

CLARIN-DSpace upgrade



Timeline

- **11/2016** - WG DSpace UI started
- **07/2021** - DSpace v7.0 release
- **11/2021** - **clarin-dspace upgrade feasibility analysis**
 - LINDAT/CLARIAH-CZ
- **02/2022** - DSpace v7.2 release
- **03/2022** - clarin-dspace upgrade started
 - CLARIN-UK and LINDAT/CLARIAH-CZ
- **05/2022** - **official announcing EOL**
- **01/2023** - DSpace v5 support ends
- **02/2023** - DSpace v7.5
- **07/2023** - DSpace v6 support ends
 - v7.6 last of v7 with new features
 - aimed for feature parity with v6

Current state

- **clarin-dspace backend**

- changed files: 414, LOC: 54154
- few LOC reused, internal api changed, **tests**
- <https://github.com/ufal/clarin-dspace/pull/1025>

- **clarin-dspace frontend**

- changed files: 447, LOC: 25961
- completely new code, generated components
- <https://github.com/ufal/dspace-angular/pull/1>

- **clarin-dspace python API**

- pyclarinspace was not used, new project by dspace
- changed files: 92, LOC: 8985
- <https://github.com/dataquest-dev/dspace-rest-python/pull/1>

Current state

- **05/2023 - clarin-dspace v7.2** in testing
 - a few known bugs, fixed but not merged yet
 - a few non mandatory features missing
 - administration tools, curation tasks UI, specific crosswalks



Features

- licenses - authz, UI administration, selector
- handle table - metadata, external
- detailed download statistics
- different crosswalks
- complex input fields
- OpenAIRE support
- tombstone - UI, metadata
- look&feel - home page, search, refbox, header/footer,
- AAI/Shibboleth - authz, UI, logging/parsing
- authz - public, private, hidden, withdrawn, deleted - search/item view/oai-pmh
- item relations
- new tables done properly (services, entities, controllers, converters)
- content negotiation, CMDI
- share submissions



New DSpace Features

- complete (and stable) REST API
- DOI support
- configurable entities
 - typed items with relationships
 - authors, organisations, funding
- S3 support
- docker support (development friendly)



What happened

- overestimated v7.2 production quality
- **underestimated the complexity and amount of work put into clarin-dspace v5 UI**
 - working vs properly implemented
- the learning curve of angular "code style" used in DSpace
- vanilla migration vs custom "data pump"
 - high level of control required
- changed coding practices
 - added many unittests and blackbox testing

Takeaways

- in depth required knowledge for CLARIN B centre is substantial and not common
- sustainable upgrade of a CLARIN B repository will require a team with expertise
 - currently used UI frameworks are nice but complex
 - finding and keeping UI developers (in academia) is difficult
 - unless very specific conditions are met, it may not be commercially interesting
- active community
- security



Thank you



<http://lindat.cz>

