

Sustainable linguistic repository built on DSpace

Jozef Mišutka
LINDAT/CLARIN dev team



Institute of Formal and Applied Linguistics
Charles University in Prague, Czech Republic



LINDAT/CLARIN

- Common Language Resources and Technology Infrastructure
- LINDAT/CLARIN = CLARIN centre in the Czech Republic
- one of the offered services is a digital repository for sharing linguistic data and tools
- must meet requirements for interoperability, AAI, preservation and sustainability



Let us improve DSpace together

- we want to take an **active** part in merging our changes if there is an **active** developer interest
- links to more information do not need to reflect the absolute latest development



Our contributions and know-how

- user interface using bootstrap
- control panel, weekly reports, logger
- submission workflow, ACL fields, drag'n'drop
- data protection law
- global user base - AAI
- licensing model
- PID - handle subsystem
- backup and replication
- metadata, export formats and QA



User Interface

- new XMLUI theme built on **bootstrap** library
 - UFALHome - home page theme
 - UFAL - general theme
- using project specific header/footer
- pages can include additional libraries using page metadata “include-library”
- deployed in January 2014

<https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/repository/revisions/master/show/sources/dspace/modules/xmlui/src/main/webapp/themes>



User Interface - home page theme

The screenshot shows the LINDAT CLARIN home page. At the top is a navigation bar with links: Home, Repository, PML TreeQuery, Tools & Services, Clarin, META-net, and Contact. Below this is a header section with a magnifying glass icon and the text "Find Linguistic Data and NLP Tools Citation Support (with Persistent IDs)", and the LINDAT CLARIN logo on the right. A search bar with a "Search" button is centered below the header. Underneath the search bar are links for "Browse" and "Advanced Search".

Author	Subject	Language (ISO)
Hajič, Jan (21)	corpus (17)	ces (65)
Bojar, Ondřej (17)	machine translation (10)	eng (28)
Pajas, Petr (11)	treebank (10)	slk (9)
Straňák, Pavel (11)	Czech (9)	hin (6)
Štěpánek, Jan (10)	parallel corpus (9)	deu (5)
... View More	... View More	... View More

Below the table is a "What's New" section. The first item is "AKCES 5 (CzeSL-SGT)" under the "Corpus" category. It lists authors: Šebesta, Karel; Bedřichová, Zuzanna; Šormová, Kateřina; Štindlová, Barbora; Hrdlička, Milan; Hrdličková, Tereza; Hana, Jiří; Petkevič, Vladimír; Jelinek, Tomáš; Škodová, Svatava; Poláčková, Marie; Janeš, Petr; Lundáková, Kateřina; Skoumalová, Hana; Sládek, Šimon; Pierscieniak, Piotr; Toufarová, Dagmar; Richter, Michal; Straka, Milan; Rosen, Alexandr. The description states: "Essays written by non-native learners of Czech, a part of AKCES/CLAC – Czech Language Acquisition Corpora. CzeSL-SGT stands for Czech as a Second Language with Spelling, Grammar and Tags. Extends the "foreign" (ciz) part ...". A note indicates "This item contains 2 files (9.8 MB)". A "Publicly Available" badge is present at the bottom right of the item.

On the right side of the page, there is a "What can you do?" section with icons for "DEPOSIT" and "CITE". Below this is a "Browse" section with a dropdown menu set to "All of the Repository". At the bottom of this section is a "My Account" section with links for "Login" and "edulD.cz Login".

User Interface - searching

LINDAT/CLARIN Repository Home > Search

[Advanced Search](#)

Limit your search

Author

Subject

Rights

PUB (92)

ACA (9)

RES (2)

Language (ISO)

Type

Contain Files

Community

Showing 1 through 10 out of 112 results

1 2 3 4 5 > 12

LanguageDescription

LINDAT / Clarin

Czech Models (Morfflex CZ + PDT) for MorphoDiTa

(Charles University in Prague, UFAL, 2013-11-12)

Author(s):

Straka, Milan; Straková, Jana

[This item contains 1 file \(67.61 MB\).](#)



Publicly Available

Corpus

LRT + Open Submissions

SYN2010: balanced corpus of written Czech

(Charles University in Prague, Faculty of Arts, Institute of the Czech National Corpus, 2010)

Author(s):

Křen, Michal; Bartoň, Tomáš; Cvrček, Václav; Hnátková, Milena; Jelínek, Tomáš; Koček, Jan; Novotná, Renata; Petkevič, Vladimír; Procházka, Pavel; Schmiedtová, Věra; Skoumalová, Hana



What can you do?



Browse

> All of the Repository

My Account

Login

edulD.cz Login

Help Desk

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

User Interface - item view

LINDAT/CLARIN Repository Home > View Item

Please use the following URI to cite or link to this item. <http://hdl.handle.net/11858/00-097C-0000-0023-95B1-E>

AKCES 5 (CzeSL-SGT)

LINDAT / Clarin Data & Tools

Authors Šebesta, Karel ; Bedřichová, Zuzanna ; Šormová, Kateřina ; Štindlová, Barbora ; Hrdlička, Milan ; Hrdličková, Tereza ; Hana, Jiří ; Petkevič, Vladimír ; Jelínek, Tomáš ; Škodová, Svatava ; Poláčková, Marie ; Janěš, Petr ; Lundáková, Kateřina ; Skoumalová, Hana ; Sládek, Šimon ; Pietscieniak, Piotr ; Toufarová, Dagmar ; Richter, Michal ; Straka, Milan ; Rosen, Alexandr

Project URL <http://utkl.ff.cuni.cz/learncorp/>

Date issued 2014-05-26

Type corpus

Language(s) ces

Description Essays written by non-native learners of Czech, a part of AKCES/CLAC – Czech Language Acquisition Corpora. CzeSL-SGT stands for Czech as a Second Language with Spelling, Grammar and Tags. Extends the “foreign” (ciz) part of AKCES 3 (CzeSL-plain) by texts collected in 2013. Original forms and automatic corrections are tagged, lemmatized and assigned error labels. Most texts have metadata attributes (30 items) about the author and the text.


Publisher Charles University in Prague

Subject(s) learner corpus Czech as a foreign language Czech language acquisition corpora AKCES non-native speakers second language acquisition


Show full item record Export to BibTeX CMLI


Files in this item Download all files in item (9.8 MB)

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

Name	2014-czesl-sgt-en.zip	
Size	9.7 MB	
Format	application/zip	
Description	corpus data and metadata, zipped	

What can you do?

DEPOSIT 

CITE 

Browse

- > All of the Repository

My Account

- > My Exports
- Logout
- Profile
- Help Desk
- Submissions


Context

- > Edit this item
- > Export item
- > Export Metadata

Administrative

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

User Interface - mobile



Search

[Browse](#) | [Advanced Search](#)

Author

- Hajič, Jan (21)
- Bojar, Ondřej (17)
- Pajas, Petr (11)
- Straňák, Pavel (11)
- Štěpánek, Jan (10)
- ... View More

Subject

- corpus (17)
- machine translation (10)
- treebank (10)

Corpus LINDAT / Clarin Data & Tools

AKCES 5 (CzeSL-SGT)

Author(s):
Šebesta, Karel; Bedřichová, Zuzanna; Šormová, Kateřina; Štindlová, Barbora; Hrdlička, Milan; Hrdličková, Tereza; Hana, Jiří; Petkevič, Vladimír; Jelínek, Tomáš; Škodová, Svatava; Poláčková, Marie; Janeš, Petr; Lundáková, Kateřina; Skoumalová, Hana; Sládek, Šimon; Pierscieniak, Piotr; Toufarová, Dagmar; Richter, Michal; Straka, Milan; Rosen, Alexandr

Description:
Essays written by non-native learners of Czech, a part of AKCES/CLAC – Czech Language Acquisition Corpora. CzeSL-SGT stands for Czech as a Second Language with Spelling, Grammar and Tags. Extends the "foreign" (ciz) part ...

This item contains 2 files (9.8 MB).

Publicly Available 

Corpus LINDAT / Clarin Data & Tools

HamleDT 2.0

Author(s):
Zeman, Daniel; Mareček, David; Mašek, Jan; Popel, Martin; Ramasamy,

Dependency Treebank).

Publisher
Charles University in Prague, UFAL


Subject(s)
MorphoDiTa Czech morphological analysis morphological generation PoS tagging

[Show full item record](#)

Export to [BibTeX](#) [CMDI](#)

Files in this item

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




Name
czech-morfflex-pdt-131112.zip

Size
67.61 MB

Format
application/zip

Description
Czech Models (MorFFlex CZ + PDT) for MorphoDiTa

[Download file](#)



User Interface - replaced by

Czech Parliament Meetings

 Authors	Pražák, Aleš ; Šmídl, Luboš
 Date issued	2011-12-15
 Type	corpus
 Language(s)	ces
 Description	The corpus consists of recordings from the Chamber of Deputies of the Parliament of the Czech Republic. It currently consists of 88 hours of speech data, which corresponds roughly to 0.5 million tokens. The annotation process is semi-automatic, as we are able to perform the speech recognition on the data with high accuracy (over 90%) and consequently align the resulting automatic transcripts with the speech. The annotator's task is then to check the transcripts, correct errors, add proper punctuation and label speech sections with information about the speaker. The resulting corpus is therefore suitable for both acoustic model training for ASR purposes and training of speaker identification and/or verification systems.
 Subject(s)	speech corpus acoustic model speaker identification and verification

 This item is replaced by a newer submission:
<http://hdl.handle.net/11858/00-097C-0000-0005-CF9C-4>
Please refer to the submission above for the latest available data. If you nevertheless need the original data, please click [here](#).

[Show full item record](#)

 Export to [BibTeX](#) [CMDI](#)

-  Brov
- > All i
-  My /
- > My
-  L
-  P
-  F
-  S
-  Con
- > Edi
- > Exp
- > Exp
-  Adm

User Interface - embargo

[Home](#)

Repository

[PML TreeQuery](#)

[Tools & Services](#)

[Clarín](#)

[META-net](#)

[LINDAT/CLARIN Repository Home](#) > [Choose Login](#)

The file is restricted

The file you are attempting to access is a restricted file. The author placed an embargo date to the associated bitstream(s) which means that the bitstreams will be available in future. You can either try to contact the author directly or use our [Help Desk](#).

Available after (year-month-day): 2020-09-12



User Interface - eperson list

LINDAT/CLARIN Repository Home > Manage E-people

E-person management

Actions

Create a new E-Person: [Click here to add a new E-Person.](#)
Browse E-People: [Click here to browse all E-People.](#)

Search for E-People:

Showing 1 thro

1 2 3 4 5 >

Del	ID	Name	Email	More information
	1	Kar	va z	Can login Password set Not logged in yet Cannot Delete Submitter of item(s) netid: null
	2	Kar	kar cz	Cannot login Password not set Not logged in yet Cannot Delete Submitter of item(s) netid: null
	3	Joz	jr	Can login Password set Last login: Thu Jun 12 20:55:35 CEST 2014

Control Panel

- truly **control** and inspect DSpace
- no more if-else and hard to debug code
- enumerate classes representing tabs in controlpanel.cfg
- extensible and easily configurable
- support for asynchronous “checks”

<https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/wiki/Controlpanel>



Control Panel implementation

- controlpanel.cfg

```
controlpanel.tabs = Test Panel
plugin.named.....administrative.ControlPanelTab = \
  cz.cuni.mff.ufal.ControlPanelTestTab = Test Panel
```

- ControlPanelTestTab.java

```
public class ControlPanelTestTab extends AbstractControlPanelTab {
  @Override
  public void addBody(Map objectModel, Division div) throws WingException {
    div.addPara("This is a tab.");
  }
}
```



Control Panel

🏠 Home **Repository** PML TreeQuery Tools & Services Clarin META-net Contact

LINDAT/CLARIN Repository Home > Control panel

Control Panel



[Java Information](#) [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](#)

Use the tabs above to select the information to display

DEPOSIT  

 Browse

> All of the Repository

● My Account



Control Panel - programs

Control Panel

[Java Information](#) [Configuration](#) [SystemWide Alerts](#) [Programs](#) [PID](#) [Shibboleth](#) [Backup](#) [IRODs Replic](#)
[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 Item](#)
[Bitstream items](#) [Specific Metadata](#) [Metadata Quality](#) [Embargoed items](#) [Oldest users](#)

These programs are usually called from installation bin directory 'bin/dspace' and the appropriate classes are taken from solr 'dspace/config/launcher.xml'

Standalone programs to run

SOLR backend (discovery) (update-discovery-index):

Clean then reindex solr (-f)

Update solr (without clean!) (-b)

Clean (force) solr (-c -f)

Optimise solr (you sure?) (-o)

Searching postgresql index (sort of solr facets cache?):

Create tables (no index) (-t)

Index (-i)

Clean and reindex (-r -f)

Control Panel - backup

Control Panel

[Java Information](#) [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](#)

Available Backups

Returned stdout:

backup2l v1.5 by Gundolf Kiefer

Summary

=====

	Backup	Date	Time	Size	Skipped	Files+D	New	Obs.	Err.
	all.1	2014-06-03	02:00	111.7G	0	23078	23078	0	72
	all.11	2014-06-04	02:00	21.1M	0	23083	121	116	0
	all.12	2014-06-05	02:00	21.2M	0	23090	117	110	0
	all.13	2014-06-06	02:00	21.2M	0	23083	96	103	0
	all.14	2014-06-07	02:00	21.2M	0	23114	114	83	0



Control Panel - replication

Control Panel

[Java Information](#) [Configuration](#) [SystemWide Alerts](#) [Programs](#) [PID](#) [Shibboleth](#) [Backup](#)
[Cron Jobs](#) [OAIPMH Validators](#) [Harvesting](#) [Release Notes](#) [Statistics](#) [Licenses](#) [Signed Lic](#)
[Current Activity](#) [Checks](#) [Verify Logging](#) [Dspace Log\(s\)](#) [User Logins](#) [Shib Raw Logins](#) [Un](#)
[Bitstream items](#) [Specific Metadata](#) [Metadata Quality](#) [Embargoed items](#) [Oldest users](#)

Execute

[Configuration](#)

[Home Dir](#)

[Replicas](#)

[To Replicate](#)

[Not Replicated](#)

[Turn on/off](#)

[Replicate missing \(async.\)](#)

i Number of items to replicate, use with caution as it may be resource intensive

[Replicate specific handle](#)

i E.g., 11858/00-097C-0000-000D-F696-9

[Delete remote path](#)

i e.g., ./dspace_1.8.2/11858_00-097Z-0000-0022-E46E-8.zip or empty when checkboxes are use

Control Panel - replication

Control Panel

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

Execute

Configuration

Home Dir

Replicas

To Replicate

Not Replicated

Turn on/off

Replicate missing (async.)

i Number of items to replicate, use with caution as it may be resource intensive

Replicate specific handle

i E.g., 11858/00-097C-0000-000D-F696-9

Delete remote path

i e.g., ./dspace_1.8.2/11858_00-097Z-0000-0022-E46E-8.zip or empty when checkboxes are use

Control Panel - replicas

Download replica

i e.g., /dspace_1.8.2/11858_00-097Z-0000-0022-E46E-8.zip

Replicated files [#90 / 74 GB]

#	STATUS	ITEM	SIZE REPLICA/ORIG	INFO	Delete
1	✔	/dspace_1.8.2_ufal_point/11858_00-097C-0000-0023-95B1-E.zip	9 MB / 9 MB	[ADMIN_Status : Archive_ok] [EUDAT_PID : /1 e3-877c-525400276111] [EUDAT_ROR : http://hdl.handle.net/11858/00-097C-0000-0023-95B1-E] [INFO_Checksum : 3c69c07: 20160095ca1] [INFO_TimeOfTransfer : 2014-05-26.15:46:03] [OTHER_AckEmail : lindat-ff.cuni.cz] [OTHER_From : https://lindat.mff.cuni.cz/repository/xmlui [OTHER_original_checksum : 3c69c0i: 160095ca1] [OTHER_original_filesize : 10288059]	●
2	✔	/dspace_1.8.2_ufal_point/11858_00-097C-0000-0023-466F-C.zip	95 MB / 95 MB	[ADMIN_Status : Archive_ok] [EUDAT_PID : /1 e3-a652-525400276111] [EUDAT_ROR : http://hdl.handle.net/11858/00-097C-0000-0023-466F-C]	●

- Lic
- Ha
- i Gener
- De
- Cit
- Sul
- FAC
- i Abc



Control Panel - signed licences

LINDAT/CLARIN Repository Home > Control panel

Search

Control Panel

Java Information 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

SIGNED LICENSES (showing only first #50)

ID	DATE	USERNAME	LICENSE	ITEM	BITSTR EAM	EXTRA METADATA
2010204	2014-06-12 10:20:48	129 [lo fal. mi	CC-BY-NC-SA + LDC99T42	136	476	IP: 193
2003176	2014-05-08 01:10:34	136 [112 icdc onn	CC-BY-NC-SA + LDC99T42	136	476	IP: 195
201632	2014-04-14 03:16:32	136 [112 :dc onn	CC-BY-NC-SA + LDC99T42	136	476	IP: 89
100501	2014-06-10 05:07:52	90 [rr :z]	Czech National Corpus (Shuffled Corpus Data)	195	698	IP: 195

Control Panel - log parsing


Control Panel

[Java Information](#) [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](#)

Choose different file ▾

 File: [dspace.log.2014-06-12] Warnings/Errors: [17]

 File: [cocoon.log.2014-06-12] Warnings/Errors: [29]

 File: [dspace.ufal.metashare-schema-errors.log.2014-06-12] Warnings/Errors: [1]

File: [authentication.log.2014-06-12] Warnings/Errors: [0]

Control Panel - unpublished items

[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](#)

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
321	258	4	str [redacted] ni.cz	Not Submitted	2014-06-10 14:30:29.794
322	259	1	jrr [redacted] om	Not Submitted	2014-06-05



Control Panel - bitstream list

[Current Activity](#) [Checks](#) [Verify Logging](#) [Dspace Log\(s\)](#) [User Logins](#) [Site Raw Logins](#) [Unpublished Items](#)
[Bitstream items](#) [Specific Metadata](#) [Metadata Quality](#) [Embargoed items](#) [Oldest users](#)

BITSTREAMS

Show not deleted unknown

Showing 1 through 10 out of 539 results

1 2 3 4 5 > 54



ID	NAME	DESCRIPTION	SIZE (MB)	FORMAT	DELETED
1			0.00 MB	Unknown	true
10	license.txt		0.00 MB	License	false
11	data.zip	The annotation of PTD with CWN synsets	2.44 MB	zip	false
21	Czech_WordNet_1.9_PDT.zip		0.43 MB	zip	false
24	CoNLL2009-ST-Czech-trial.zip		0.06 MB	zip	false



Control Panel - specific metadata

Bitstream items **Specific Metadata** Metadata Quality Embargoed items Oldest users

IS REPLACED BY ITEMS (dc.relation.isreplacedby)

#	NAME	VALUE
1	Czech Parliament Meetings	http://hdl.handle.net/11858/00-097C-0000-0005-CF9C-4
2	Prague Dependency Treebank 2.0 (PDT 2.0)	http://hdl.handle.net/11858/00-097C-0000-0006-DB11-8
3	Dspace modifications for use of EPIC handles	http://hdl.handle.net/11858/00-097C-0000-0023-4087-6
4	AKCES 2	http://hdl.handle.net/11858/00-097C-0000-0023-3FBB-3
5	Czech Named Entity Corpus 1.0	http://hdl.handle.net/11858/00-097C-0000-0023-1B04-C
6	Prague Dependency Treebank 2.5	http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3
7	Morče	http://hdl.handle.net/11858/00-097C-0000-0023-43CD-0

RELATION (dc.relation.*)

#	NAME	relation	VALUE
1	Prague Dependency Treebank 2.5	dc.relation.isbase	http://hdl.handle.net/11858/00-097C-0000-0004-0000-5

© Browse
> All of t
My Acc
> My Ex
Log
Profi
Help
Sub
Admini
Acc
Reg
Item
Coll
With
Con
Stati
Imp
Cur

Control Panel - metadata quality

Metadata QA

Process up to: 1000

Select metadata field:

- dc.contributor.advisor
- dc.contributor.author**
- dc.contributor.editor
- dc.contributor.illustrator
- dc.contributor.other
- dc.contributor
- dc.coverage.placeName
- dc.coverage.spatial
- dc.coverage.temporal
- dc.creator
- dc.date.accessioned
- dc.date.available
- dc.date.copyright
- dc.date.created
- dc.date.issued
- dc.date.submitted
- dc.date.updated
- dc.date
- dc.description.abstract
- dc.description.embargo

Show metadata values

Occurrences in items

Hajič, Jan:	149 / 469	<input type="text" value="Hajič, Jan"/>	<input type="button" value="Update"/>
Skoumalová, Hana:	105 / 469	<input type="text" value="Skoumalová, Hana"/>	<input type="button" value="Update"/>



Weekly reports

- cron based email report on the internals and statistics of a DSpace repository including the results of curation tasks and integrity checking
- not only for administrators

<https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/repository/visions/master/entry/sources/dspace-api/src/main/java/cz/cuni/mff/ufal/Report.java>



Logger

- crucial objects should use a logger that also sends emails when an exception is logged and not fail silently
 - still have to filter out “insignificant” ones
- trivial but proved very useful
- replace `Logger.getLogger` with `yourLogger.getLogger`

<https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/repository/revisions/master/entry/sources/dspace-api/src/main/java/cz/cuni/mff/ufal/Logger.java>



Licensing

- create new licences and define restrictions
- attach licences to submissions (bitstreams) in submission workflow
- require authentication and user signature with additional information if necessary
- send download link if necessary

https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/wiki/License_administration



Licensing

Home **Repository** PML TreeQuery Tools & Services Clarin META-net

LINDAT/CLARIN Repository Home > Licenses

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"/>	Czech National Corpus (Shuffled Corpus Data)	http://ucnk.ff.cuni.cz/doc/CNC_license.php	Ask only once		ACA	BY NC	9
<input type="checkbox"/>	CC-BY-NC-SA + LDC99T42	http://ufal.mff.cuni.cz/pcedt2.0/en/pcedt_license.html	Ask only once		RES	CC BY SA NC	3
<input type="checkbox"/>	BSD 3-Clause "New" or "Revised" license	http://opensource.org/licenses/BSD-3-Clause	Not required		PUB	BY BSD	2
<input type="checkbox"/>	BSD 2-Clause "Simplified" or "FreeBSD" license	http://opensource.org/licenses/BSD-2-Clause	Not required		PUB	BY BSD	5
<input type="checkbox"/>	Artistic License (Perl) 1.0	http://opensource.org/licenses/Artistic-Perl-1.0	Not required		PUB	OSI	1
<input type="checkbox"/>	Artistic License 2.0	http://opensource.org/licenses/Artistic-2.0	Not required		PUB	OSI	1
<input type="checkbox"/>	GNU General Public Licence, version 3	http://opensource.org/licenses/GPL-3.0	Not required		PUB	GPLv3	9

Licensing

LINDAT/CLARIN Repository Home > Item > Licence Agreement

Search

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.

Q CC-BY-NC-SA + LDC99T42

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	<input type="text"/>
User ID	<input type="text" value="mis"/> <input type="text" value="z"/>
Item handle	<input type="text" value="11858/00-097C-0000-0015-8DA"/>
Bitstream	<input type="text" value="PCEDT-2-full-DVD.zip"/>
IP Address	<input type="text" value="192.130.197.20"/>

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

? What can you do?



© Browse

> All of the Repository

My Account

Logout

Profile

Help Desk

Submissions

i General Information

Deposit

Quote Cite

Submission Lifecycle

? FAQ

! About



Submission Workflow

- new UI
- suggest metadata values from other submissions
- autocomplete for vocabularies
- user can leave us a note
- metadata field permission

<https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/wiki/FineGrainedFieldPermissions>



SW - suggest & autocomplete

The screenshot displays the 'Item submission' page of the LINDAT/CLARIN Repository. At the top, a navigation bar includes 'Home', 'Repository', 'PML TreeQuery', 'Tools & Services', 'Clarín', 'META-net', and 'Contact'. A user profile 'Profile: Jozef Miš' is visible in the top right. Below the navigation bar, a breadcrumb trail shows 'LINDAT/CLARIN Repository Home > Item submission'. A search bar is located on the right side of the page.

The main content area features a progress bar for the submission process, consisting of eight steps: 1. Item Type (completed), 2. Describe (active), 3. Upload, 4. License, 5. Details, 6. Note, 7. Review, and 8. Complete.

The 'Describe Item' section contains a 'Title' input field with a placeholder 'Enter the main title of the item.' Below it is an 'Authors' section with an input field containing 'Hajič'. An autocomplete dropdown menu is open, showing two suggestions: 'Hajič, Jan (used by #21)' and 'Hajičová, Eva (used by #4)'. To the right of the input field are two labels: 'Last name, e.g. Smith' and 'First name(s) + "Jr", e.g. Donald Jr'. A help icon and text below the authors section state: 'Enter the names of the authors of this item. Start typing the author's last name and use autocomplete form that will appear if applicable. End your input by pressing ESC if you don't want to use the preselected value.'

On the right side of the page, there is a sidebar with a search bar and a 'What can you do?' section. This section includes a 'DEPOSIT' button with a document icon, a 'CITE' button with a quote icon, and a 'Browse' section with a dropdown menu set to '> All of the Repository'. Below this are 'My Account' options: 'Logout', 'Profile', 'Help Desk', and 'Submissions'. At the bottom of the sidebar is a 'Context' section with options: '> Edit Collection', '> Item Mapper', '> Export Collection', and '> Export Metadata'. Finally, there is an 'Administrative' section with a gear icon.

SW - drag'n'drop

The screenshot shows the 'Item submission' page in the LINDAT/CLARIN Repository. The page has a navigation bar with links: Home, Repository, PML TreeQuery, Tools & Services, Clarin, META-net, and Contact. Below the navigation bar, there is a breadcrumb trail: LINDAT/CLARIN Repository Home > Item submission. The main content area is titled 'Item submission' and features a progress indicator with three steps: 'Item Type' (checked), 'Describe' (checked), and 'Upload' (active). Below the progress indicator is the 'Upload File(s)' section, which includes a 'File' upload button and a text input field. A modal dialog titled 'File upload' is open, displaying details for two files being uploaded. The first file is 'OR2014.misutka.pdf' (application/pdf) with a progress of 0 bytes / 0.29 MB. The second file is 'misutka.dspace.pptx' (application/vnd.openxmlformats-officedocument.presentationml.presentation) with a progress of 0 bytes / 0.07 MB. A 'Start Upload' button is located at the bottom right of the modal. The right sidebar contains a search bar, a 'DEPOSIT' button, and a list of navigation options: 'Browse', 'My Account', 'Context', and 'Administrative'.

Home Repository PML TreeQuery Tools & Services Clarin META-net Contact

LINDAT/CLARIN Repository Home > Item submission

Search

What can you do?

DEPOSIT

Item submission

Item Type Describe Upload

Upload File(s)

File

Vybrat súbory Nie je vybratý žiadny súbor

Please enter the full path of the file to allow you to select the file from your local system

When uploading language resource files, please use the following naming convention: [Standards](#)

Uploading files larger than 2 GB is not supported. Thank you for your understanding.

Drop file(s) here.

File upload

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

Filename: OR2014.misutka.pdf
Type: application/pdf
Describe the file:

Progress: 0 bytes / 0.29 MB

Filename: misutka.dspace.pptx
Type: application/vnd.openxmlformats-officedocument.presentationml.presentation
Describe the file:

Progress: 0 bytes / 0.07 MB

Start Upload

Browse

All of the Repository

My Account

Logout

Profile

Help Desk

Submissions

Context

Edit Collection

Item Mapper

Export Collection

Export Metadata

Administrative

SW - field permission

Home Repository PML TreeQuery Tools & Services Clarin META-net Contact

LINDAT/CLARIN Repository Home > Item submission

Search

Item submission

1. Item Type 2. Describe 3. Upload 4. License 5. Details 6. Note 7. Review 8. Complete

Even More Information

Special fields

Replaces (Admin only field)

Add

A related resource that is supplanted, displaced, or superseded by the described resource.

Is replaced by (Admin only field)

Add

A related resource that supplants, displaces, or supersedes the described resource.

What can you do?

DEPOSIT

Browse

> All of the Repository

My Account

Logout

Profile

Help Desk

Submissions

Context

> Edit Collection

> Item Mapper

> Export Collection

> Export Metadata

Administrative

Global AAI

- shibboleth
 - part of NF, CLARIN's SPF, eduGAIN
- local users as fallback mechanism
- reimplementing of shibboleth in DSpace
 - we accept more IDs and **require an email**
- CP integration
- IdP selection in a separate subproject
 - our discovery script used by discojuice



Global AAI

The screenshot displays the LINDAT/CLARIN Repository website's navigation bar with links for Home, Repository, PML TreeQuery, Tools & Services, Clarin, META-net, and Contact. The main heading is "New user registration", accompanied by a progress indicator showing three steps, with the first step being active. Below the heading, a registration form is visible, featuring an "Email Address" input field and a "Register" button. A note states: "This address will be verified and used as your login name." To the right, a "DEPOSIT" button is present. An overlay window titled "Sign in to LINDAT/CLARIN Repository" is open, showing a list of identity providers (IdPs) such as "Univerzita Karlova v Praze", "RedCLARA", "GÉANT Identity Provider", "Clarin.eu website account", "CESNET, a. i. e.", "FN Pízen", "University of West Bohemia", "Jan Evangelista Purkyně University in Ústí nad Labem", and "Institute of Chemical Technology Prague". A search bar at the bottom of the overlay allows for finding providers by name.



Privacy concerns - data protection

Welcome to LINDAT/CLARIN Repository

Privacy Statement

The information released by your IdP (home organisation) is shown below.

shib-identity-provider:	https://idp.csc.fi/idp/shibboleth
shib-session-index:	261ff 78441287f266z
shib-session-id:	_96al ca3bl
shib-application-id:	default
sn:	Matt
shib-authentication-instant:	2014-01-13T09:58:01.160Z
eppn:	matl
shib-authncontext-class:	urn:oasis:names:tc:SAML:2.0:ac
persistent-id:	Gfg3Si/xUwYb
cn:	Ma
shib-authentication-method:	urn:oasis:names:tc:SAML:2.0:ac
mail:	m

We use only the required attributes as stated in <http://lin>
However, we may log the attributes and keep them for a

[Return to Home](#)

[LINDAT/CLARIN Repository Home](#) > [Control panel](#)

Control Panel

- [Java Information](#)
- [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](#)

Last Login (showing only the first #200)

58.	Firstname Lastname113	user113@nowhere.no	Wed Dec 18 16:13:40 CET 2013
59.	Firstname Lastname114	user114@nowhere.no	Wed Jan 08 16:07:14 CET 2014
60.	Firstname Lastname81	user81@nowhere.no	Thu Jan 09 10:05:14 CET 2014
61.	Firstname Lastname115	user115@nowhere.no	Thu Jan 09 11:54:42 CET 2014
62.	Firstname Lastname37	user37@nowhere.no	Fri Jan 10 21:40:36 CET 2014

PIDs

- LINDAT/CLARIN uses handles
 - local (more prefixes on one instance)
 - remote through EPIC using REST API
- we support external handles too
 - not connected to DSpace Item but to URL

https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/wiki/New_handle_functionality_in_DSpace



PIDs

Handle List

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

Showing 1 through 5 out of 134 results

1 2 3 4 5 > 27


	Handle	Internal	URL	Resource type	Resource ID
<input type="radio"/>	11858/00-097C-0000-0001-486F-D	Yes	https://ufal-point-dev.ms.mff.cuni.cz/demo/xmlui/handle/11858/00-097C-0000-0001-486F-D	COMMUNITY	1
<input type="radio"/>	11858/00-097C-0000-0001-4870-7	No	http://hdl.handle.net/11858/00-097C-0000-0001-4870-7		
<input type="radio"/>	11858/00-097C-0000-0001-4872-3	Yes	https://ufal-point-dev.ms.mff.cuni.cz/demo/xmlui/handle/11858/00-097C-0000-0001-4872-3	ITEM	3
<input type="radio"/>	11858/00-097C-0000-0001-4877-A	Yes	https://ufal-point-dev.ms.mff.cuni.cz/demo/xmlui/handle/11858/00-097C-0000-0001-4877-A	COLLECTION	2
<input type="radio"/>	11858/00-097C-0000-0001-487A-4	Yes	https://ufal-point-dev.ms.mff.cuni.cz/demo/xmlui/handle/11858/00-097C-0000-0001-487A-4	ITEM	4

1 2 3 4 5 > 27


New external handle


Edit handle


Delete handle

 Browse


> All of the Re

 My Account

 Logout


 Profile

 Help Desk

 Submission

 Administrative

 Access Control

 Registries

 Items

 Collection

 Withdrawn

 Control Panel

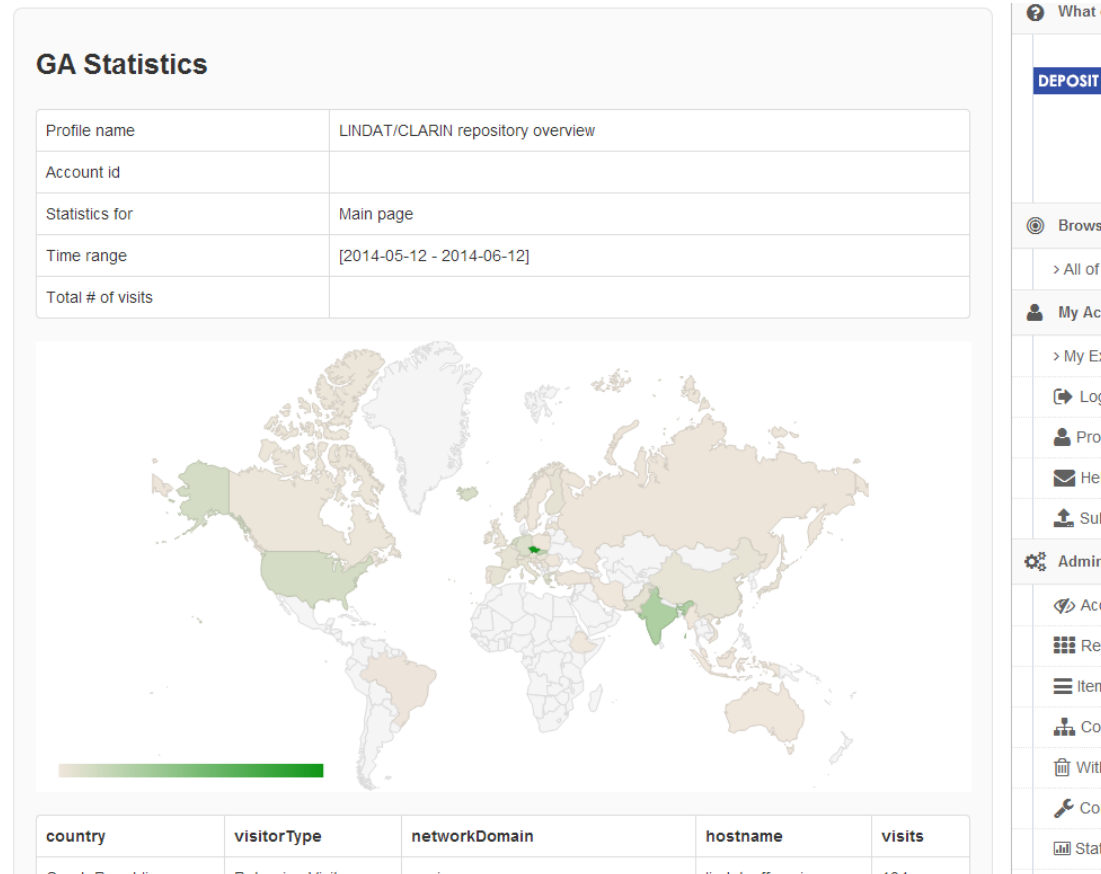
 Statistics

 Import Metadata

 Curation

 Licenses

GA Statistics



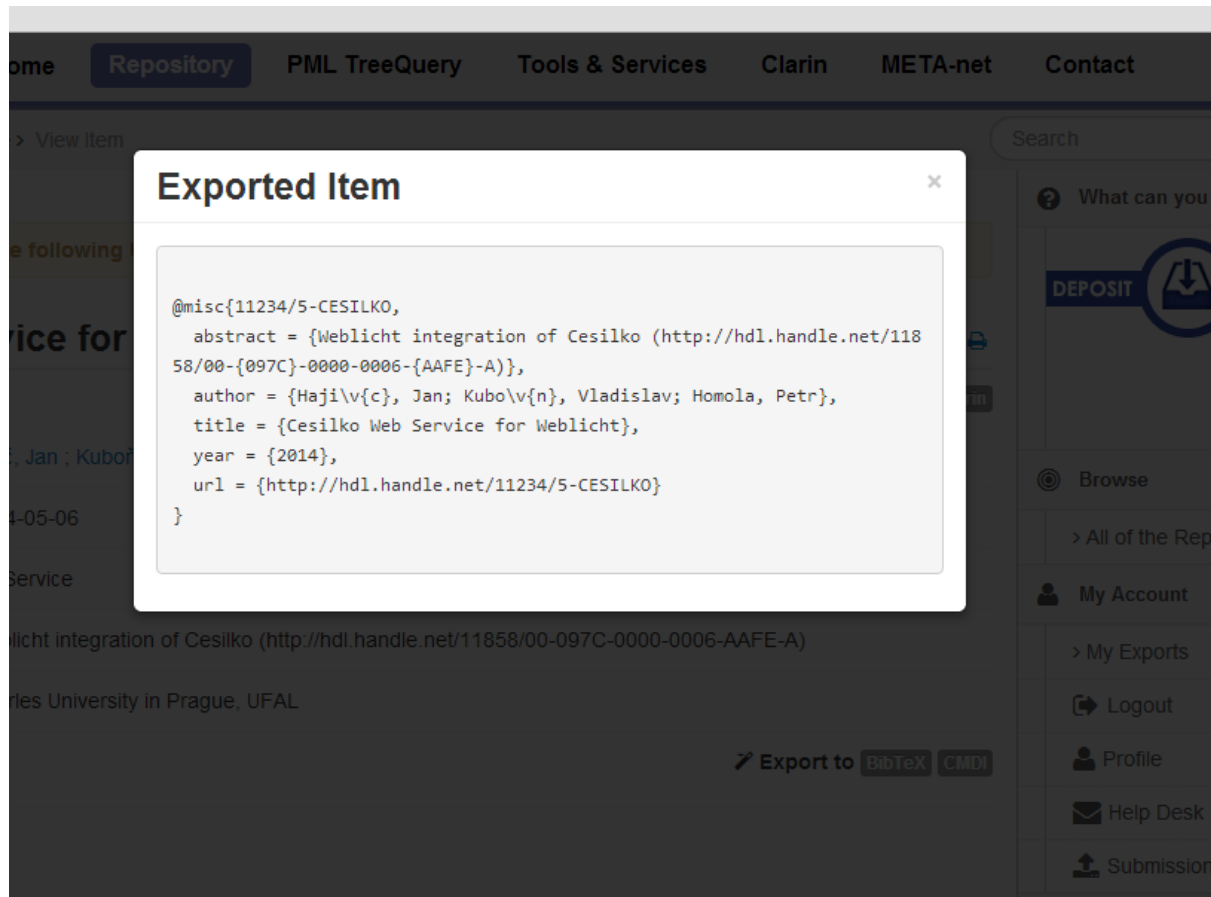
<https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/repository/changes/sources/dspace-xmlui/dspace-xmlui-api/src/main/java/cz/cuni/mff/ufal/dspace/app/xmlui/aspect/statistics/GAStatisticsTransformer.java?rev=master>

Metadata and dissemination

- basic support of hierarchical metadata - CMDI
- use of local.* metadata not shown to user
- provenance for every change
- exporting to CMDI/bibtex using stripped version of oai processor
 - available at dspace.url/oaistripped



Export to bibtex in Item View



The screenshot shows a web interface with a dark header containing navigation links: 'Repository', 'PML TreeQuery', 'Tools & Services', 'Clarín', 'META-net', and 'Contact'. A search bar is visible on the right. A central dialog box titled 'Exported Item' displays the following BibTeX code:

```
@misc{11234/5-CESILKO,  
  abstract = {Weblicht integration of Cesilko (http://hdl.handle.net/118  
58/00-097C-0000-0006-AAFE-A)},  
  author = {Haji\v{c}, Jan; Kubo\v{n}, Vladislav; Homola, Petr},  
  title = {Cesilko Web Service for Weblicht},  
  year = {2014},  
  url = {http://hdl.handle.net/11234/5-CESILKO}  
}
```

Below the dialog box, the background interface shows a 'DEPOSIT' button with a download icon, a 'Browse' section with a link to 'All of the Repo...', and a 'My Account' section with links for 'My Exports', 'Logout', 'Profile', 'Help Desk', and 'Submissions'. At the bottom right, there are buttons for 'Export to BibTeX' and 'CMDI'.




Curation








- allow curation on Workflow objects
- run curation from CP
 - updated error reporting



Curation














LRT + Open Submissions Data & Tools



 Authors	Susanne Alt, salt@atilf.fr
 Project URL	http://freebank.loria.fr/corpus.php
 Type	corpus
 Language(s)	fra
 Description	written texts and spoken dialogues, mixed
 Publisher	ATILF
 Collection(s)	LRT + Open Submissions Data & Tools

Actions you may perform on this task:

Show item record in simple / full format	Show full item record
If you have reviewed the item and it is suitable for inclusion in the collection, select "Approve".	Approve item
Validate this item with the curation framework.	Curate item
If you have reviewed the item and found it is not suitable for inclusion in the collection, select "Reject". You will then be asked to enter a message indicating why the item is unsuitable, and whether the submitter should change something and resubmit.	Reject item
Select this option to change the item's metadata. You'll be asked to add all the required fields.	Edit metadata
Move item to another collection	Move...
Return the task to the pool so that another user may perform the task.	Return task to pool

-  Browse
 - > All of the Repository
-  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



Curation

[Home](#) **Repository** [PML TreeQuery](#) [Tools & Services](#) [C](#)

[LINDAT/CLARIN Repository Home](#) > Curation Tasks

System Curation Tasks

Handle of DSpace Object

i Hint: Enter [your-handle-prefix]/0 to run a task across entire site (not all tasks may support this capability)

Task

- Profile Bitstream Formats
- Profile Bitstream Formats**
- Check for Required Metadata
- Check Links in Metadata
- Check Item Handles
- Check Relations in Metadata

Perform **Queue** **Run all curation tasks at once**



Searching

- boost matches in title and sort according to relevancy
 - hack but really important (proxy to SOLR)
- add different facets
 - use of local.* metadata automatically generated e.g., collection name

<https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/repository/revisions/master/entry/sources/dspace-discovery/dspace-discovery-solr/src/main/java/org/dspace/discovery/SolrServiceImpl.java>



Searching

The screenshot displays the LINDAT search interface. At the top, there is a search bar with a magnifying glass icon and a 'Search' button. Below the search bar, the 'Selected filters' section shows 'Author: Hajič, Jan' and a 'Clear All' button. The 'Advanced Search' section includes an 'Add Filters' button and a search criteria field: 'Title' contains 'kh'. A blue button labeled 'Khresmoi (2)' is visible below the criteria. The 'Limit your search' section on the left has dropdown menus for 'Author', 'Subject', 'Rights', 'Language (ISO)', and 'Type'. The main results area shows 'Showing 1 through 10 out of 21 results' and a pagination control with buttons for 1, 2, 3, and a next arrow. A 'Sort items by' dropdown menu is open, listing options: Relevance (checked), Title Asc, Title Desc, Issue Date Asc, and Issue Date Desc. The first result is 'Khresmoi Query Translation Test Data 1.0' by Charles University in Prague, UFAL, 2013-10-10, with author(s) Pecina, Pavel; Dušek, Ondřej; Hajič, Jan; Urešová, Zdeňka. A link indicates 'This item contains 1 file (55.35 KB)'. On the right, a sidebar menu contains sections: 'DEPOSIT' with a deposit icon, a quote icon, 'Browse' with '> All of the Repository', 'My Account' with '> My Exports', 'Logout', 'Profile', 'Help Desk', 'Submissions', 'Administrative' with 'Access Control', 'Registries', 'Items', 'Collections & Communities', 'Withdrawn Items', 'Control Panel', and 'Statistics'.

Demo available

- <http://hdl.handle.net/11234/6-DEMO>
 - demo-admin
 - demo-user
- master-dev branch



Additional information

- contact us: clarin-list @ ufal.mff.cuni.cz
- issue tracker, wiki, repository
 - <http://hdl.handle.net/11234/6-PM>
- read-only git repository
 - <http://hdl.handle.net/11234/6-VCS>
- subprojects
 - <https://redmine.ms.mff.cuni.cz/projects/lindat-aa>
 - in progress - license selector



Thank you

