# Investigating Evolving Collection Support with Simple Tools


IT | SCHOOL OF IT

*Hussein Suleman*
*Department of Computer Science*
*School of IT*
*University of Cape Town*

*hussein@cs.uct.ac.za*

*Last edited: 11/26/2022*

*Slide count:*

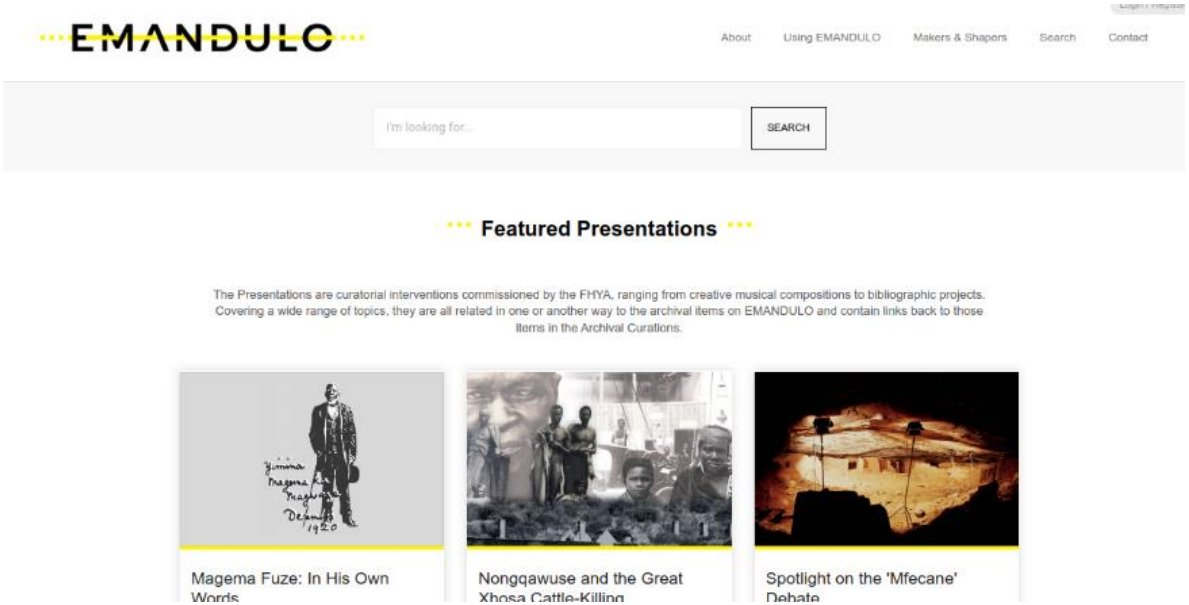
# Simple DL

- An experimental simple Digital Library toolkit.
- Offline static files and views, for low-resource environments and simple preservation.
- File stores and minimal software.
- 3 processes:
  - Import data from spreadsheets.
  - Generate views and website.
  - Index metadata for search
  - and browse.


*https://github.com/slumou/simpledl*

# FHYA and Evolving Collections

- Five Hundred Year Archive is an evolving archive of historical data - collecting, reconceptualising and reframing pre-colonial history on an ongoing basis.



# Types of Evolving Collections

- Single item update
  - Metadata for one item is updated with new information.
- Multiple item update
  - Licence fields of all items in a sub-collection are updated.
- Reorganization of content
  - Some items are removed from different sub-collections and collected into a new sub-collection.

# Supporting Evolving Collections 1/2

- Spreadsheets as norm
  - All metadata is in spreadsheets.
  - Users know how to use spreadsheets and tools are mature for making changes.
- Incremental and clean builds
  - All files can be processed, or only changes can be processed.
  - This support fast minor changes, as well as major structural changes, such as collection refactoring.

# Supporting Evolving Collections 2/2

- Local and Global Identifier Resolution
  - A simple static identifier resolver is updated with all local changes.
  - A global identifier does partial resolution and redirection.
  - External links will be valid across all types of collection evolution.

# Performance Tests

- Evolving static digital libraries could require a lot of work, since all processing happens in advance.
- Full update (8k records, 111 spreadsheets) on a basic Xeon server:
  - 6s to import metadata, 15s to generate views
- Update in one subcollection
  - 3s to import metadata, 1s to generate views
- No changes
  - 3s to import metadata, 1s to generate views

# Concluding thoughts

- Digital libraries are often seen as storing final and immutable objects but not in all contexts.
- Some types of collections are constantly evolving, both structurally and content-wise.
  - Simultaneous evolution of content on one layer (basic objects) with evolution of content on higher layers (exhibitions and presentations).
- Modern tools need to allow evolution while maintaining link validity, preservation, migration, etc.

# Questions / Comments / Discussion

hussein@cs.uct.ac.za


IT | SCHOOL OF IT