

LINDAT / CLARIN DSpace Repository

History and state-of-the-art of CLARIN DSpace



<https://github.com/ufal/clarin-dspace>



- LINDAT dev team
 - ? - ? Petr Pajas
 - 2011: Jozef Mišutka, Karel Vandas
 - 2012: Jozef Mišutka, Amir Kamran, Bushra Jawaid, Michal Sedlák, Ondřej Košarko
 - 2013: + Loganathan Ramasamy, -Bushra, +- Michal
 - 2014: + Michal Josífko
 - 2015-2016: Jozef, Amir, Ondřej
- dspace versions
 - started with dspace 1.6
 - moved to 1.8.x in 2012
 - backporting interesting features from newer DSpace
 - rebased on 5.x branch in 2015
- tools
 - vcs/issue tracker: svn/Trac, svn/Redmine, git/Redmine, git/github, git/github (proper fork)
 - CI: jenkins, selenium (for ui tests), now travis
 - apache -> nginx, tomcat (6,7,8)...



- automation via makefiles
- configuration files
- backups



- epic api v1, v2
- per community prefix
- handle management
- (shortref.org support via dspace-rest)



List and edit handles

Handle List

This is the list of handles currently resolvable by this DSpace instance.

Showing 1611 through 1612 out of 1612 results

1 < 160 161 162



	Handle	Internal	URL	Resource type	Resource ID
<input type="radio"/>	11234/1-1822	Yes	https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1822	ITEM	3089
<input type="radio"/>	11234/1-1823	Yes	https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1823	ITEM	3090

1 < 160 161 162



New external handle

Edit handle

Delete handle



New Handle

Handle

URL

Save

Cancel



- import script
- metadata QA
- withdrawn (tombstone)



- collections as metadata
- revamped mirage template
- responsive with bootstrap
- interfaces for new features
- submission (later)
- RefBox
- faceted search
- revived autocomplete
- download all zipped
- basic archive preview



Deposit Free and Safe
License of your Choice (Open licenses encouraged)
Easy to Find
Easy to Cite



Search

Advanced Search

Author	Subject	Language (ISO)
Hajjić, Jan (24)	Germanistik (47)	English (198)
Simov, Kiril (23)	corpus (25)	German (122)
Bojar, Ondřej (20)	parallel corpus (14)	Czech (104)
Branco, António (18)	treebank (14)	Dutch (89)
Johannessen, Janne Bondi (18)	machine translation (12)	Finnish (86)
... View More	... View More	

Attractive

Practical

What's New

ToolService

Linguistic digital repository based on DSpace 5.2

Authors: Miskutka, Jozef ; Karman, Amir ; Kostariko, Ondřej ; Josifko, Michal ; Ramasamy, Loganathan ; Straňák, Pavel ; Hajjić, Jan

Description: One of the goals of LINDAT/CLARIN Centre for Language Research Infrastructure is to provide technical background to institutions or researchers who wants to share their tools and data used for research in linguistics or ...

MSTperl parser (2015-05-19)

Authors: Rosta, Rudolf

Description: MSTperl is a Perl implementation of the MST parser of Ryan McDonald (http://www.sesl.upenn.edu/~sectrn/MSTParser/MSTParser.html). MST parser (Maximal state-of-the-art natural language ...)

What can you do?

- DEPOSIT
- CITE
- Browse
- All of the Repository
- My Account
- Cite
- Submission Lifecycle
- FAQ
- About

Responsive and Mobile Friendly

Czech-English Parallel Corpus 1.0 (CzEng 1.0)

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

Bojar, Ondřej; Zdobych, Zdeněk; Dušek, Ondřej; Galuščáková, Petra; Malík, Martin; Mareš, David; Mareš, Jiří; Novák, Michal; Papež, Martin and Těmčková, Aneta. 2015. Czech-English Parallel Corpus 1.0 (CzEng 1.0). LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague. <http://hdl.handle.net/11234/1488>

Authors: Bojar, Ondřej ; Zdobych, Zdeněk ; Dušek, Ondřej ; Galuščáková, Petra ; Malík, Martin ; Mareš, David ; Mareš, Jiří ; Novák, Michal ; Papež, Martin ; Těmčková, Aneta

Project URL: <http://hdl.handle.net/11234/1488>

Demo URL: http://lindat.mff.cuni.cz/services/kontext/hun-og/first_form?corpus=czeng_10_es_8

Date issued: 2015-12

Type: corpus

Size: 206442315 tokens

Language(s): Czech ; English

Easy to Cite

REFBox

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

BIBTEX CMDI

Bejček, Eduard; Hajičová, Eva; Hajič, Jan; Jínová, Pavlína; Kettnerová, Václava; Kolářová, Veronika; Mikulová, Marie; Mirovský, Jiří; Nedoluzhko, Anna; Panevová, Jarmila; Poláková, Lucie; Ševčíková, Magda; Štěpánek, Jan and Zikánová, Šárka, 2013, *Prague Dependency Treebank 3.0*, LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, <http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3>.



Persistent Identifier

Share:   

Social Media Integration

Prague Dependency Treebank 3.0

Press **ctrl + c** to copy

```
<cmd:CMD xmlns:cmd="http://www.clarin.eu/cmd/" xmlns:olac="http://www.clarin.eu/cmd/" xmlns:sm="http://www.clarin.eu/cmd/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" CMDVersion="1.1" xsi:schemaLocation="http://www.clarin.eu/cmd/ http://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/profiles/clarin.eu:cr1:p_1349361150622/xsd">
  <cmd:Header>
    <cmd:MdCreationDate>2016-04-06</cmd:MdCreationDate>
    <cmd:MdSelfLink>http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3&format=cmdi</cmd:MdSelfLink>
    <cmd:MdProfile>clarin.eu:cr1:p_1349361150622</cmd:MdProfile>
    <cmd:MdCollectionDisplayName>LINDAT / CLARIN Data &amp; Tools</cmd:MdCollectionDisplayName>
  </cmd:Header>
  <cmd:Resources>
    <cmd:ResourceProxyList>
      <cmd:ResourceProxy id="lp_200">
        <cmd:ResourceType>LandingPage</cmd:ResourceType>
        <cmd:ResourceRef>http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3</cmd:ResourceRef>
      </cmd:ResourceProxy>
    </cmd:ResourceProxyList>
  </cmd:Resources>
</cmd:CMD>
```

CMDI

Prague Dependency Treebank 3.0

Press **ctrl + c** to copy

```
@misc{11858/00-097C-0000-0023-1AAF-3,
  title = {Prague Dependency Treebank 3.0},
  author = {Bejček, Eduard and Hajičová, Eva and Hajič, Jan and Jínová, Pavlína and Kettnerová, Václava and Kolářová, Veronika and Mikulová, Marie and Mirovský, Jiří and Nedoluzhko, Anna and Panevová, Jarmila and Poláková, Lucie and Ševčíková, Magda and Štěpánek, Jan and Zikánová, Šárka},
  year = {2013},
  url = {http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3},
  doi = {10.1093/lpl/200},
  copyright = {Attribution-(NonCommercial)-(ShareAlike) 3.0 Unported ((CC) (BY)-(NC)-(SA) 3.0)},
}
```

BIBTEX

Export to different formats



<https://github.com/ufal/lindat-common>

Easy to Integrate

Integrate other services

HamleDT 3.0




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

BIBTEX CMDI

Zeman, Daniel; Mareček, David; Mašek, Jan; Popel, Martin; Ramasamy, Loganathan; Rosa, Rudolf; Štěpánek, Jan and Žabokrtský, Zdeněk; 2015, *HamleDT 3.0*, LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, <http://hdl.handle.net/11234/1-1508>.



 This resource is also integrated in following services:

Share:   

PML-TQ

KonText

LINDAT / CLARIN

 Authors

Zeman, Daniel ; Mareček, David ; Mašek, Jan ; Popel, Martin ; Ramasamy, Loganathan ; Rosa, Rudolf ; Štěpánek, Jan ; Žabokrtský, Zdeněk

 Project URL

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



Easy to Find

Searching and Browsing

Easy to Search

Advanced Search

Search

Limit your search

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

Showing 1 through 10 out of 1113 results

1 2 3 > 112

Corpus LINDAT / CLARIN

[Czech-English Parallel Corpus 1.0 \(CzEng 1.0\)](#)
(Charles University in Prague, UFAL / 2011-12)

Author(s):
Bojar, Ondřej ; Žabokrtský, Zdeněk ; Dušek, Ondřej ; Galuščáková, Petra ; Majliš, Martin ; Mareček, David ; Maršík, Jiří ; Novák, Michal ; Popel, Martin ; Tamchyna, Aleš

[This item contains 6 files \(61.72 GB\).](#)

Publicly Available

ToolService LINDAT / CLARIN


[Fairytale child \(2014-11-21\)](#)

Search Listing with clearly visible Licensing, Files and Size information

Search Facets to quickly filter out the search results

Bitstream Preview

Preview what's inside a zip archive or a text file
without downloading



The screenshot shows a file preview window for 'hamledt-3.0.tar'. The file details are as follows:

Name	hamledt-3.0.tar
Size	842.13 MB
Format	application/x-tar
Description	HamleDT 3.0

Below the details are two buttons: 'Download file' and 'Preview'. The 'Preview' button is active, showing a 'File Preview' window with a tree view of the archive's contents:

- hamledt3
 - bg-ud11
 - conllu
 - dev
 - 003.conllu.gz (15 kB)
 - 002.conllu.gz (14 kB)
 - 006.conllu.gz (17 kB)
 - 001.conllu.gz (10 kB)
 - 005.conllu.gz (16 kB)
 - 009.conllu.gz (17 kB)

Easy to Retrieve Item View and Bitstream Access

Czech-English Parallel Corpus 1.0 (CzEng 1.0)

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

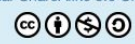
Bojar, Ondřej; Galuščáková, Petra; Majlíš, Martin; Mareček, David; Maršík, Jiří; Novák, David; Maršík, Jiří; Novák, Michal; Popel, Martin; Tamchyna, Aleš, 2011, *Czech-English Parallel Corpus 1.0 (CzEng 1.0)*, LINDAT/CLARIN digital repository of Formal and Applied Linguistics, Charles University in Prague, <http://hdl.handle.net/11234/1-1458>.




Visible Citing Information

Authors	Bojar, Ondřej ; Žabokrtský, Zdeněk ; Dušek, Ondřej ; Galuščáková, Petra ; David ; Maršík, Jiří ; Novák, Michal ; Popel, Martin ; Tamchyna, Aleš
Project URL	http://ufal.mff.cuni.cz/czeng
Demo URL	http://lindat.mff.cuni.cz/services/kontext/run.cgi/first_form?corpname=
Date issued	2011-12
Type	corpus
Size	206442315 tokens
Language(s)	Czech , English

This item is **Publicly Available** and licensed under:
Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0)

Clear Licensing Information



Name	data-plaintext-format.0.tar	
Size	114.09 MB	
Format	application/x-tar	
Description	Sample corpus data in plain text format	
Download file		
Name	data-plaintext-format.tar	
Size	1.09 GB	
Format	application/x-tar	
Description	Complete corpus data in plain text format	
Download file		
Name	data-export-format.0.tar	
Size	875.76 MB	
Format	application/x-tar	
Description	Sample corpus data in export format	

Files in this item

[Download all files in item \(66.13 MB\)](#)

Large Size

The requested files are being packed into one large file. This process can take some time, please be patient.

[Continue](#) [Cancel](#)

Download All



- used hibernate
- licensing framework
- subscription based monthly download/visit summary statistics for items in pdf format



Easy to Distribute Licensing Framework

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"/>	Licence Universal Dependencies v1.1	https://lindat.mff.cuni.cz/repository/xmitui/page/licence-UD-1.1	Not required		PUB	CC GPLv2	3
<input type="checkbox"/>	Creative Commons - Attribution 3.0 Unported (CC BY 3.0)	http://creativecommons.org/licenses/by/3.0/	Not required		PUB	CC BY	12
<input type="checkbox"/>	Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0)	http://creativecommons.org/licenses/by-sa/3.0/	Not required		PUB	CC BY SA	150
<input type="checkbox"/>	Attribution-NoDerivs 3.0 Unported (CC BY-ND 3.0)	http://creativecommons.org/licenses/by-nd/3.0/	Not required		PUB	CC BY ND	2
<input type="checkbox"/>	Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0)	http://creativecommons.org/licenses/by-nc-sa/3.0/	Not required		PUB	CC BY SA NC	89

With our Licensing Framework it is **easy to add new licenses**.

Licenses have labels **PUB** (public), **RES** (restricted), **ACA** (academic). The labels decides the access mechanism of the licenses resources e.g. for PUB label licenses the user is not required to login.

It is also possible to force the user to **sign the license**, with options like sign only once or always.

Licenses can be configured to ask users **extra information** on signing.

Manage Licenses

All Licenses Define License Define License Label

Define new licence

License name

License definition URL

License requires confirmation

License Labels Publicly Available (PUB)
 Academic Use (ACA)
 Restricted Use (RES)

License Labels Attribution Required (BY)
 Share Alike (SA)
 Noncommercial (NC)
 Redeposit Modified (ReD)
 No Derivative Works (ND)
 Inform Before Use (Inf)
 Distributed under Creative Commons (CC)
 No Copyright (ZERO)
 GNU General Public License, version 3.0 (GPLv3)
 GNU General Public License, version 2.0 (GPLv2)
 BSD (BSD)
 The MIT License (MIT)

Easy to Distribute Licensing Framework

License Signing Process

Licence Agreement

The requested content is distributed under one or more licence(s). You have to agree to the licence(s) below before you can obtain the content. Please, read the licence(s) by clicking on the button(s) carefully.

Czech National Corpus (Shuffled Corpus Data)

The information below identifies you, the SIGNER. If the information is incorrect, please contact our [Help Desk](#). This information will be stored as part of the electronic signatures.

Signer	Kamran;Kamran
User ID	g.kamran@lindat.cz
Item handle	11858/00-097C-0000-0023-119E-8
Bitstream	syn2005.gz
IP Address	217.193.6.119

By clicking on the button below I, SIGNER with the above ID, agree to the LICENCE(s) restricting the usage of requested BITSTREAM(s).



I AGREE

Clear visibility of what information we will store

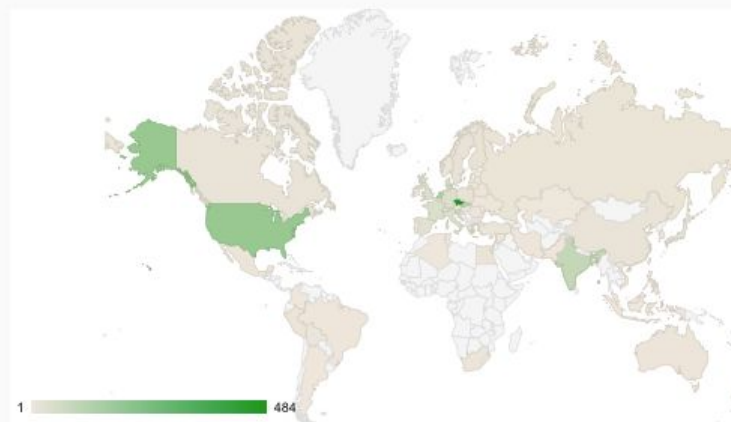
SIGNED LICENSES (showing only first #50)

ID	DATE	USERNAME	LICENSE	ITEM	BITS TREA M	EXTRA METADATA
2015051809-3259-211-730	2015-06-18 09:32:09	217 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	210	730	IP: 190.83.184.217
2015051809-2547-211-698	2015-06-18 09:25:47	217 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	195	698	IP: 190.83.184.217
2015051809-2535-211-701	2015-06-18 09:25:35	217 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	196	701	IP: 190.83.184.217
2015051809-2515-211-692	2015-06-18 09:25:15	217 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	190	692	IP: 190.83.184.217
2015051403-4052-247-849	2015-05-14 03:40:52	217 [NLP Centre Web Corpus License]	NLP Centre Web Corpus License	58	849	IP: 196.116.18.214
2015051106-2747-37-730	2015-05-11 06:27:47	97 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	210	730	IP: 196.116.20.81
2015051104-3945-37-698	2015-05-11 04:39:45	97 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	195	698	IP: 196.116.20.81
2015042311-5740-242-698	2015-04-11 11:57:40	217 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	195	698	IP: 195.115.59.68
2015030210-1901-86-692	2015-03-02 10:18:01	86 [Czech National Corpus (Shuffled Corpus Data)]	Czech National Corpus (Shuffled Corpus Data)	190	692	IP: 196.116.18.209

In Control Panel the administrators can view details of signed licenses.

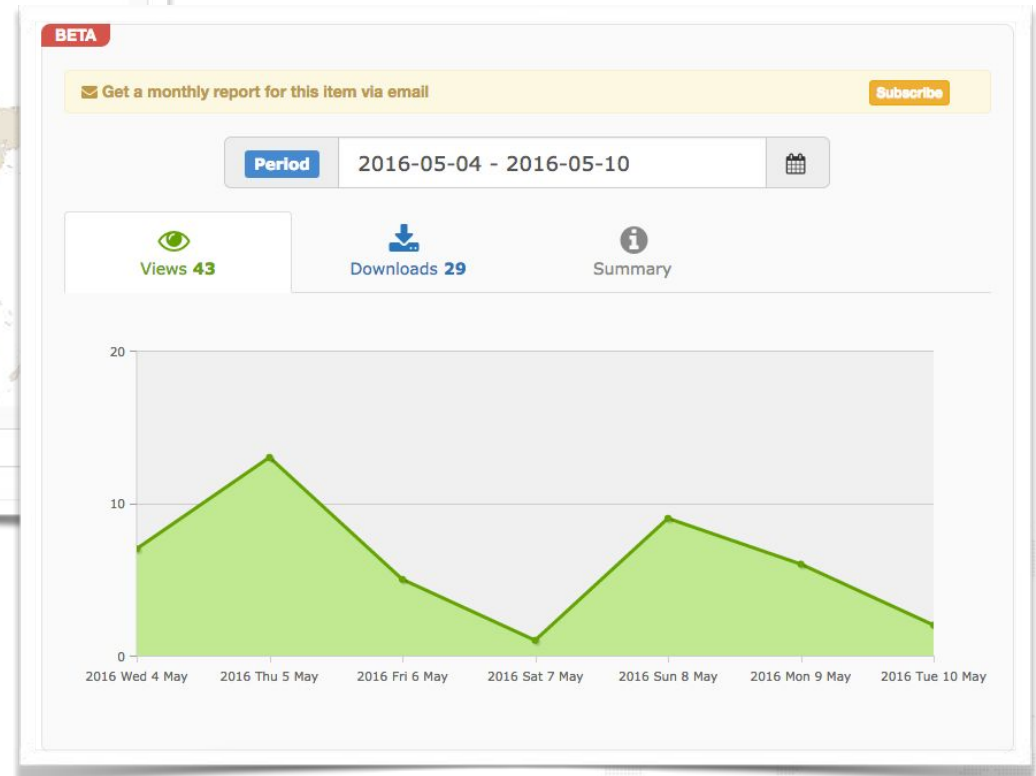
GA Statistics

Profile name	LINDAT/CLARIN repository overview		
Account id	27008245		
Statistics for	Main page		
Time range	<input type="text" value="2015-04-27"/>	<input type="text" value="2015-05-27"/>	<input type="button" value="Update"/>
Total # of visits	1611		



country	visitorType	networkDomain	hostname
Czech Republic	Returning Visitor	cuni.cz	(not set)

- Google Analytics
- PIWIK Integration



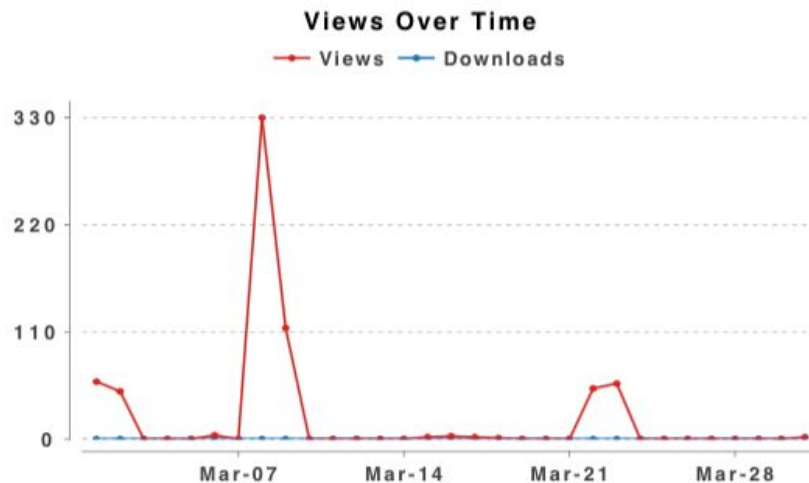
Monthly Item Statistics

Mar, 2015



Testing branding metadata value 2

<http://hdl.handle.net/11858/1458>



Visitors By Country

Netherlands	44
Czech Republic	2
Slovakia	2

Summary

675 pageviews, 101 unique pageviews

48 visits, 21 unique visitors

0 downloads, 0 unique downloads

- Control panel custom tabs (**present in DSpace 6**)
 - settings/commands/logs/metadata
qa/b2safe backup
- healthcheck (**present in DSpace 6**)
- curation tasks
 - also on workflow (editor) items
- exception emails



Control Panel

Control Panel

- Java Information
- Extra Java Info
- Configuration
- Extra Configuration
- SystemWide Alerts
- Programs
- PID
- Shibboleth
- Backup
- IRODs Replication
- Cron Jobs
- OAIPMH Validators
- Harvesting
- Release Notes
- Statistics
- Licenses
- Signed Licenses
- Current Activity
- Checks
- Verify Logging
- Dspace Log(s)
- User Logins
- Shib Raw Logins
- Unpublished Items
- Bitstream items
- Specific Metadata
- Metadata Quality
- Embargoed items
- Oldest users

Java and Operating System

Java Runtime Environment Version:	1.8.0_45
Java Runtime Environment Vendor:	Java HotSpot(TM) 64-Bit Server VM
Operating System Name:	Linux
Operating System Architecture:	amd64
Operating System Version:	2.6.32-19-pve

Control Panel

Verify Logging

- ✓ Logging to /opt/lindat-dspace/installation/log/dspace.log.2016-05-10 OK.
- ✓ Logging to /opt/lindat-dspace/installation/log/resources.log.2016-05-10 OK.
- ✓ Logging to /opt/lindat-dspace/installation/log/curator.log.2016-05-10 OK.
- ✓ Logging to /opt/lindat-dspace/installation/log/checker.log.2016-05-10 OK.
- ✓ Logging to /opt/lindat-dspace/installation/log/dspace.u
- ✓ Logging to /opt/lindat-dspace/installation/log/utilities.l
- ✓ Logging to /opt/lindat-dspace/installation/log/dspace.u
- ✓ Logging to /opt/lindat-dspace/installation/log/cocoon.l
- ✓ Logging to /opt/lindat-dspace/installation/log/authentic

Monitor errors in real time

Choose different file ▾

⚠ File: [dspace.log.2016-05-10] Warnings/Errors: [2]

Exception at line:	Content
1821	2016-05-10 02:25:05,609 ERROR org.dspace.app.xmlui.aspect.discovery.AbstractSearch @ can not sort on multivalued field: dc.date.issued org.dspace.discovery.SearchServiceException: can not sort on multivalued field: dc.date.issued at org.dspace.discovery.SolrServiceImpl.search(SolrServiceImpl.java:1643)
3187	2016-05-10 03:50:10,868 ERROR org.hibernate.engine.jdbc.spi.SqlExceptionHelper @ ERROR: canceling statement due to statement timeout

✓ File: [dspace.ufal.metashare-schema-errors.log.2016-05-10] Warnings/Errors: [0]

✓ File: [curator.log.2016-05-10] Warnings/Errors: [0]

⚠ File: [solr.log.2016-05-10] Warnings/Errors: [1]

Control Panel

Verify Logins

Choose different file ▾

File: [authentication.log] User logins: [1] successful: [1]

Absolute file: [/opt/lindat-dspace/installation/log/authentication.log.2016-05-10]

✓ amirkamran@msn.com;kamran@ufal.mff.cuni.cz @ 2016-05-10 03:14:18,61 => logged in successfully

Showing 41 through 50 out of 989 results

1 < 3 4 5 6 7 > 99

ID	NAME	DESCRIPTION	SIZE (MB)	FORMAT	DELETED
150	pdt2.zip	Prague Dependency Treebank 2.0 data	218.50 MB	zip ▾	false
151	pdt2sample.zip	A sample of the full PDT 2.0 data	0.52 MB	zip ▾	false
152	license.txt		0.00 MB	License ▾	false
154	English-Hindi-without-Emille.tgz	The complete parallel data, including the patch for the Emille corpus	12.16 MB	tar.gz,tgz ▾	false
155	license.txt		0.00 MB	License ▾	false
156	CoNLL2009-ST-Czech.zip	All the Czech cospus files + documentation and the official eval script	14.84 MB	zip ▾	false

View all bitstreams in the repository

Monitor Item progress

WORKFLOW ITEMS

Following items are currently in workflow mode i.e. waiting for an approval. You can view the details of the item by clicking the itemID

WORKFLOW ID	ITEM ID	COLLECTION	CURRENT STATE	ASSIGNED TO	LAST MODIFIED
-------------	---------	------------	---------------	-------------	---------------

WORKSPACE ITEMS

Following items are currently in workspace mode i.e. in the process of submission. You can view the details of the item by clicking the itemID

WORKSPACE ID	ITEM ID	STAGE REACHED	SUBMITTER EMAIL	REJECTED	LAST MODIFIED
208	154	3	amirkamran@msn.com;kamran@ufal.mff.cuni.cz	Not Submitted	2015-05-27 14:12:41.261
217	163	5	ko_ok@centrum.cz	Not Submitted	2015-05-27 09:01:32.112
2993	2921	1	jmisutka@gmail.com	Not Submitted	2015-05-25 22:05:10.742
2991	2919	1	jorg.tiedemann@lingfil.uu.se	Not Submitted	2015-05-19 22:54:03.405
2978	2907	5	ko_ok@centrum.cz	Rejected	2015-05-15 14:04:49.584
2970	2899	5	gdepoorter@taalunie.org	Rejected	2015-05-12 15:02:24.642



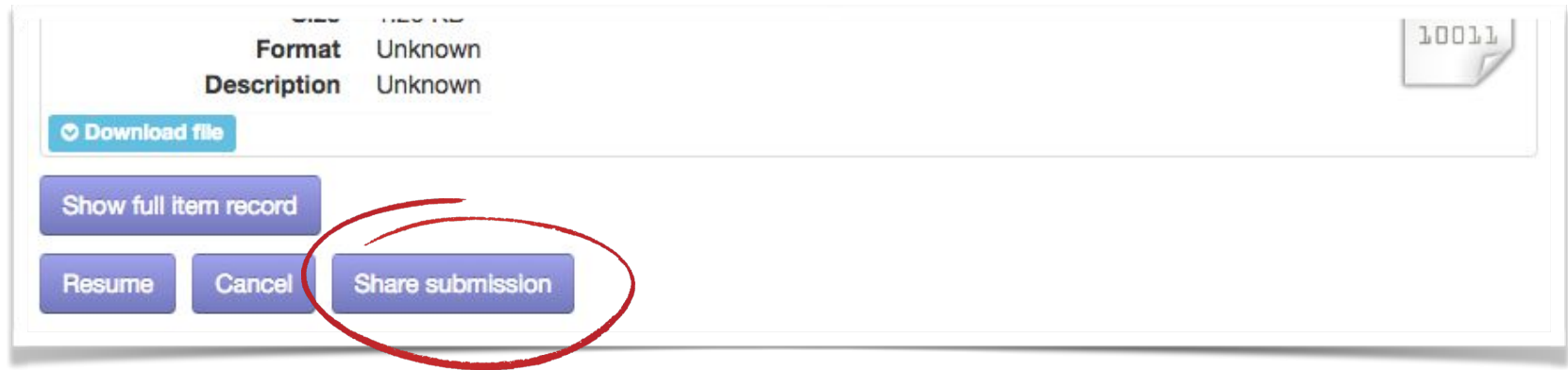
- refactored shibboleth
 - e.g. possibility to fill email if not in attributes
- CoCo welcome message
- (discojuice)
- spinoff projects
 - log parser
 - idp QA checker
 - attribute aggregator



- tweaking the workflow
 - required fields
 - pages
 - comments/hints
- input-forms
 - localization
 - acl for fields
 - regexps
- file uploads
 - d'n'd + multiple files
 - admin uploads (no size limits)
- “sharing” submissions



Work collaboratively - pass submission to another user



The screenshot shows a submission management interface. At the top, there are fields for 'Format' and 'Description', both labeled 'Unknown'. To the right, there is a document icon with the number '10011'. Below these fields, there are several buttons: 'Download file' (light blue), 'Show full item record' (dark blue), 'Resume' (dark blue), 'Cancel' (dark blue), and 'Share submission' (dark blue). The 'Share submission' button is circled in red.

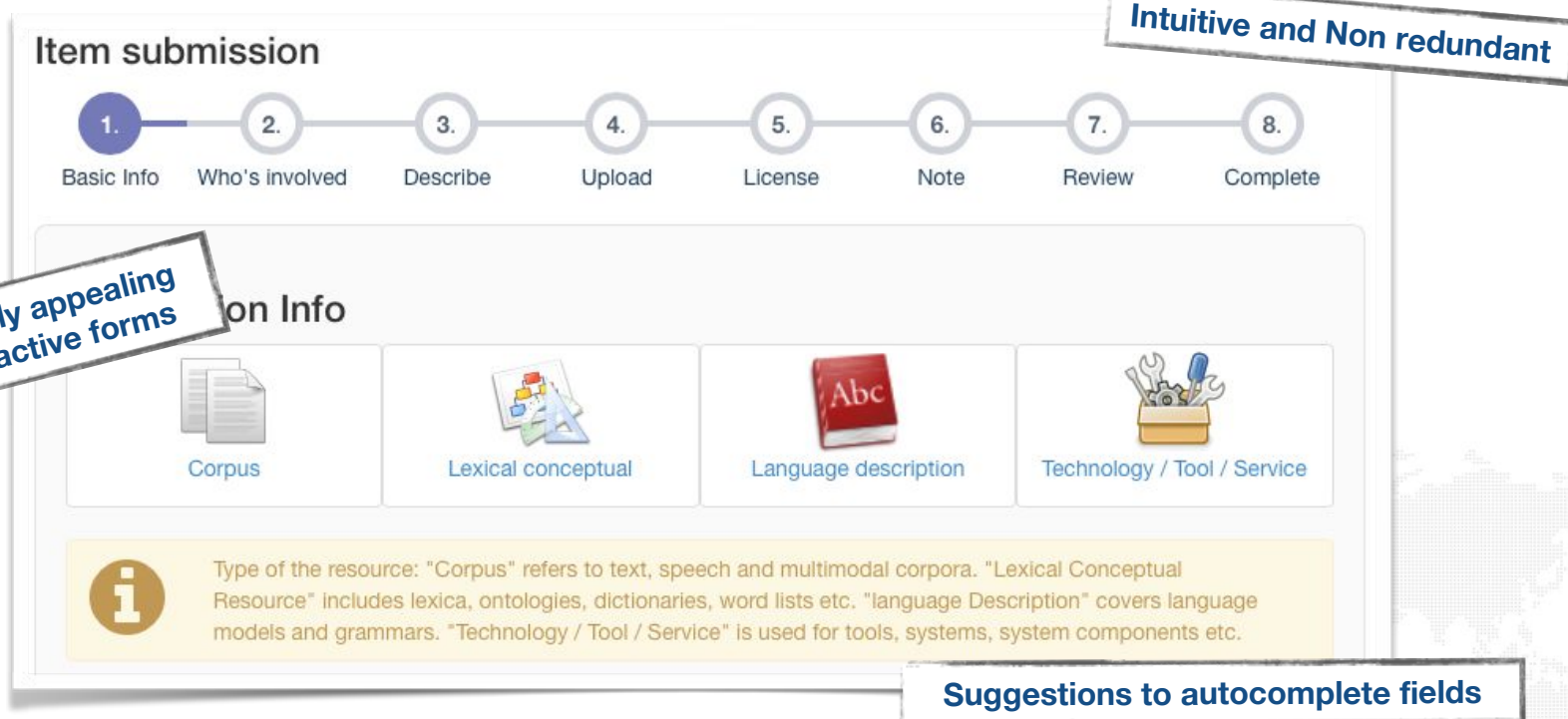
Share link

To pass your submission to another user give them the following link: https://lindat.mff.cuni.cz/repository/xmlui/submit?workspaceID=3037&share_token=b8847556e4df6e9d464a1aead59e47b3. You will also receive an email with this link.

Easy to Deposit

Submission Process

- The DSpace submission process consists of a series of "steps", where each "step" corresponds to one or more UI pages.
- We have carefully designed the steps to make sure that a submitter provides as much metadata as possible.



Auto Complete Submission Process

Item submission

Submission Progress Bar



People and organizations involved

Authors

Last name, e.g. *Smith*

Hajič, Jan (used by #25)

Tufiš, Dan (used by #22)

Bojar, Ondřej (used by #20)

Bondi (used by #19)

Branco, António (used by #18)

Žabokrtský, Zdeněk (used by #16)

Start typing the author's last name and use autocomplete form that will suggest names. Press ESC if you don't want to use the preselected value.

Autocomplete Suggestions

OpenAIRE Integration

Submission Process

- OpenAIRE is a platform for accessing open repositories with research outputs and evaluation.

Funding

Funding type

EU

Grant no. or funding project code

HORIZON

Funding code:	EC/FP7/604965
Project name:	NUGENIA-PLUS - Preparing NUGENIA for HORIZON2020
Funding code:	EC/FP7/268363
Project name:	HM-PP - Advances in Horizontal Merger Policy and Selected Public Policy Topics
Funding code:	EC/FP7/611517
Project name:	CONECTA 2020 - COoperation and NETworking between Europe and Latin America and the Caribbean to
Funding code:	EC/FP7/246798
Project name:	HORIZONCF - New horizons in organo-fluorine chemistry

Validation and Help

Submission Process

Item submission



Resource details

Description


* Please give us a description

 Enter a description of the submitted data.

Language

* Please choose a language for the resource.

Add

 Select the language of the main content of the item. Multiple languages are possible. Start typing to see an autocomplete form that will appear if applicable.

Validation and Error Messages

Proper Help Messages

File Upload Submission Process

Support for large files

Item submission

- Basic Info
- Who's involved
- Describe
- 4. Upload**
5. License
6. Note
7. Review
8. Complete

Upload File(s)

File
Choose File no file selected

i Please enter the full path of the file on your computer corresponding to your item. If a file selection window will allow you to select the file from your computer.

When uploading language resources, please try to use one of the recommended formats: [Standards](#)

Uploading files larger than 100 MB requires special handling. Please contact [Help Desk](#) about this for more understanding.

Drop file(s) here.

Please note, that you can select licences in the License step of this submission after you upload at least one file.


File Upload

Please fill in the description(s) and hit the "Start Upload" button. Then wait till the file(s) are uploaded.

Filename: Form_A_Passport_Form.pdf
Type: application/pdf
Describe the file:

Progress: 0 bytes / 0.24 MB

File Upload Progress




Licensing

Submission Process


- The **License Step** is one of the most critical one.
- Clear and visible **Distribution License Agreement**, we want to make sure that the user is fully aware of how the deposited resources will be distributed by the repository.
- The user must also select the **Resource License** under which the resources will be distributed. We support major licenses and with our **advance licensing framework**, it is easy to add new licenses with labels and extra metadata.

Read and accept the [Distribution License Agreement](#)

Click to accept By checking this box, you agree to the Distribution License Agreement for this repository to reproduce, translate and distribute your submissions worldwide.

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

Select the resource license

 The License Selector will provide you visual assistance to select the most appropriate license for your data or software. For the list of all supported licenses and their details visit [License List Page](#).

[OPEN License Selector](#)

- OR -

If you already know under which license you want to distribute your work, please select from the dropdown below.

Select a License ...

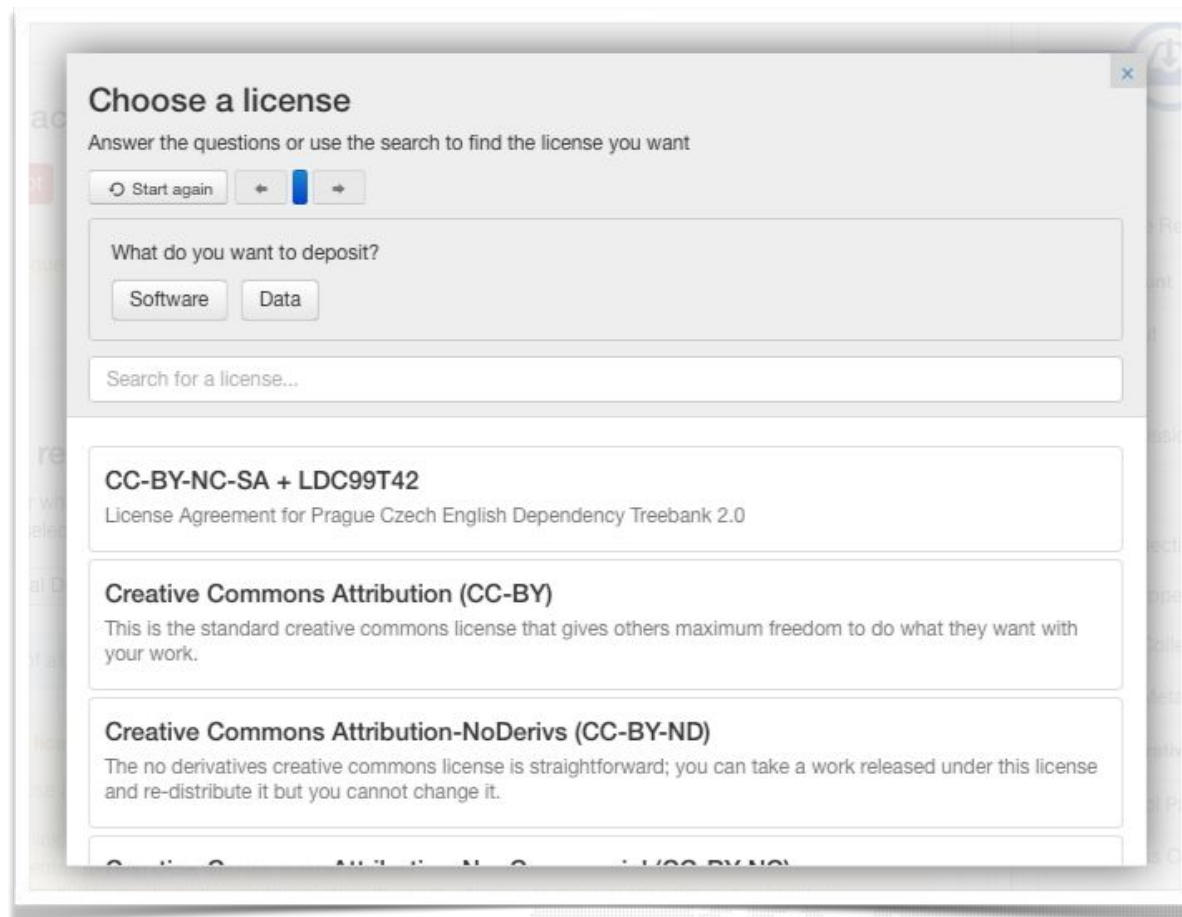


Licensing

Submission Process

To make sure that the submitter choose the correct license, we develop a **License Selector** tool.

A simple question/answer screen will guide the user to decide which license is most appropriate for his/her requirements.



The screenshot shows a web interface titled "Choose a license". It includes a "Start again" button, a progress indicator, and a question "What do you want to deposit?" with "Software" and "Data" options. Below is a search bar and a list of license options:

- CC-BY-NC-SA + LDC99T42**
License Agreement for Prague Czech English Dependency Treebank 2.0
- 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.
- Creative Commons Attribution-NoDerivs (CC-BY-ND)**
The no derivatives creative commons license is straightforward; you can take a work released under this license and re-distribute it but you cannot change it.



<https://github.com/ufal/public-license-selector>

- unified metadata profile
- reasonable set for various harvesters
- cmdi output
 - “hardcoded” profile
 - uploaded file
- OpenAIRE Data Archives
- “technical” metadata
 - provenance tracks changes, files counts/branding



- versions vs revisions
- different meaning than dspace
 - different PID
 - copy of submission
- replaces/replaced by



Acknowledgement

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: MSM 0021620825

Project name: Jazyk jako lidská činnost, její produkt a faktor

Univerzita Karlova v Praze

Project code: P10 – Lingvistika

Project name: Program rozvoje vědních oblastí na Univerzitě Karlově P10 – Lingvistika, modul Osvojování a vývoj jazykové a komunikační kompetence u populace ČR, řešeno od r. 2012

Subject(s)

youth language

classroom

language acquisition corpus

AKCES



This item is replaced by a newer submission:

<http://hdl.handle.net/11858/00-097C-0000-0023-3FBB-3>

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

- DSpace code not performing well on big files
- Tomcat not suitable for big file downloads
- avoid Tomcat for publicly available files
 - real path in metadata
 - careful because of statistics



- external (github wiki)
- internal (static repository pages)
 - user documentation
 - policies
 - processes



- issue reports
- bug fixes
- features
- DSpace (upstream)
 - xoai
 - nginx-shib/nginx-http-shibboleth,
 - openresty/headers-more-nginx-module
- our code on github
- presenting at Open Repositories conference

