

## Motivations for moving to DSpace7

**(UoA)** The integration of Open Access and infrastructure with the Open Access mandate. Noted the decoupling features, patching support and improved security. Also given the Symplectic Elements integration and our regular upgrades in this space, DSpace required to keep pace.

**(Kim Shephard on why an institute might care to migrate to v7)** DSpace7 is recommended to keep DSpace running, the IT team happy regarding patches and security. V7 is the version with the most support by the DSpace community so the continuation of security improvements, patches and support might be enough of a motivator to move to V7.

## Architecture details and challenges in DSpace7 migration

**(AUT)** In the space of DSpace architecture and open access management, AUT service on-premise running frontend and backend applications and 1 database across three servers and have the ability to add an additional server. They are running a docker with the frontend and database servers but not in the backend. Regarding whether AUT are doing database replication or standalone, they are currently running stably on standalone and are not sure of the benefits of replication.

Regarding deciding on the size, how many items vs server specifications, they did not perform any specific calculation. They considered the capabilities needed for their then current version, v5, and referenced the hardware requirements in the upgrade documentation. From there they are monitoring usage patterns. Generally, a good experience from the recommended data packs from the documentation.

AUT had an issue with data loss in authorship fields, with items moved from dc.contributor.author to dark.contributor.author – a potential risk related to the schema in Elements or crosswalks/RT2 deposit crosswalk.

**(Lincoln)** Lincoln have an Atmire hosted open repository DSpace7 as well as some additional features through Atmire hosting. Initially the migration was a bit rocky but had regular support from their product manager. Their DSpace7 environment was live as of last week (i.e. week of September 28<sup>th</sup> 2023). Migrating required a lot of data tidying to conform to the prefix structure in DSpace7.

Mentioned they have an inspirational list of data fields for tidying ensuring consistent data and terminology.

Also mentioned they can share some of their configuration details but not their files due to being hosted with Atmire. With the move to being hosted, configuration was basically rebuilt from scratch and Lincoln gave a specific list of essential configurations including information for admin only visibility to avoid the erroneous display of student personal information.

Hosting was also mentioned in the discussion around filtering of good and bad crawlers, with that responsibility lying with Atmire. Supported crawlers specifically mentioned was Elements, digital.nz and the library search.

For disaster recovery discussion, they noted that at the point of an issue occurring, Atmire would automatically restart the server whilst the user was placed into a read-only standby for 15 minutes. Additionally, they have backing up practices and a TEST environment populated from production data.

**(UOA)** Holdover xml files from RT1 requires attending to clearance of 1.2M files, with a curation project that was tasked with data cleaning before the move from disc storage to cloud hosting.

## CC licensing fields

dc.rights.uri and dc.rights.name had been in use. However, it had been populated with other sata and links too, thus DSpace was confused. Led to the creation of new fields, dc.rights.cc.uri and dc.rights.cc.name, and adding of CC information into those fields whilst other fields were used for any other rights or restricted licenses etc.

CC license fields were managed with the DSpace config. It was noted that with Elements crosswalks, to ensure it is also configured and displaying properly. If those fields only contain CC information, then it should work however the config needs to align.

For UoA, custom configuration with CC licensing broke pre-v6 and will need to be addressed with their move to v7.

## Statistics

**(UoA)** With Elements Discovery profiles direct download link of DSpace items, would there be a change in the statistics reported in DSpace?

**(Lincoln)** Elements is managed by another team in the university but monitor Google analytics. Whilst there has been a decrease in total usage, it is still too early to definitively say given the update was so recent. They are intending to continue monitoring traffic coming through Elements Discovery for the PDF rather than via DSpace.

**(Kim Shepard)** Institutions have different practices in regard to retaining statistics in the upgrade, some opting to not bring stats, others to restrict the time period migrated (i.e., the last year). Depending on version of DSpace, schema settings in stats may need changes in formatting to be able to be brought across.

## Multiple front ends

DSpace7 has the ability to have different angular frontends run over the same backend and there was the opportunity to expose different collections via this feature. However, no one had yet explored this.

Running multiple frontends and docker swarms were noted to be a major improvement on processing load.

## Solr search

Handling of titles with colons was noted to be poor with Solr, potentially due to the use of colons in its control vocabulary. However, Solr should be able to handle searching despite the use of colons, pluses and quotations for controlled vocabulary and this may be a bug. Advanced search settings may be something to consider setting up.

In regard to using quotations around full titles which utilise colons (i.e., "Full title: continuation of title"), concerns were raised that these are public repositories and users may not know to utilise phrase searching operators.

## Browsing by supervisor functionality

**(Craig)** It was noted that searching by supervisor was not an out-of-the-box functionality. Whilst browsing by author is and supervisors are listed often there, there was no ability to filter by author as student and author as supervisor. By supervisor browsing was an original feature using the dc.contributor.adviser field, however, that piece of configuration did not migrate.

**(Lincoln)** They have retained the feature using the `lu.thesis.contributor.supervisor` field described in the extended metadata. However, they had specifically asked this of Atmire who handled this configuration.