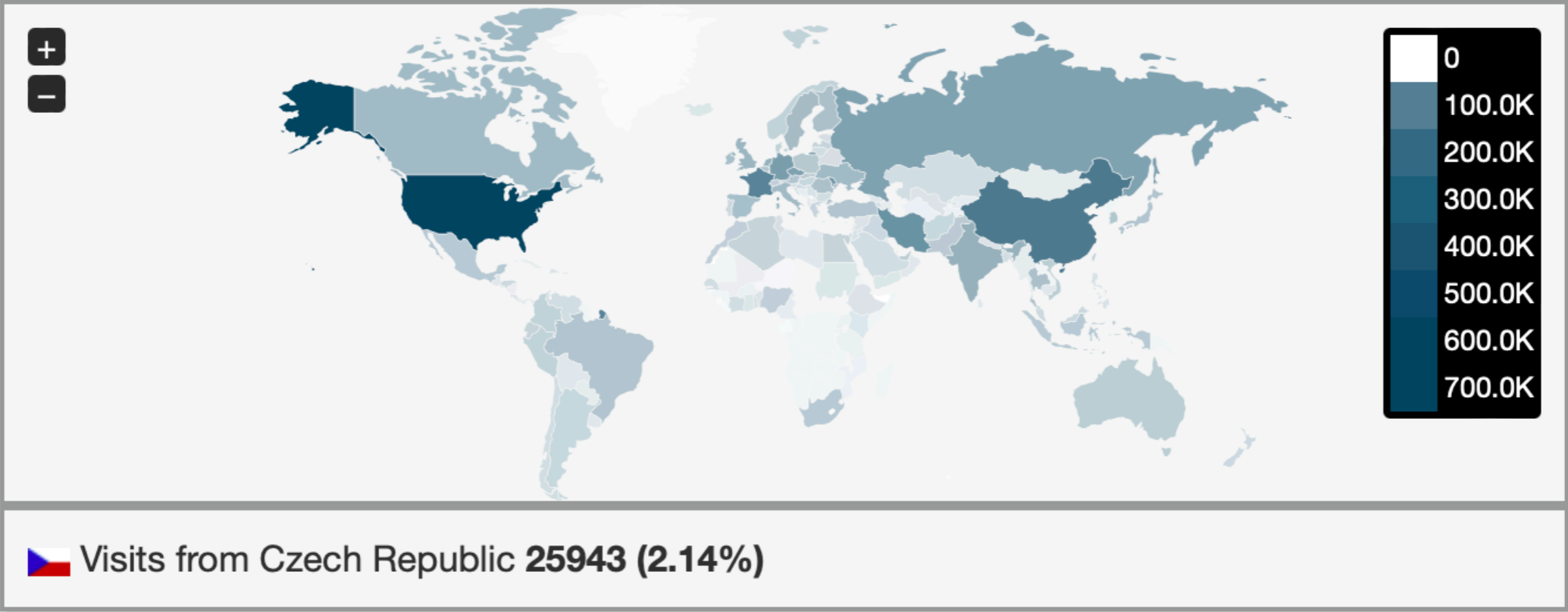
### Important Metrics

Downloads	<b>4137487</b>	Unique Downloads	<b>2960822</b>
Visits	<b>1211775</b>	Unique Visitors	<b>1211201</b>

Hover on a metric name to show its definition.

# Statistics and Reporting

## Country Wise Visits



# Statistics and Reporting – per item

BETA

Get a monthly report for this item via email

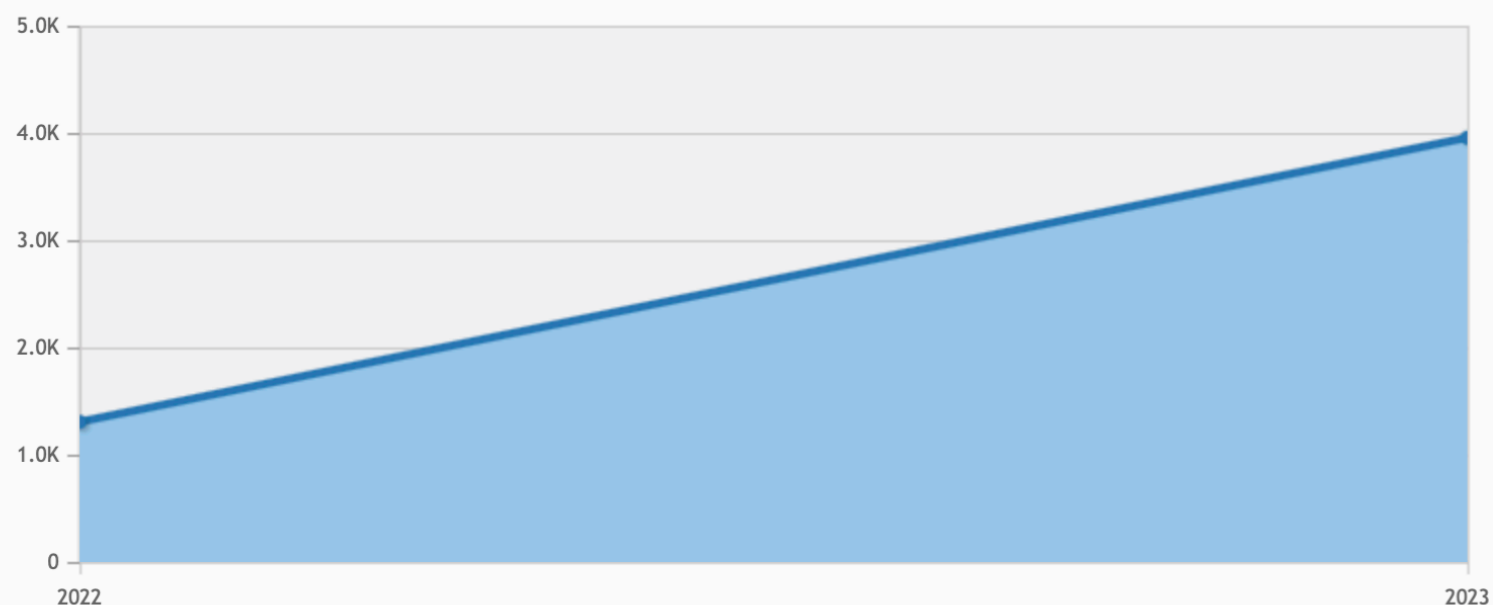
Subscribe

Click on a data point to summarize by year / month.

Views **4266**

Downloads **5264**

Statistics for years 2022 to 2023



## Filewise Statistics

2022

**801** ud-treebanks-v2.11.tgz

**244** allzip

**133** ud-documentation-v2.11.tgz

**128** ud-tools-v2.11.tgz

**1** ud-treebanks-v2.0.tgz



Browse

> All of the Repository

My Account

Logout

Profile

Submissions

Context

> Edit this item

> Export Item

> Export Metadata

Administrative

Control Panel

Access Control

> Content Administration

Registries

Collections & Communities

Statistics

Curation Tasks

Licenses

Handles

# Statistics and Reporting – per item

BETA

Get a monthly report for this item via email

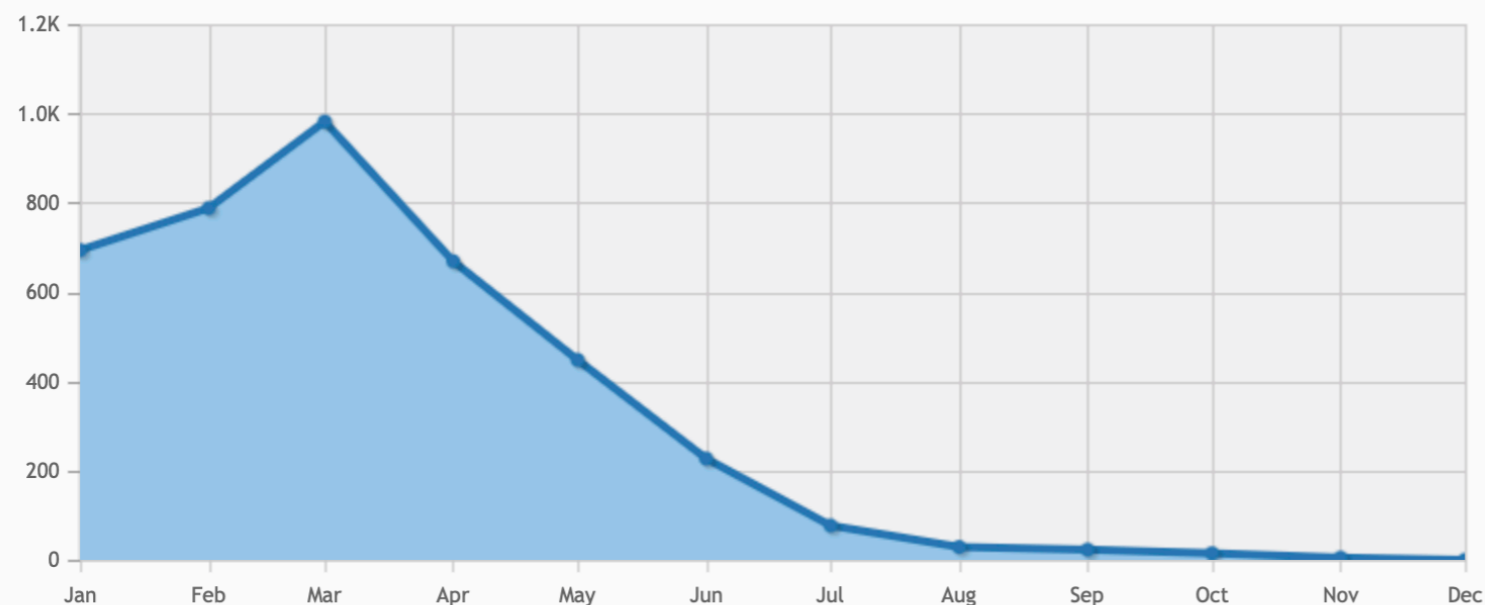
Subscribe

Click on a data point to summarize by year / month.

 Views **3309**

 Downloads **3957**

Statistics for the year 2023 [Back](#)



## Filewise Statistics

2283	ud-treebanks-v2.11.tgz
839	allzip
412	ud-documentation-v2.11.tgz
412	ud-tools-v2.11.tgz
4	ud-documentation-v2.11.tgz
4	ud-tools-v2.11.tgz



 Browse

> All of the Repository

 My Account

 Logout

 Profile

 Submissions

 Context

> Edit this item

> Export Item

> Export Metadata

 Administrative

 Control Panel

 Access Control

> Content Administration

 Registries

 Collections & Communities

 Statistics

 Curation Tasks

 Licenses

 Handles

Thank you!

[stranak@ufal.mff.cuni.cz](mailto:stranak@ufal.mff.cuni.cz)

