

SimpleDL

Building digital libraries differently



Hussein Suleman <hussein@cs.uct.ac.za>

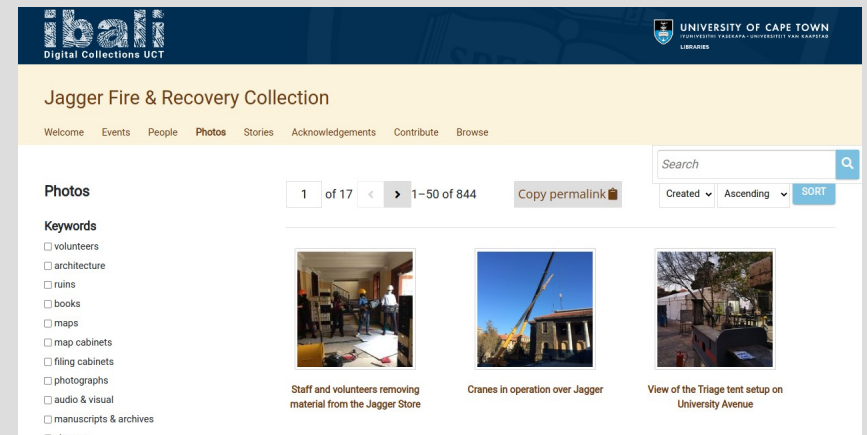
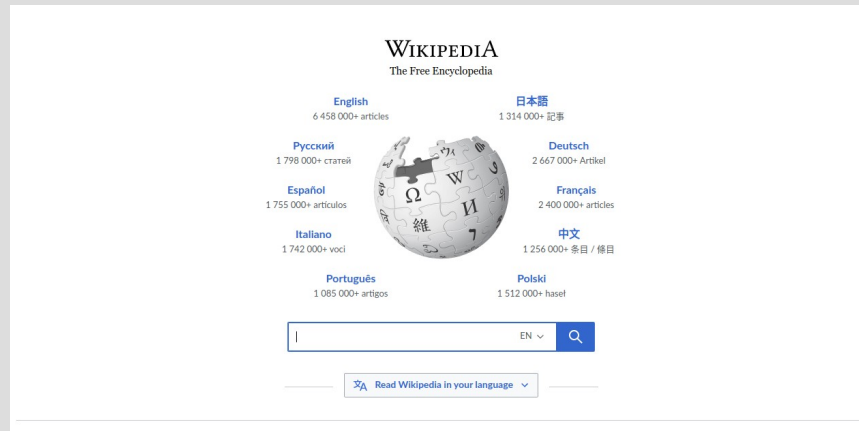
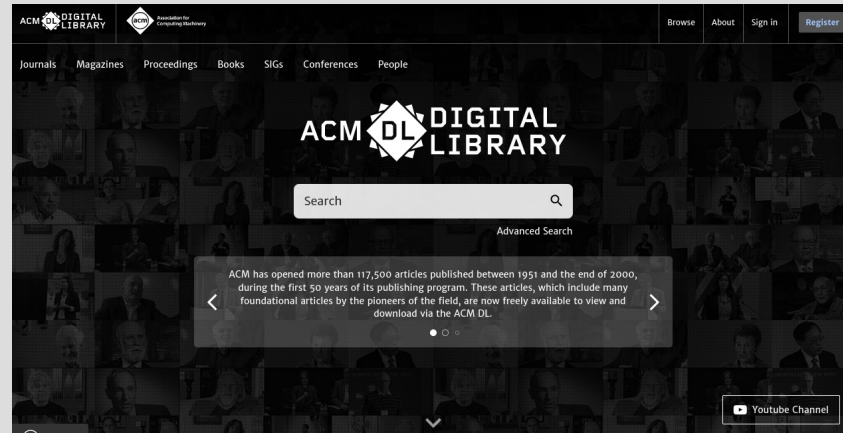
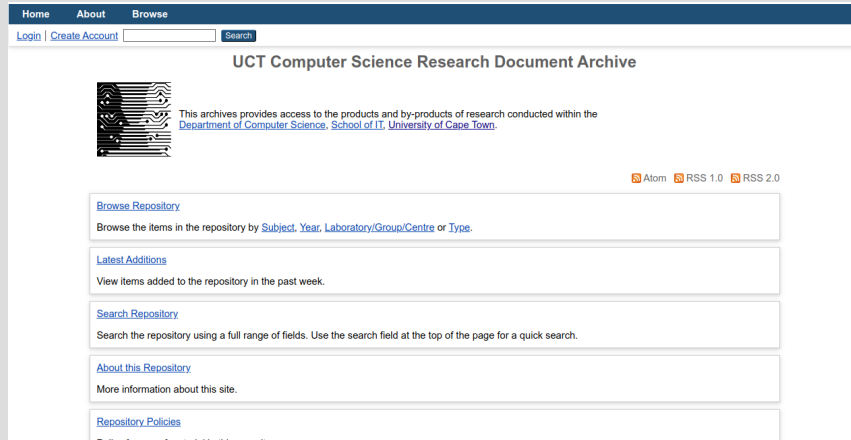
*Digital Libraries Laboratory
Department of Computer Science
School of IT
University of Cape Town
November 2021*



Last edited: 4 August 2022

Slide count: 48

What is a Digital Library?



Archives and Repositories

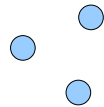
- ❑ Repositories are places to put stuff.
- ❑ Archives are long-term preservation mechanisms.
- ❑ Digital Libraries can be the study of digital archiving.
 - Blurred distinction...

- ❑ Archives are a critical component of all modern research:
 - Primary datasets
 - Publications – Open Access
 - Theses and dissertations
 - Source code repositories
 - Etc.



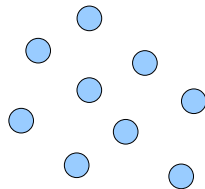
A Short (and Oversimplified) History of Digital Information

In the beginning ...



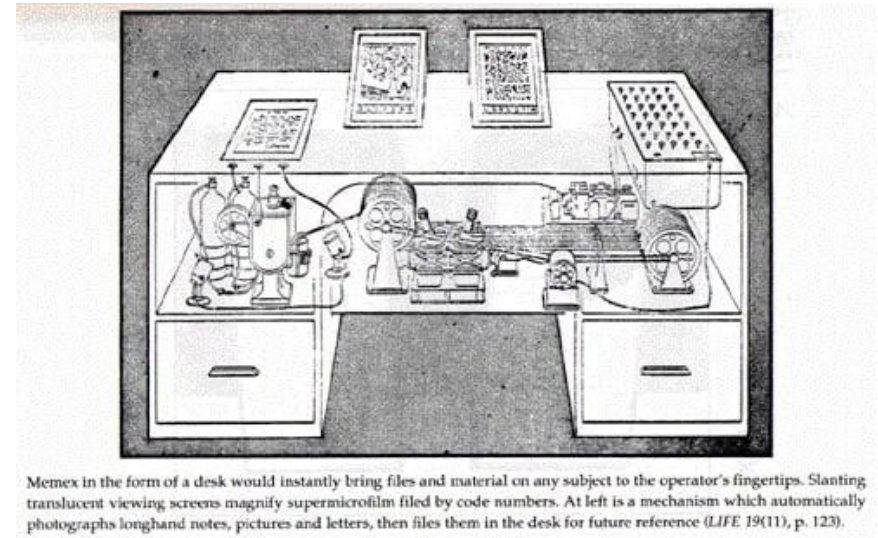
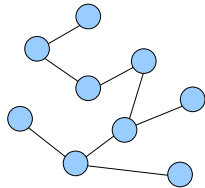
A Short (and Oversimplified) History of Digital Information

With more digital information ...



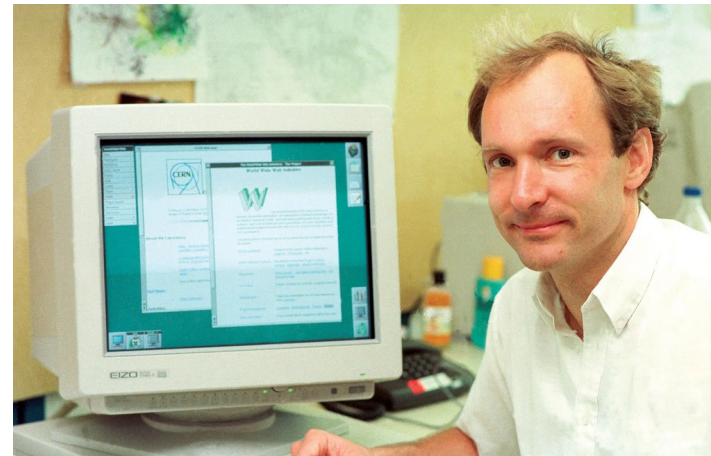
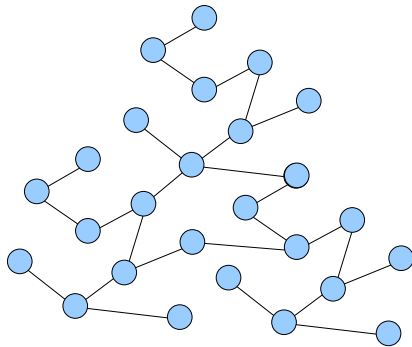
A Short (and Oversimplified) History of Digital Information

Vannevar Bush (1945)
Information Retrieval (1960s)
Hypertext (1965)



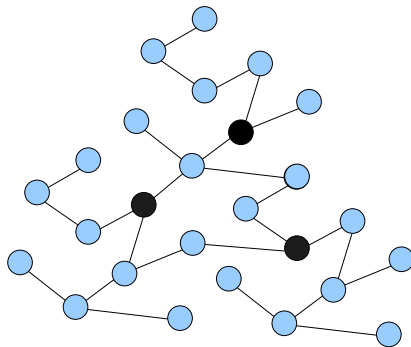
A Short (and Oversimplified) History of Digital Information

World Wide Web (early 1990s)



A Short (and Oversimplified) History of Digital Information

Content instability ...

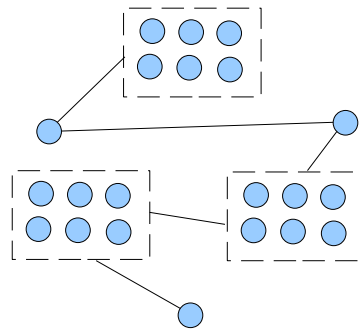


A Short (and Oversimplified) History of Digital Information

NSF Digital Library Initiative (mid 1990s)

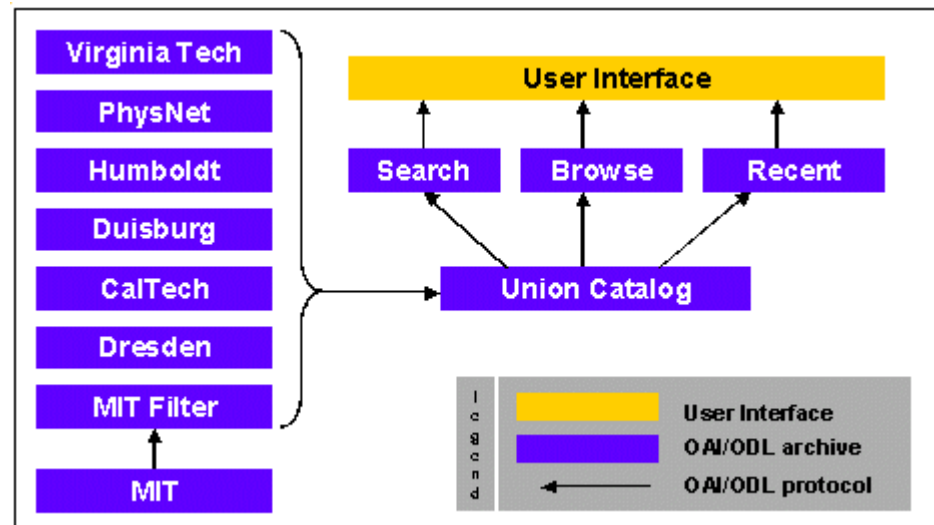
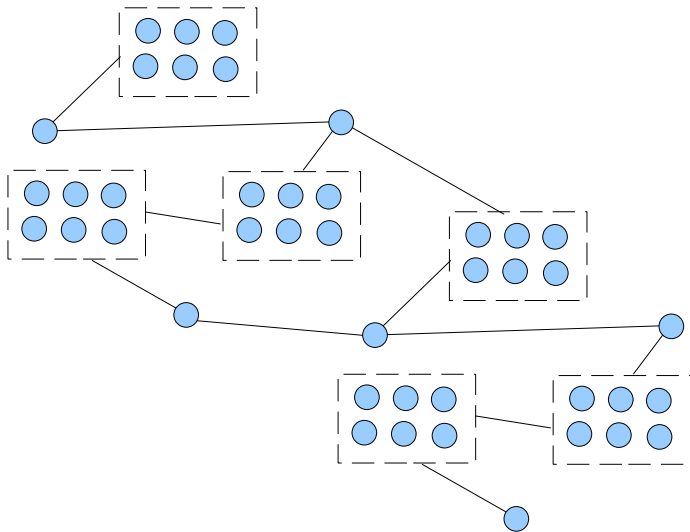
Open Archives Initiative (2001)

Digital Repository Toolkits (2002), CMSes/Wikis (~2000)



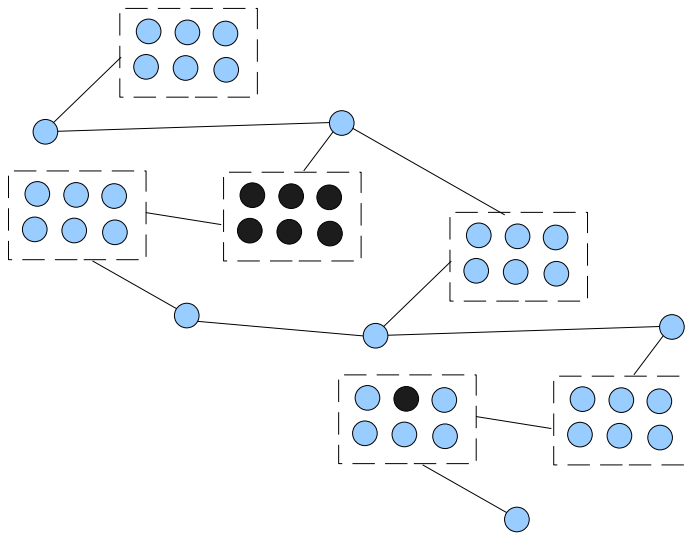
A Short (and Oversimplified) History of Digital Information

Thousands of archives (2010s)



A Short (and Oversimplified) History of Digital Information

When archives disappear ... (2020)



South African Digitization Initiative (SADI)

- ❑ Going back about 10 years ...
 - Collaboration of government agencies (NRF, DST, Nat Lib, Nat Archives, ...), museums (Iziko, ...) and universities (UCT, UP, Wits, ...).
 - Build supporting infrastructure for research, especially in humanities.
 - Archives everywhere + central portal.
- ❑ Europeana sent people to help.
- ❑ Some interested external funders.
- ❑ But no progress ...

Resources to build archives



Money



Time



Skilled People



Institutions

Low-Resource Environments

- Poor countries
 - Most of Africa.
- Specific areas
 - Rural areas all over the world.
- Specific organisations
 - NGO/NPOs relying on donor funding all over the world.
- Specific projects
 - Archivists who get funding for digitization but not maintenance.



Impact of low-resource environment

- ❑ Archives that are shut down (thesis archives in Africa).
- ❑ Archives at risk of being shut down.
- ❑ Archive divide (rich vs. poor).
- ❑ External providers (@mire).
- ❑ No archive.

How to overcome low-resource issues?

- Can we create a more suitable software architecture for low resource environments?
 - Resilience
 - Preservation
 - Rescue

- ... as an alternative to existing opaque architectures for software, based on the fashion of the day ...

Prior Work on Alternative Architectures

□ Greenstone

- archives distributed on CDROM.

□ Project Gutenberg

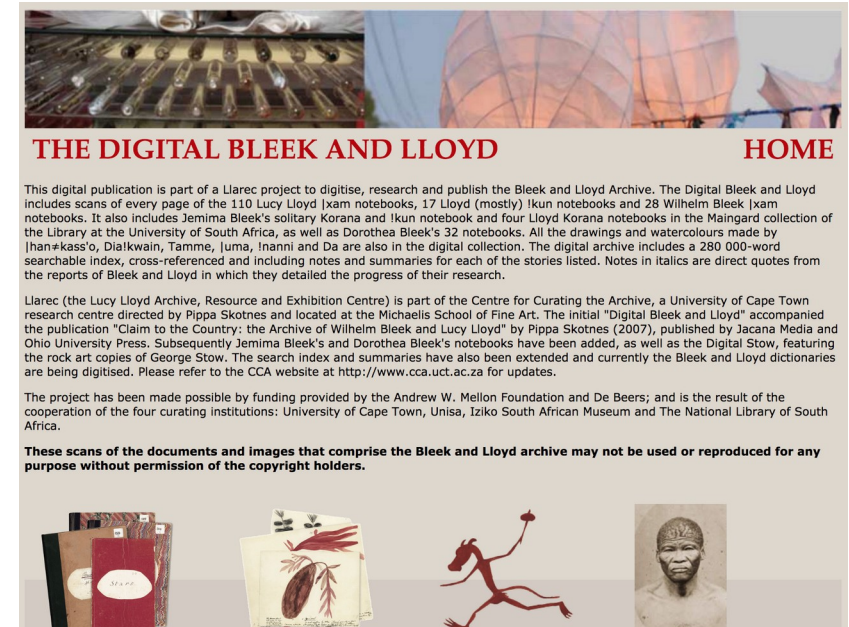
- long term preservation using simple formats.

□ LOCKSS

- enables rescue by making many copies.

□ Bleek and Lloyd Collection

- entirely offline archive on DVD.



Principles and Design Goals 1/2

- ❑ **Simplicity/Minimalism** – only as much system as needed
- ❑ **No imposition on users** – no logins unless necessary
- ❑ **No Internet** - should work even when offline
- ❑ **Simple preservation** – as simple as copying files
- ❑ **Flexibility in objects/services** – no arbitrary restrictions
- ❑ **Superimposed and hierarchical information** – because we understand this
- ❑ **Platform agnosticism** – any OS, any hardware
- ❑ **Pre-processing** – fewer cycles needed when live



Principles and Design Goals 2/2

- It may be naive to assume perfect digital preservation.
- Remember that preservation is
 - Informed by evolving research,
 - Actually unproven and unprovable,
 - And implemented by practitioners.

- **Design for Rescue**
 - What design decisions enable cheaper/faster/easier rescue after a disaster?
 - This was the philosophy behind LOCKSS.

Technical Options: Alternative Data Model

□ XML/XSLT

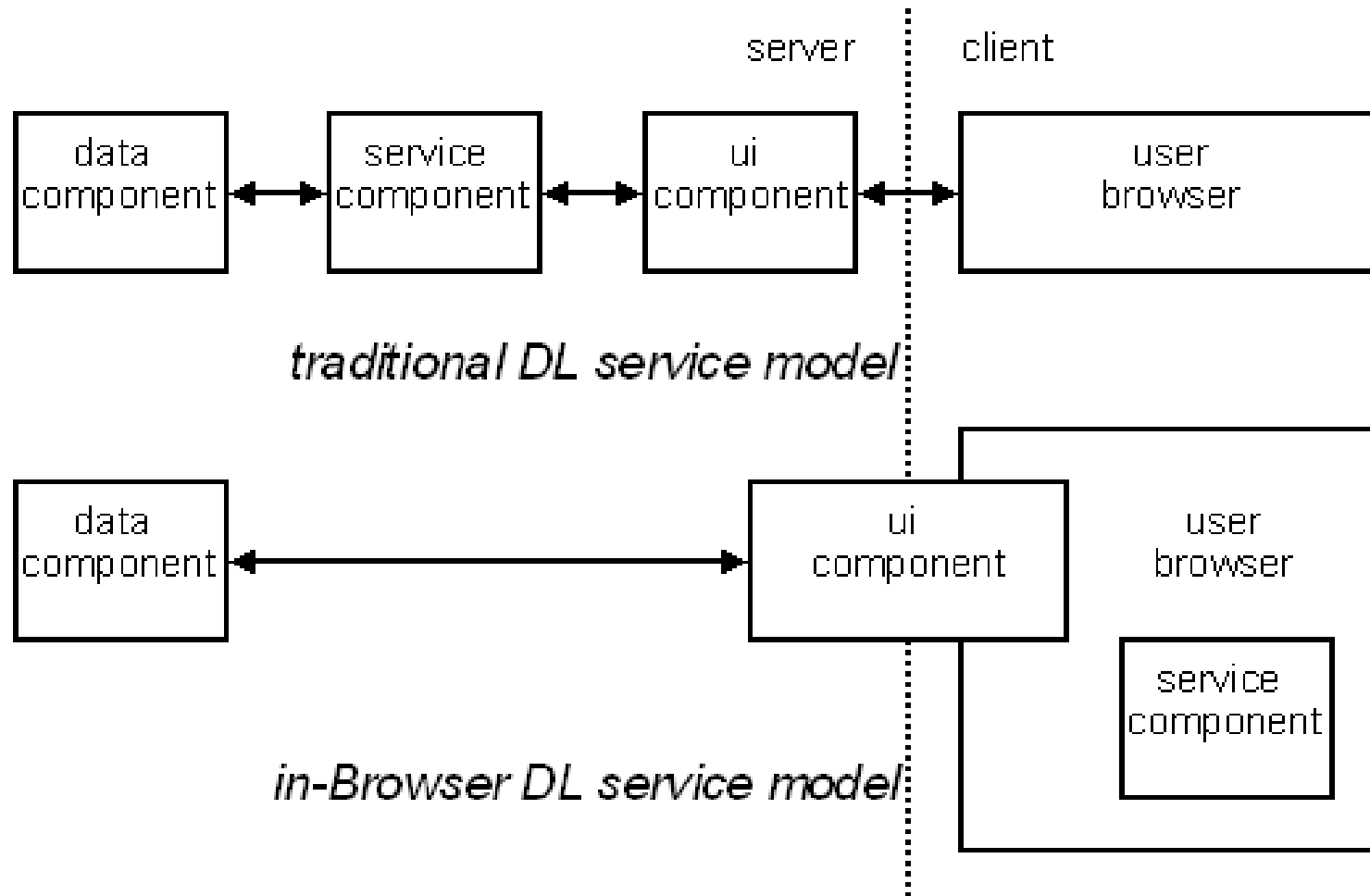
- Store all metadata/data in simple files in hierarchical directories. No database needed, so DBMS is not a risk.

□ Static generation

- Site pre-generation (source->metadata->site) so software is not a long-term risk.



Technical Options: Alternative Service Model



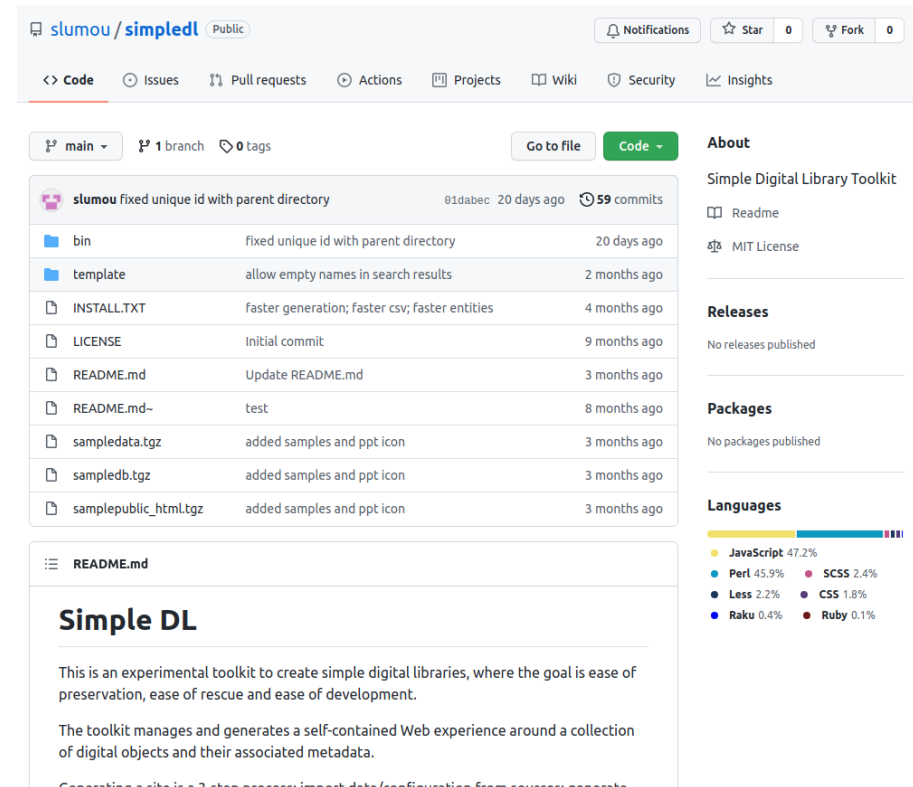
Simple Archives

- Proposition:
 - If we simplify the design of archives ...
- It will cost less to set up and maintain.
- It will be faster to create.
- It can be maintained with fewer skilled people.
- It can be migrated/rescued/preserved easily.



Simple DL

- ❑ Toolkit to create simple archives.
- ❑ Based on many successful ideas in DL architecture, but built for low-resource environments and for adaptation.
- ❑ Experimental but:
 - Can change how we develop systems.
 - Provides a platform to experiment.



slumou / **simpliedl** Public

< Code Issues Pull requests Actions Projects Wiki Security Insights

main 1 branch 0 tags Go to file Code

File	Description	Commit
bin	fixed unique id with parent directory	20 days ago
template	allow empty names in search results	2 months ago
INSTALL.TXT	faster generation; faster csv; faster entities	4 months ago
LICENSE	Initial commit	9 months ago
README.md	Update README.md	3 months ago
README.md~	test	8 months ago
sampledata.tgz	added samples and ppt icon	3 months ago
sampledb.tgz	added samples and ppt icon	3 months ago
samplepublic_html.tgz	added samples and ppt icon	3 months ago

slumou fixed unique id with parent directory 01dabec 20 days ago 59 commits

Simple DL

This is an experimental toolkit to create simple digital libraries, where the goal is ease of preservation, ease of rescue and ease of development.

The toolkit manages and generates a self-contained Web experience around a collection of digital objects and their associated metadata.

Generation a site is a 3-step process: import data/configuration from sources; generate

About
Simple Digital Library Toolkit
Readme
MIT License

Releases
No releases published

Packages
No packages published

Languages

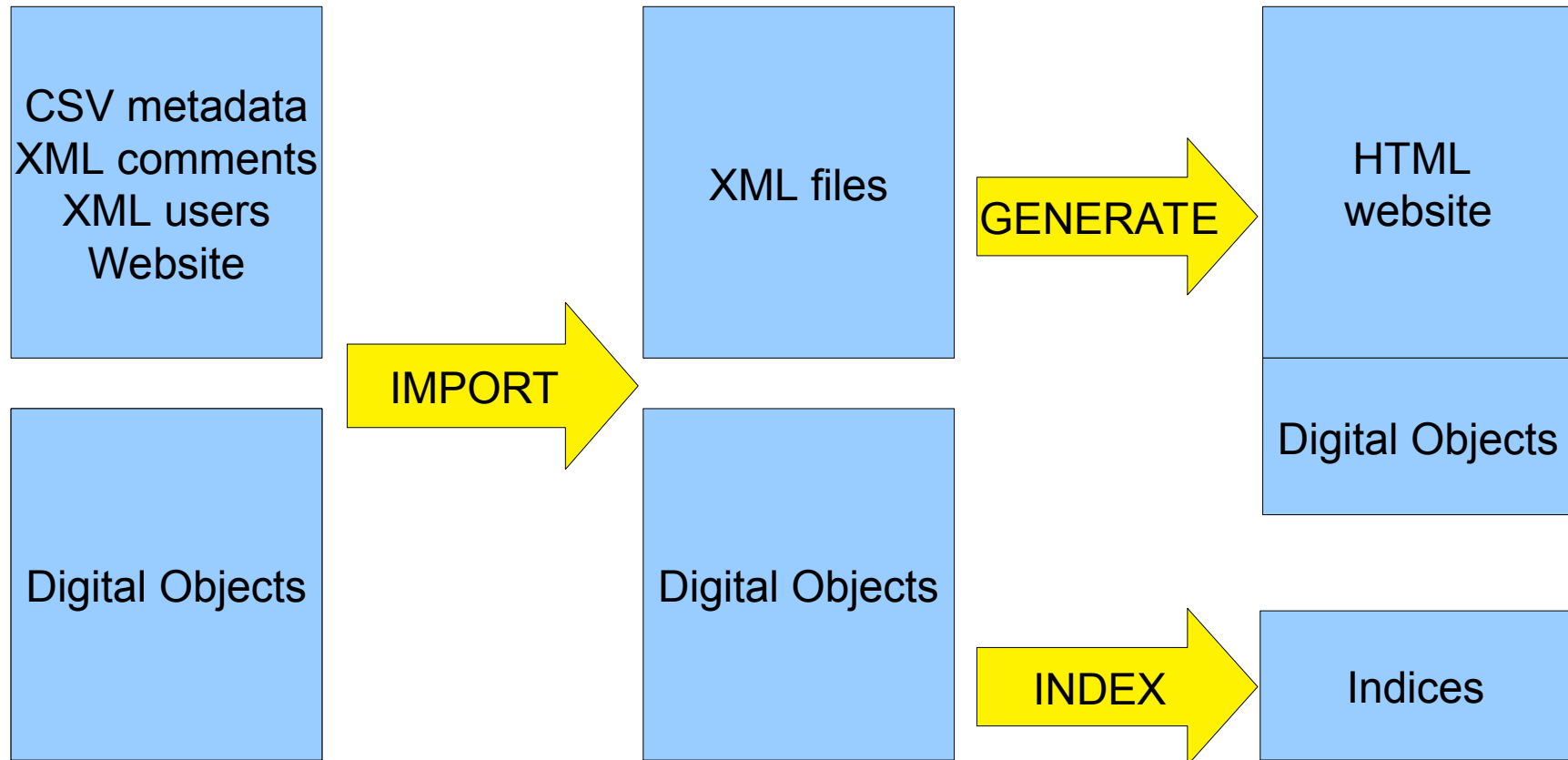
Language	Percentage
JavaScript	47.2%
Perl	45.9%
Less	2.2%
Raku	0.4%
SCSS	2.4%
CSS	1.8%
Ruby	0.1%

<https://github.com/slumou/simpliedl>

Major Features

- ❑ Offline pre-generated website.
- ❑ Metadata created in spreadsheets and XML files.
- ❑ Search/Browse in Javascript.
- ❑ Any metadata format (DC, ICA-Atom).
- ❑ Minimal use of Web applications.
- ❑ XSLT + CSS for styling.
- ❑ User and entity management.
- ❑ Submissions, comments, moderation.
- ❑ Globally unique identifiers/URLs.

Basic Idea



Management Interface

- Web UI to manage files, moderate submissions and import/generate/index.

Simple DL

Logged in: | [Edit Profile](#) | [Admin](#) | [Logout](#)

[Site Home](#)

Manager Options:

Moderate

 |

Manage

 |

Import

metadata

 force ☐ clean ☐ |

Generate

metadata

 force ☐ |

Index

Manage Datasets

Datasets: [spreadsheets] [\[collection\]](#) [\[carousel\]](#)

Options: Create [F]older, [D]elete file/folder, [U]pload file or ZIP

Select a file to attach (for uploads):

Choose file

 No file chosen

[.] ROOT/ [f]

[] ... AMAFA/ [f] [d]

[.] ... JAG/ [f] [d]

[] [JAG_institutional_materials_v2.csv](#) [8839 bytes, 09 May 2021 14:31:01] [d]

[] [Brenthurst_v3.csv](#) [365457 bytes, 09 May 2021 14:31:01] [d]

[] [Maritz_v2.csv](#) [526790 bytes, 09 May 2021 14:31:01] [d]

[] [Jaques_v2.csv](#) [25425 bytes, 09 May 2021 14:31:01] [d]

[] [Brodie.csv](#) [7073 bytes, 09 May 2021 14:31:01] [d]

[] [Horstmann_v2.csv](#) [271199 bytes, 09 May 2021 14:31:01] [d]

Users and Entities

- ❑ Users have accounts on the system.
- ❑ Entities are extracted from metadata.
- ❑ Every entity page can links to authority records and items.

Makers and Shapers Profile

Details

Name	Macebo, Dyer D
Type	Commissioned Contributor
Dates of Existence	1870 - YYYY
History	[Source - Chloe Rushovich for FHYA, 2017, using KCAL materials: Dyer D. Macebo was from the Umvoti mission station. His father was of the Tulwane regiment but was a Christian. He was interviewed by James Stuart in 1898. Macebo was 28 years old when he was interviewed by Stuart.]

Activity

Type ↑↓	Date ↑↓	Item ↑↓
Interlocution	1898/11/02	Macebo, Dyer D. Testimony from 'The James Stuart Archive of Recorded Oral Evidence Relating to the History of the Zulu and Neighbouring Peoples', Volume 2 (Mab-Maz)
Interlocution	1898/11/02	Placeholder for relevant pages from John Wright's annotated photocopies of James Stuart's handwritten notes pertaining to Macebo, Dyer D
Interlocution	1898/11/02	Placeholder for Macebo, Dyer D, Hyperlinked Archival Research Tool
Interlocution	1898/11/02	Placeholder for Macebo, Dyer D: Notes of his conversations with James Stuart, as collated by John Wright

Submissions and Comments

- Users can add comments and attachments; or make new submissions.

Contributions

Name: [ettore morelli](#)

Date: 2021-06-24T14:30:44+0200

This research note is best read together with the note written by John Wright on the same subject. The two research notes were produced during the creation of the EMANDULO presentation 'Exploring uMgungundlovu' between the end of 2020 and the beginning of 2021.

The research note by Wright is here:

<http://emandulo.apc.uct.ac.za/metadata/uMgungundlovu/28/1352/index.html>

And the presentation 'Exploring uMgungundlovu' is here:

<http://emandulo.apc.uct.ac.za/cgi-bin/view/Presentations/httpsstudio-emanduloutcazafhya-exploring-umgungundlovu.zip/studio-emandulo.uct.ac.za/https://studio-emandulo.uct.ac.za/fhya-exploring-umgungundlovu/index.html>

They both relate to the edited and interactive map of uMgungundlovu that is here attached in static form.

Attached File:

[Exploring uMgungundlovu Map Static](#)



Search and Browse

- Faceted search implemented in Javascript.

Search Results

Enter your search terms:
bowl

Custodian
All

Level of description
All

Material type
Object

Makers and shapers
All

Sub-collection
All

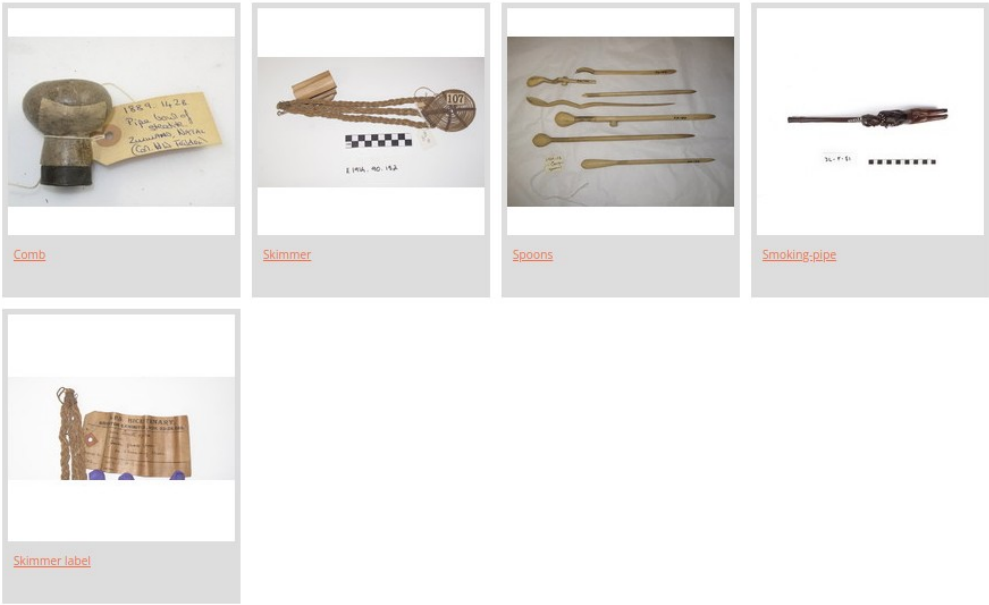
Date
All

Arrange by...
relevance

Search through...
☒ metadata +fulltext +comments
☐ metadata +comments
☐ metadata

Results
Showing 1 to 5 of 5 results

Page: << 1 >> Results per page: 20



The search results display five items in a grid. The first row contains four items: a 'Comb' (a wooden comb with a tag), a 'Skimmer' (a long-handled wooden skimmer), 'Spoons' (a set of wooden spoons), and a 'Smoking-pipe' (a long, thin pipe). The second row contains one item: a 'Skimmer label' (a small label with text). Each item has a thumbnail image and a caption below it.

Evaluation

- ❑ Evaluation of software design is hard!
- ❑ Scalability:
 - Advanced faceted search queries:
 - ❑ Multiple search terms + multiple browse facets
 - ❑ 30000 items searched in 0.3s
 - Up to 100000 items handled in sub-second times
- ❑ Case Studies:
 - Emandulo, Re-source
 - New Bleek and Lloyd
 - NDLTD Document Archive, CS-HEAP

Case Study 1: Emandulo 1/2

...EMANDULO...

[About](#) [Using EMANDULO](#) [Makers & Shapers](#) [Search](#) [Contact](#)

Login / Register

... Featured Presentations ...

The Presentations are curatorial interventions commissioned by the FHYA, ranging from creative musical compositions to bibliographic projects. Covering a wide range of topics, they are all related in one or another way to the archival items on EMANDULO and contain links back to those items in the Archival Curations.



Case Study 1: Emandulo 2/2

...EMANDULO...

Login / Register

About

Using EMANDULO

Makers & Shapers

Search

Contact

Home

- ↳ Early KZN Materials collected by James Stuart and associated items
- ↳ FHYA curation of the testimonies of individual interlocutors from six published volumes of the James Stuart Archive
- ↳ FHYA curation of the testimonies of individual interlocutors from Volume 1 of the James Stuart Archive (Ant-Lyl)



FHYA curation of the testimonies of individual interlocutors from Volume 1 of the James Stuart Archive (Ant-Lyl)

... Scope and Content ...

[Source - Debra Pryor for FHYA, 2019: Series contains front and back matter and the individual testimonies from 'The James Stuart Archive of Recorded Oral Evidence Relating to the History of the Zulu and Neighbouring Peoples', Volume 1 (Ant-Lyl), edited by Colin de B. Webb and John Wright.

The volume was edited in the period April 1971 to November 1975. Wright was responsible for translating into English words and passages recorded by Stuart in isiZulu. He discussed points of translation with Professor A.T. Cope, head of the Department of Bantu Languages at the University of Natal in Durban. The volume was published in March 1976. It has since been reprinted.]

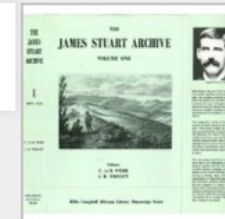
... Metadata ...

Title

FHYA curation of the testimonies of individual interlocutors from

... Contents ...


File: **James Stuart Archive, Volume 1, Front matter**



File: **Antel, Mr.**




Case Study 2: New Bleek and Lloyd 1/3



THE DIGITAL
BLEEK & LLOYD
General editor Pippa Skotnes


ABOUT USABOUT THE DIGITAL BLEEK & LLOYDCOLLECTIONS ▾RESOURCES ▾




centre for
CURATING
the archive

Search


COLLECTIONS




NOTEBOOKS




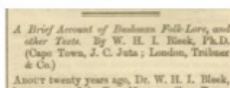
DRAWINGS, WATERCOLOURS,
OBJECTS




LEXICONS









CONTRIBUTORS




Wilhelm Bleek




Lucy Lloyd





!alkunta




!|kabbo









Case Study 2: New Bleek and Lloyd 2/3

Search Results

Enter your search terms:

Collection
All

Contributor
Inanni

Category
artefact and dress

Keyword
All

Sub-collection
All


Date
All

Arrange by...
relevance


Results

Showing 1 to 20 of 29 results


Page: << 1 >> Results per page: 20




[NLLB D ln116](#)



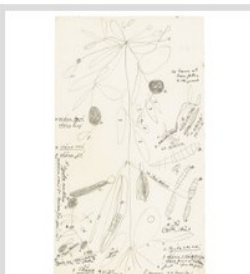
[NLLB D ln113](#)




[NLLB D ln069](#)




[NLLB D ln014](#)




[NLLB D ln064](#)



[NLLB D ln062](#)




[NLLB D ln103](#)




[NLLB D ln136](#)

Case Study 2: New Bleek and Lloyd 3/3



THE DIGITAL
BLEEK & LLOYD
General editor Pippa Skotnes

ABOUT USABOUT THE DIGITAL BLEEK & LLOYDCOLLECTIONS ▾RESOURCES ▾



centre for
CURATING
the archive


Search

Home > Dictionaries > English to |xam > F > Face

Face

A B C D E **F** G H I
J K L M N O P Q R
S T1 T2 U V W X Y
Z

Face
Face
Face
Face (faces)
Face (the)
Face (the)
Face (the)



Metadata

Title	Face
Contributor	LC Lloyd
Type	envelope
Medium	ink on manila envelope
First letter / click	F
Keyword	Face

Case Study 3: Historical Equipment Archive Project 1/2

UCT Computer Science

Historical Equipment Archive Project

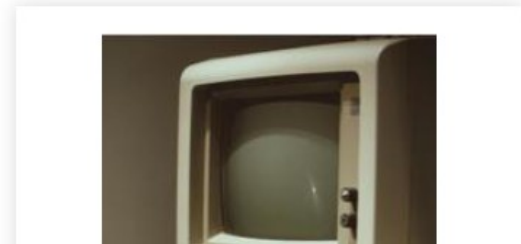
[Login / Register](#)

[HOME](#) | [ABOUT HEAP](#) | [PEOPLE](#) | [SEARCH](#) | [CONTACT US](#)

Search HEAP

Enter your search terms:

Categories of hardware



Case Study 3: Historical Equipment Archive Project 2/2



Workstations

Metadata

Title	Workstations
Description	Workstations are high performance computers that ...
Source	Department of Computer Science

Contents

Item: **NeXT Personal Computer**



Item: **Olivetti PC - M24**



The Bootstrapping Philosophy

- This is a bootstrapping project:
 - Quickly create new archives for small collections in low-resource environments.
 - Meet the needs of most of the people most of the time.
 - Support experimentation and customisation.
 - Support eventual transition to other systems.

- This is not a solution for all problems.
 - One size does not fit all.
 - One design/architecture does not meet all needs.

Some Reflections 1/2

- ❑ How people engage with data is a rapidly evolving and unsettled idea.
- ❑ Context matters and bias is real.
- ❑ An archive is not the final location for an object.
- ❑ An archive is one temporary location for an object.
- ❑ We are not all online or online all the time.

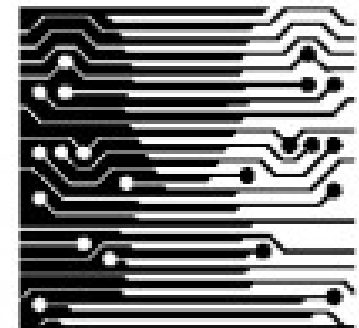
Some Reflections 2/2

- This is about compromises:
 - Archiving/Preservation can be simpler.
 - Scalability is not as good (if you have millions of objects).
- This might help us get to a national heritage portal faster.
- Thinking about resource limits suggests radically different system design approaches.
 - This will ultimately benefit everyone in the world!
 - And it goes well beyond digital archives.

Digital Libraries Laboratory @ UCT

- ❑ 15+ years old
- ❑ We study the organization of information for research and education.
- ❑ Focus areas:
 - Cultural heritage preservation
 - African language information retrieval
 - Educational technology
- ❑ Currently
 - 2 Postdocs, 3-6 PhD students, 9 Masters students, +Honours +UG +Assistants

**digital libraries
laboratory**



Digital Libraries Laboratory M/D/PD Members/Alumni

- **Current:** Aslam Safla, Abayomi Agbeyangi, Bernard Akhigbe, Gavin Asary Ryan le Roux, Tinashe Makwande, Tumelo Makgaka, Mustafa Ali, Julian Nkuna, Jarryd Dunn, Toky Raboanary, Frances Gillis-Webber, Jecton Anyango, Zola Mahlaza, Jackson Moji, Jivashi Nagar, Sunkanmi Olaleye
- **Alumni:** Khanyisa Mtombeni, Joseph Telemala, Soham Singh, Kinsley Ndenge, Janine Ritchie, Devan Govender, Selvas Mwanza, Maryam Al-Mukhaylid, Jessica Mantel, Pascal Ngaboyisonga, Catherine Chavula, Mvurya Mgala, Morebodi Modise, Ayodeji Olojede, Anesh Kalan, Sinini Ncube, Josiah Chavula, Job King'ori Maina, Mohato Lekena, Sarah Brittan, Sean Packham, Haji Ali Haji, Deena Chetty, Charity Chao Mbogo, Grace Ssekakubo, Jorgina Paihama, Mushashu Mwansa Lumpa, Ngoni Munyaradzi, Lebeko Poulo, Marius Nel, Kathryn van Niekerk, Lighton Phiri, Michelle Havenga, Kyle Williams, Lisemelo Tlale, Morwan Mohammed Noor, Stefano Rivera, Mohammed Mustafa Ali, Hisham Abdalla, Nils Körber, Mongezi Ngindana, Kieran Sharpey-Schafer, Christopher Parker, Ndapandula Nakashole, Themba Munalula, Mayumbo Nyirenda, Ashwinkoomarsing Balluck, Mpho Sello, Masha du Toit, Candice Cloete, Muammar Omar, Siyabonga Mhlongo, Linda Eyambe



Ongoing Work

- ❑ Complex object authoring tools – to create visualisations of concepts maps
 - *External company, using Wordpress*
- ❑ Complex object support – archiving entire presentations/exhibitions
 - *using a Web downloader*
- ❑ Archive archiving – to manage versions of evolving knowledge structures
 - *ARCH2 2022 Honours project*
- ❑ Virtualisation – to create/manage simple archives remotely
 - *Masters project – Mustafa Ali*

More Ongoing Work

- Bookmarking facilities – personal and community-oriented
 - *2021 Honours project*
- Measuring feasibility of archives on a personal/mobile device – takeaway archives
 - *Masters project – Tumelo Makgaka*
- Automatic Entity linking in historical documents
 - *Masters project – Jarryd Dunn*
- Virtual-Physical archive linkages (CS-HEAP)
 - *Postdoc – Bernard Akhigbe*
- Abstracting the architecture beyond digital libraries
 - *Postdoc – Abayomi Agbeyangi*



Open Ideas for 2023 and beyond

- ❑ A national infrastructure for heritage archives.
 - Potential partners...
 - ❑ Robben Island Museum, Mayibuye Centre, various UCT projects, etc.
 - ❑ KZN Museum – PMB and Durban
 - ❑ Nelson Mandela Foundation, Desmond&Leah Tutu Foundation, Constitution Hill, etc.
- ❑ Backing up and repatriating the country's heritage collections.
- ❑ Measuring the rescuability/migration of data.
- ❑ Developing simpler and more effective interoperability.
- ❑ Simple collections for academic/education archives (e.g., Vula sites).
- ❑ Developing theoretical models for alternative software development approaches in low-resource environments.



Joining the DL Lab

- ❑ Weekly or fortnightly group meetings, reading groups
 - Critical discussion guaranteed :)
- ❑ Links to related groups around the world
- ❑ Branding (!)
- ❑ Social/Cultural Excursions

- ❑ If you are interested, talk with me.

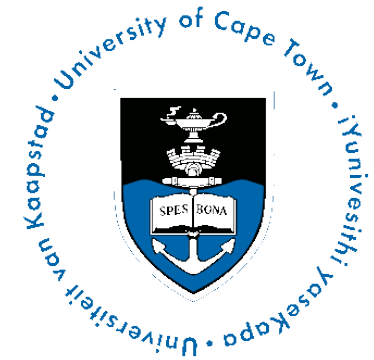


References

- Suleman, H. (2021) Simple DL: A toolkit to create simple digital libraries, to appear in Proceedings of Asian Digital Libraries Conference (ICADL), December 2021.
- Suleman, H. (2019) Investigating the effectiveness of client-side search/browse without a network connection, Proceedings of 21st International Conference on Asia-Pacific Digital Libraries (ICADL), 4-7 November 2019, Kuala Lumpur, Malaysia, Springer. Available: <http://pubs.cs.uct.ac.za/id/eprint/1333/>
- Suleman, H. (2019) Reflections on Design Principles for a Digital Repository in a Low Resource Environment, Proceedings of HistoInformatics Workshop 2019, 13 September 2019, Oslo, Norway, CEUR. Available: <http://pubs.cs.uct.ac.za/id/eprint/1331/>
- Phiri, L. and H. Suleman (2013). Flexible Design for Simple Digital Library Tools and Services. In Machanick, Philip and Mosiuoa Tsietsi, Eds. Proceedings SAICSIT '13 2013 South African Institute for Computer Scientists and Information Technologists, pages 160-169, East London, South Africa. Available <http://pubs.cs.uct.ac.za/archive/00000891/>
- Phiri, L., K. Williams, M. Robinson, S. Hammar and H. Suleman (2012) Bonolo: A General Digital Library System for File-based Collections. In Chen, Hsin-Hsi and Gobinda Chowdhury, Eds. Proceedings 14th International Conference on Asia-Pacific Digital Libraries 7634, pages 49-58, Taipei, Taiwan. Available <http://pubs.cs.uct.ac.za/archive/00000784/>
- Suleman, H., M. Bowes, M. Hirst and S. Subrun (2010), Hybrid Online-Offline Collections, in Proceedings of Annual Conference of the South African Institute for Computer Scientists and Information Technologists (SAICSIT 2010), Bela Bela, South Africa, 11-13 October 2010, ACM. Available <http://pubs.cs.uct.ac.za/archive/00000655/>
- Suleman, H. (2007). Digital Libraries Without Databases: The Bleek and Lloyd Collection", in Kovacs, Laszlo, Norbert Fuhr and Carlo Meghini (eds): Proceedings of Research and Advanced Technology for Digital Libraries, 11th European Conference (ECDL 2007), pp. 392-403, 16-19 September, Budapest, Hungary, pp. 392-403, Springer. Available <http://pubs.cs.uct.ac.za/archive/00000433/>
- Suleman, H. (2007). in-Browser Digital Library Services, in Kovacs, Laszlo, Norbert Fuhr and Carlo Meghini (eds): Proceedings of Research and Advanced Technology for Digital Libraries, 11th European Conference (ECDL 2007), 16-19 September, Budapest, Hungary, pp.462-465, Springer. Available <http://pubs.cs.uct.ac.za/archive/00000434/>



Thank you for listening!



<https://github.com/slumou/simplifiedl>

Questions, comments and suggestions are welcome...

Contact me at: hussein@cs.uct.ac.za

*Digital Libraries Laboratory
Department of Computer Science
School of IT
University of Cape Town*