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

http://www.lindat.cz
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

User Interface - home page theme

http://www.lindat.cz
User Interface - searching
User Interface - item view

http://www.lindat.cz
User Interface - mobile

http://www.lindat.cz
### Czech Parliament Meetings

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Pražák, Aleš; Šmídl, Luboš</td>
</tr>
<tr>
<td>Date Issued</td>
<td>2011-12-15</td>
</tr>
<tr>
<td>Type</td>
<td>corpus</td>
</tr>
<tr>
<td>Language(s)</td>
<td>ces</td>
</tr>
<tr>
<td>Description</td>
<td>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.</td>
</tr>
<tr>
<td>Subject(s)</td>
<td>speech corpus, acoustic model, speaker identification and verification</td>
</tr>
</tbody>
</table>

This item is replaced by a newer submission: [http://hdl.handle.net/11858/00-097C-0000-0005-CF9C-4](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](http://www.lindat.cz).
User Interface - embargo

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

http://www.lindat.cz

### 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.

#### Showing 1 thru

<table>
<thead>
<tr>
<th>Del</th>
<th>ID</th>
<th>Name</th>
<th>Email</th>
<th>More information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>1</td>
<td>Kar</td>
<td>va</td>
<td>Can login, Password set, Not logged in yet, hnetc: null, Submitter of item(s)</td>
</tr>
<tr>
<td>Ø</td>
<td>2</td>
<td>Kar</td>
<td>km</td>
<td>Cannot login, Password not set, Not logged in yet, hnetc: null, Submitter of item(s)</td>
</tr>
<tr>
<td>Ø</td>
<td>3</td>
<td>Joz</td>
<td>jr</td>
<td>Can login, Password set, Last login: Thu Jun 17 20:55:35 CEST 2014</td>
</tr>
</tbody>
</table>
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”

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

Use the tabs above to select the information to display.
Control Panel - programs

Control Panel

Java Information  | Configuration  | SystemWide Alerts | Programs | PID | Shibboleth | Backup | IRODs Replica
Cron Jobs  | OA/PMH 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

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

Standalone programs to run

**SOLR backend (discovery) (update-discovery-index):**
- Clean then reindex solr (-t)
- 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)

http://www.lindat.cz
Control Panel - backup

Available Backups

Returned stdout:
backup2l v1.6 by Gundolf Kiefer

Summary

<table>
<thead>
<tr>
<th>Backup</th>
<th>Date</th>
<th>Time</th>
<th>Size</th>
<th>Skipped</th>
<th>Files+D</th>
<th>New</th>
<th>Obs.</th>
<th>Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>all.1</td>
<td>2014-06-03</td>
<td>02:00</td>
<td>111.7G</td>
<td>0</td>
<td>23076</td>
<td>23076</td>
<td>0</td>
<td>72</td>
</tr>
<tr>
<td>all.11</td>
<td>2014-06-04</td>
<td>02:00</td>
<td>21.1M</td>
<td>0</td>
<td>23083</td>
<td>121</td>
<td>116</td>
<td>0</td>
</tr>
<tr>
<td>all.12</td>
<td>2014-06-05</td>
<td>02:00</td>
<td>21.2M</td>
<td>0</td>
<td>23090</td>
<td>117</td>
<td>110</td>
<td>0</td>
</tr>
<tr>
<td>all.13</td>
<td>2014-06-06</td>
<td>02:00</td>
<td>21.2M</td>
<td>0</td>
<td>23083</td>
<td>96</td>
<td>103</td>
<td>0</td>
</tr>
<tr>
<td>all.14</td>
<td>2014-06-07</td>
<td>02:00</td>
<td>21.2M</td>
<td>0</td>
<td>23114</td>
<td>114</td>
<td>83</td>
<td>0</td>
</tr>
</tbody>
</table>

http://www.lindat.cz
Control Panel - replication

Control Panel

Java Information  Configuration  SystemWide Alerts  Programs  PID  Shibboleth  Backup  Cron Jobs  OAIPMH Validators  Harvesting  Release Notes  Statistics  Licenses  Signed License  Current Activity  Checks  Verify Logging  Dspace Log(s)  User Logins  Shib Raw Logins  Unsupervised Bibtags  Bibtags Items  Specific Metadata  Metadata Quality  Embargoed Items  Oldest Users

Execute

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Home Dir</th>
<th>Replicas</th>
<th>To Replicate</th>
<th>Not Replicated</th>
<th>Turn on/off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Replicate missing (async)

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

Replicate specific handle

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

Delete remote path

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

http://www.lindat.cz
Control Panel - replication

Control Panel

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

Execute

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Home Dir</th>
<th>Replicas</th>
<th>To Replicate</th>
<th>Not Replicated</th>
<th>Turn on/off</th>
</tr>
</thead>
</table>

Replicate missing (async.)

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

Replicate specific handle

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

Delete remote path

E.g.: /dspace_1.8.2/11858_00-097Z-0000-0022-E46E-E.zip or empty when checkboxes are use

http://www.lindat.cz
## Control Panel - replicas

### Replicated files [#90 / 74 GB]

<table>
<thead>
<tr>
<th>#</th>
<th>STATUS</th>
<th>ITEM</th>
<th>SIZE REPLICA/ORG</th>
<th>INFO</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✔️</td>
<td>/dspace_1_8.2_ufsl_point/11858_00-097C-0000-0023-95B1-E.zip</td>
<td>9 MB / 9 MB</td>
<td>[ADMIN_Status : Archive_OK] [EUDAT_PID : 18930-097C-0000-0023-95B1-E] [EUDAT_ROR : <a href="http://hdl.handle.net/11858/00-097C-0000-0023-95B1-E">http://hdl.handle.net/11858/00-097C-0000-0023-95B1-E</a>] [NINFO_Checksum : 0303007_f6c20180b000e3991a86b678e81b1e91] [INFO_TimeOfTransfer : 2014-05-26 15:49:07] [OTHER_AckEmail : <a href="mailto:lindat-IT@com.cz">lindat-IT@com.cz</a>] [OTHER_From : <a href="https://lindat.mff.cuni.cz/repository/xmls">https://lindat.mff.cuni.cz/repository/xmls</a>] [OTHER_original_checksum : 0303007_1800e3991a86b678e81b1e91] [OTHER_original_filesizze : 102800059]</td>
<td>![Delete]</td>
</tr>
<tr>
<td>2</td>
<td>✔️</td>
<td>/dspace_1_8.2_ufsl_point/11858_00-097C-0000-0023-8881-C.zip</td>
<td>95 MB / 95 MB</td>
<td>[ADMIN_Status : Archive_OK] [EUDAT_PID : 18930-097C-0000-0023-8881-C] [EUDAT_ROR : <a href="http://hdl.handle.net/11858/00-097C-0000-0023-8881-C">http://hdl.handle.net/11858/00-097C-0000-0023-8881-C</a>] [NINFO_Checksum : 0303007_f6c20180b000e3991a86b678e81b1e91] [INFO_TimeOfTransfer : 2014-05-26 15:49:07] [OTHER_AckEmail : <a href="mailto:lindat-IT@com.cz">lindat-IT@com.cz</a>] [OTHER_From : <a href="https://lindat.mff.cuni.cz/repository/xmls">https://lindat.mff.cuni.cz/repository/xmls</a>] [OTHER_original_checksum : 0303007_1800e3991a86b678e81b1e91] [OTHER_original_filesizze : 102800059]</td>
<td>![Delete]</td>
</tr>
</tbody>
</table>

http://www.lindat.cz
Control Panel - signed licences

http://www.lindat.cz

<table>
<thead>
<tr>
<th>ID</th>
<th>DATE</th>
<th>USERNAME</th>
<th>LICENSE</th>
<th>ITEM</th>
<th>BITSTREAM</th>
<th>EXTRA METADATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2014-06-12</td>
<td>[lofal.</td>
<td>CC-BY-NC-SA + LDC99T42</td>
<td>136</td>
<td>476</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>10:20:48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2014-06-08</td>
<td>[l113redeccc</td>
<td>CC-BY-NC-SA + LDC99T42</td>
<td>136</td>
<td>476</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>01:10:34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01</td>
<td>2014-04-14</td>
<td>[112dccc</td>
<td>CC-BY-NC-SA + LDC99T42</td>
<td>136</td>
<td>476</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>03:18:32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>2014-04-14</td>
<td>[112dccc</td>
<td>CC-BY-NC-SA + LDC99T42</td>
<td>136</td>
<td>476</td>
<td></td>
</tr>
<tr>
<td>632</td>
<td>03:18:32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10050</td>
<td>2014-06-10</td>
<td>[rzzczech</td>
<td>Czech National Corpus (Shuffled Corpus Data)</td>
<td>196</td>
<td>608</td>
<td>IP: 195</td>
</tr>
<tr>
<td>080</td>
<td>06:07:52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Control Panel - log parsing


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

http://www.lindat.cz
Control Panel - user logins

<table>
<thead>
<tr>
<th>User Logins</th>
<th>Shib Raw Logins</th>
<th>Unpublished Items</th>
</tr>
</thead>
</table>

| File: [authentication.log] User logins: 15 successful: 15 |
| Absolute file: log/authentication.log.2014-06-12 |

<table>
<thead>
<tr>
<th>User</th>
<th>Log Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>jni</td>
<td>!com @ 2014-06-12 00:11:38,61 =&gt; logged in successfully</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>carlicense</th>
</tr>
</thead>
</table>

| shib-identity-provider | https://cas.cuni.cz/op/shibboleth |

<table>
<thead>
<tr>
<th>manager</th>
</tr>
</thead>
</table>

| orgunit-dn | dc=mff,dc=cuni,dc=cz |

| shib-authn-context-decl |

| accept | text/html,application/xhtml+xml,application/xml; q=0.9,image/webp,/*; q=0.8 |

| street |

http://www.lindat.cz
Control Panel - unpublished items

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

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

http://www.lindat.cz
### BITSTREAMS

Showing 1 through 10 out of 539 results

<table>
<thead>
<tr>
<th>ID</th>
<th>NAME</th>
<th>DESCRIPTION</th>
<th>SIZE (MB)</th>
<th>FORMAT</th>
<th>DELETED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>license.txt</td>
<td></td>
<td>0.00 MB</td>
<td>Unknown</td>
<td>true</td>
</tr>
<tr>
<td>10</td>
<td>data.zip</td>
<td>The annotation of PTD with CWN synsets</td>
<td>0.00 MB</td>
<td>License</td>
<td>false</td>
</tr>
<tr>
<td>11</td>
<td>Czech_WordNet_1.9_PDT.zip</td>
<td></td>
<td>2.44 MB</td>
<td>zip</td>
<td>false</td>
</tr>
<tr>
<td>21</td>
<td>CoNLL2009-ST-Czecktrial.zip</td>
<td></td>
<td>0.43 MB</td>
<td>zip</td>
<td>false</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td>0.06 MB</td>
<td>zip</td>
<td>false</td>
</tr>
</tbody>
</table>

http://www.lindat.cz
Control Panel - specific metadata

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

<table>
<thead>
<tr>
<th>#</th>
<th>NAME</th>
<th>relation</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Prague Dependency Treebank 2.5</td>
<td>dc.relation.isbasedepth</td>
<td><a href="http://hdl.handle.net/11858/00-097C-0000-0004-0000-5">http://hdl.handle.net/11858/00-097C-0000-0004-0000-5</a></td>
</tr>
</tbody>
</table>
Control Panel - metadata quality
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


http://www.lindat.cz
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

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

## Manage Licenses

<table>
<thead>
<tr>
<th>License Name</th>
<th>Definition (URL)</th>
<th>Confirmation</th>
<th>Required user info</th>
<th>License Label</th>
<th>Extended Labels</th>
<th>Used by Bitstrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech National Corpus (Shuffled Corpus Data)</td>
<td><a href="http://work.ff.cuni.cz/doc/CNC_license.php">http://work.ff.cuni.cz/doc/CNC_license.php</a></td>
<td>Ask only once</td>
<td>ACA</td>
<td>BY NC</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>CC-BY-NC-SA + LDC98742</td>
<td><a href="http://ufal.mff.cuni.cz/podst2.0/ccncd_license.html">http://ufal.mff.cuni.cz/podst2.0/ccncd_license.html</a></td>
<td>Ask only once</td>
<td>RES</td>
<td>CC BY SA NC</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>BSD 3-Clause &quot;New&quot; or &quot;Revised&quot; License</td>
<td><a href="http://opensource.org/licenses/BSD-3-Clause">http://opensource.org/licenses/BSD-3-Clause</a></td>
<td>Not required</td>
<td>PUB</td>
<td>BY BSD</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>BSD 2-Clause &quot;Simplified&quot; or &quot;FreeBSD&quot; License</td>
<td><a href="http://opensource.org/licenses/BSD-2-Clause">http://opensource.org/licenses/BSD-2-Clause</a></td>
<td>Not required</td>
<td>PUB</td>
<td>BY BSD</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Artistic License (Perl) 1.0</td>
<td><a href="http://opensource.org/licenses/Artistic-Perl-1.0">http://opensource.org/licenses/Artistic-Perl-1.0</a></td>
<td>Not required</td>
<td>PUB</td>
<td>OSI</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Artistic License 2.0</td>
<td><a href="http://opensource.org/licenses/Artistic-2.0">http://opensource.org/licenses/Artistic-2.0</a></td>
<td>Not required</td>
<td>PUB</td>
<td>OSI</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>GNU General Public License, version 3</td>
<td><a href="http://opensource.org/licenses/GPL-3.0">http://opensource.org/licenses/GPL-3.0</a></td>
<td>Not required</td>
<td>PUB</td>
<td>GPLv3</td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>
Licensing

http://www.lindat.cz
Submission Workflow

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

SW - suggest & autocomplete

http://www.lindat.cz
SW - drag’n’drop
SW - field permission

Item submission

Even More Information

Special fields

Replaces (Admin only field)

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

Is replaced by (Admin only field)

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

http://www.lindat.cz
Global AAI

- shibboleth
  - part of NF, CLARIN’s SPF, eduGAIN
- local users as fallback mechanism
- reimplementaton 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

New user registration

Register an account to subscribe to collections for email updates, and submit new items to the repository.

Email Address

Register

http://www.lindat.cz
Privacy concerns - data protection

Welcome to LINDAT/CLARIN Repository

Privacy Statement

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

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>shib-identify-provider</td>
<td><a href="https://ldp.cz/idp/lsibboleth">https://ldp.cz/idp/lsibboleth</a></td>
</tr>
<tr>
<td>shib-session-index</td>
<td>2614f</td>
</tr>
<tr>
<td>shib-session-id</td>
<td>b4412b7285b8</td>
</tr>
<tr>
<td>shib-application-id</td>
<td>default</td>
</tr>
<tr>
<td>SR</td>
<td>Matt</td>
</tr>
<tr>
<td>shib-authenticationinstant</td>
<td>2014-01-13T09:58:01.186Z</td>
</tr>
<tr>
<td>cppc</td>
<td>mail</td>
</tr>
<tr>
<td>shib-authcontext-class</td>
<td><a href="mailto:umcassie@name.cs">umcassie@name.cs</a>/uwyo</td>
</tr>
<tr>
<td>persistent-id</td>
<td>Gfd35xUvwYb</td>
</tr>
<tr>
<td>crn</td>
<td>Maa</td>
</tr>
<tr>
<td>shib-authentication-method</td>
<td><a href="mailto:umcassie@name.cs">umcassie@name.cs</a>/uwyo</td>
</tr>
<tr>
<td>mail</td>
<td>mma</td>
</tr>
</tbody>
</table>

We use only the required attributes as stated in [http://lin](http://lin). However, we may log the attributes and keep them for a reasonable period of time.

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  Ospace Log(s)  User Logins  Shib Raw Logins  Unpublished Items
Bitstream items  Specific Metadata  Metadata Quality  Embargoed items

Last Login (showing only the first #200)

<table>
<thead>
<tr>
<th>No.</th>
<th>Firstname Lastname113</th>
<th><a href="mailto:user113@nowhere.no">user113@nowhere.no</a></th>
<th>Wed Dec 18 16:13:40 CET 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>58.</td>
<td>Firstname Lastname114</td>
<td><a href="mailto:user114@nowhere.no">user114@nowhere.no</a></td>
<td>Wed Jan 08 16:07:14 CET 2014</td>
</tr>
<tr>
<td>59.</td>
<td>Firstname Lastname01</td>
<td><a href="mailto:user01@nowhere.no">user01@nowhere.no</a></td>
<td>Thu Jan 09 10:05:14 CET 2014</td>
</tr>
<tr>
<td>60.</td>
<td>Firstname Lastname15</td>
<td><a href="mailto:user15@nowhere.no">user15@nowhere.no</a></td>
<td>Thu Jan 09 11:54:42 CET 2014</td>
</tr>
<tr>
<td>61.</td>
<td>Firstname Lastname37</td>
<td><a href="mailto:user37@nowhere.no">user37@nowhere.no</a></td>
<td>Fri Jan 10 21:40:36 CET 2014</td>
</tr>
</tbody>
</table>

http://www.lindat.cz
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

# PIDs

## Handle List

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

Showing 1 through 5 out of 134 results

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

http://www.lindat.cz
GA Statistics


http://www.lindat.cz
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
Curation

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

**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

http://www.lindat.cz
Curation

System Curation Tasks

Handle of DSpace Object
11858/00-097C-0000-0007-710f

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


http://www.lindat.cz
Searching

http://www.lindat.cz
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-aai
• in progress - license selector

http://www.lindat.cz
Thank you