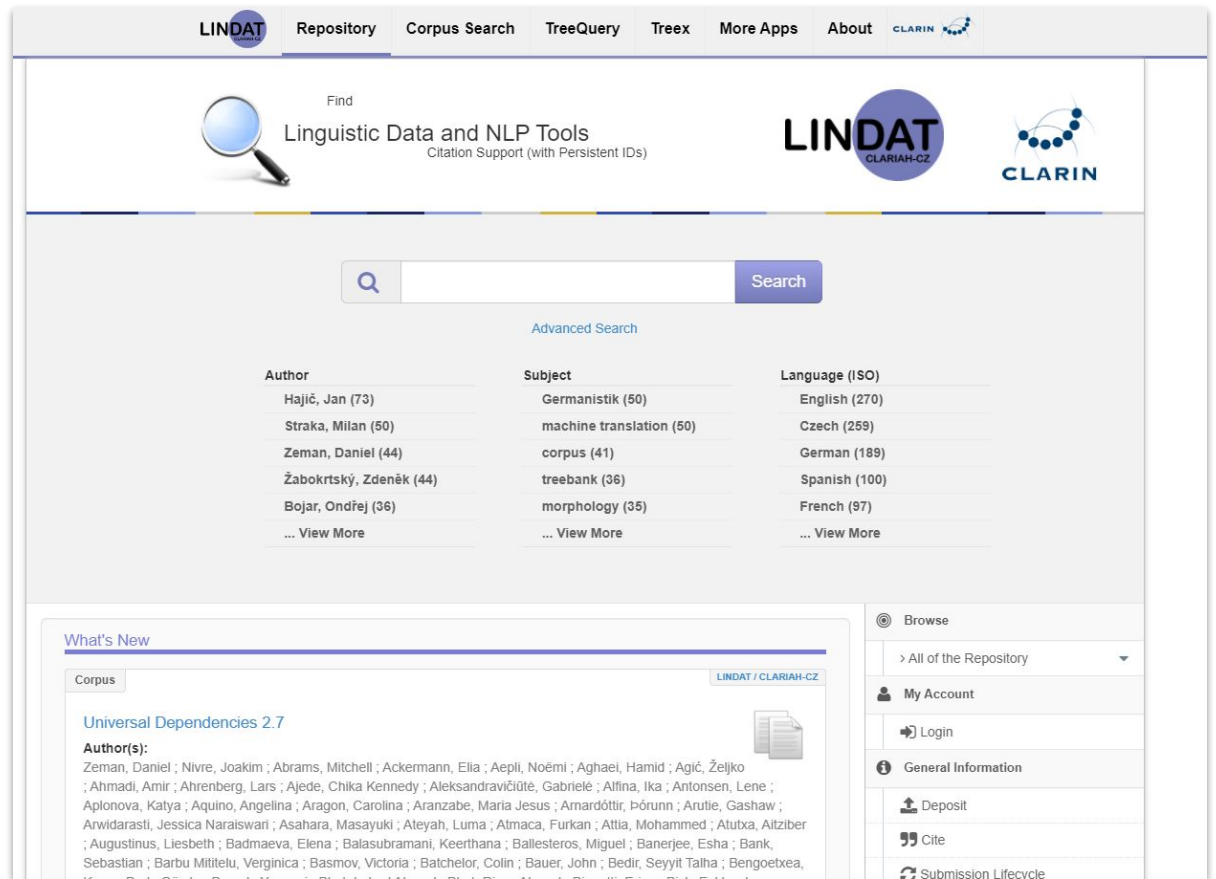


---

# CLARIN-DSpace upgrade

Jozef Mišutka, LINDAT/CLARIAH-CZ

- Czech CLARIN
  - EU infrastructure, ERIC
- **data** repository since 2011
  - based on DSpace
- domain specific for linguistics
- CoreTrustSeal certification
- CLARIN B centre certification



The screenshot shows the LINDAT/CLARIAH-CZ website interface. At the top, there is a navigation bar with links for Repository, Corpus Search, TreeQuery, Treex, More Apps, and About. The main header features the LINDAT logo and the text "Linguistic Data and NLP Tools" with "Citation Support (with Persistent IDs)" below it. A search bar is prominently displayed in the center, with a "Search" button. Below the search bar, there are three columns of data: Author, Subject, and Language (ISO). Each column lists several items with their respective counts and a "View More" link.

Author	Subject	Language (ISO)
Hajič, Jan (73)	Germanistik (50)	English (270)
Straka, Milan (50)	machine translation (50)	Czech (259)
Zeman, Daniel (44)	corpus (41)	German (189)
Žabokrtský, Zdeněk (44)	treebank (36)	Spanish (100)
Bojar, Ondřej (36)	morphology (35)	French (97)
... View More	... View More	... View More

Below the data tables, there is a "What's New" section featuring a document titled "Universal Dependencies 2.7" with a list of authors. On the right side, there is a "Browse" menu with options like "All of the Repository", "My Account", "Login", "General Information", "Deposit", "Cite", and "Submission Lifecycle".

- based on DSpace 5 (2015-2018)
  - **Security Updates:** Provided for the most recent three (3) major releases of the platform.
  - current stable release 7.2
  - <https://wiki.lyrasis.org/display/DSPACE/Releases>
- outdated technology (but still working!)
  - cocoon + xmlui + java
  - frontend

funded by

---



- supervised by **LINDAT/CLARIAH-CZ**
- partially outsourced

# upgrade analysis

---

- v5
  - only security fixes
  - no support once after DSpace 8 (vague roadmap?)
- v6
  - security fixes
  - no support once after DSpace 9
  - no new features applicable to data repositories
- v7
  - complete overhaul of the frontend
  - single frontend
  - latest technologies used - angular
  - configurable entities, video/image content streaming, IIIF, signposting, resourcesync
  - stable REST api

# upgrade plan/current state

- 9 tasks with several subtasks
  - e.g., PIDS, license information
- already started
  - proper unittests
  - up to date devops

Submission Extensions						24.1	80%		
2	Different fields for different types of submission	yes, partially	yes, complete	yes, partially	yes, partially	In DSpace8 - removed from	3.5%	3.3	77%
3	Retrieve information on complex fields	no	yes, complete	no	no		3.8%	3.6	80%
4	Hints, examples, suggestions	yes, partially	yes, complete	yes, partially	yes, partially		6.9%	6.5	80%
5	OpenAIRE	yes, partially	yes, complete	yes, partially	yes, partially		2.3%	2.6	73%
6	Sanity checks	no	yes, complete	no	yes, completely		0.3%	0.3	95%
7	Admin only fields	no	yes, complete	no	no		2.3%	2.6	80%
8	Upload CMDI file	no	yes, complete	no	no		3.5%	3.3	84%
9	User friendly settings	yes, partially	yes, complete	yes, complete	yes, partially	In DSpace7.2	0.3%	0.3	90%
10	Support files over 2GB	no	yes, complete	yes, partially	yes, completely	no	1.7%	1.6	90%
<b>Dissemination extensions</b>						<b>8.5</b>	<b>78%</b>		
1	Harvestable metadata via OAI-PMH	yes, completely	yes, complete	yes, complete	yes, completely				100%
2	TombStone	no	yes, complete	no	no		2.3%	2.6	78%
3	Item is hidden from search, harvestable only for oai	yes, partially	yes, complete	yes, partially	yes, partially		2.3%	2.6	75%
4	Harvest CMDI metadata format	no	yes, complete	no	no		1.4%	1.3	80%
5	Harvest OLAC metadata format	no	yes, complete	no	no		1.4%	1.3	75%
6	Google scholar mapping metadata	yes, completely	yes, complete	yes, complete	yes, partially	yes, completely	0.7%	0.7	90%
<b>PIDs</b>						<b>20.2</b>	<b>75%</b>		
1	PIDs associat. to metadata records	yes, completely	yes, complete	yes, complete	yes, completely		0.9%	0.8	80%
2	HTTP-accept header for content negotiation	no	yes, complete	no	no		1.7%	1.6	80%
3	Return directly cmdi	no	yes, complete	no	no		0.7%	0.7	75%
4	Config handle prefix for communities	no	yes, complete	no	no		2.3%	2.6	75%
5	Manage handle table	no	yes, complete	no	no		5.7%	5.4	80%
6	Support DOIs	yes, partially	yes, partially	yes, partially	yes, partially	In DSpace8	3.7%	9.1	70%
<b>Statistics</b>						<b>5.5</b>	<b>72%</b>		
1	Bitstream downloads	yes, partially	yes, complete	yes, partially	yes, partially		4.8%	4.6	70%
2	Stat based on Matomo Analytics	no	yes, complete	no	no		1.0%	1.0	80%
<b>Licensing Framework</b>						<b>19.2</b>	<b>82%</b>		
1	Item view - show item license	yes, partially	yes, complete	yes, complete	no		1.7%	1.6	80%
2	Technical support for restricted items	no	yes, complete	no	no		4.5%	4.2	80%
3	Choose license in submission process	yes, partially	yes, complete	yes, partially	no		4.8%	4.6	80%
4	License administrator	no	yes, complete	no	no		6.6%	6.2	85%
5	Attach/detach license to/from item	no	yes, complete	no	no		2.1%	2.0	80%
6	New search filter option in search page - Licenses	no	yes, complete	no	no		0.7%	0.7	80%
<b>Look&amp;Feel</b>						<b>8.1</b>	<b>77%</b>		
1	Support Ref Table citations - CMDI bib	no	yes, complete	no	no		4.1%	3.9	80%
2	After DB update, OAI gets updated	no	yes, complete	no	no		1.0%	1.0	85%
3	Item view shows versioning	yes, partially	yes, complete	yes, complete	yes, completely				95%
4	Possible to create new version of item	yes, partially	yes, complete	yes, complete	yes, completely		3.5%	3.3	76%
<b>Data downloads</b>							<b>80%</b>		
1	Resumable downloads, restrictive	no	yes, complete	yes, complete	yes, completely				80%
<b>AAI using Shibboleth</b>						<b>7.5</b>	<b>75%</b>		
1	Federated SSO, authorization via Shibboleth	yes, partially	yes, complete	yes, partially	yes, partially		2.1%	2.0	75%
2	Page with a list of released attributes (from IdP)	no	yes, complete	no	no		2.3%	2.6	78%
3	Item deposit by registered users only	yes, completely	yes, complete	yes, complete	yes, completely				90%
4	Is it possible to hide item?	yes, completely	yes, complete	yes, complete	yes, completely				90%
5	Implement GEANT... - DP-CoC	no	yes, complete	no	no		0.3%	0.3	95%
6	Login - select federated login	no	yes, complete	no	no		2.3%	2.6	70%
<b>Unicode Support</b>						<b>1.1</b>	<b>95%</b>		
1	DSpace should support UTF-8	yes, partially	yes, complete	yes, complete	yes, completely				80%
2	Multilingual Support	yes, completely	yes, complete	yes, complete	yes, completely		1.2%	1.1	95%
<b>Total</b>						<b>94.1</b>	<b>78.07%</b>		

- CLARIN DSpace IG (interest group)
  - slack #dspace channel
- DSpace IG meeting in OXFORD (hosted by CLARIN-UK)
  - at the end of this year
  - announced at #dspace

# Thank you

---