

*(and now for something completely different...)*

## *Low Resource Metadata/Data Repositories*

*Hussein Suleman <[hussein@cs.uct.ac.za](mailto:hussein@cs.uct.ac.za)>  
Digital Libraries Lab, Department of Computer Science  
School of IT, University of Cape Town  
24 September 2020*



# Low Resource Archives

- “2 million euros and 2 years and we can build any digital repository system”
  
- Can we use DSpace/Atom/etc.
- Can we do the same thing as everyone else?
  - or
- Can we do better?
  - Should we do better?
  - Do we have to do better?





# 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



# African Problems 1/3

## Skills and Education



- Typical archivists are not as highly skilled as counterparts elsewhere.
  - Education in digital media is still not the norm.
- Education levels of general population hinders preservation – end-user data curation is very difficult.



# African Problems 2/3

## Funding

- Typically, there is little.
- Many archiving/preservation projects are funded by external agencies, but with restrictions on data accessibility.
- There is a desperate need to do more with less.





# African Problems 3/3

## Internet Bandwidth (Digital Divide)



- Non-existent in some places and poor in most other places.
- Archiving/preservation projects designed for high bandwidth are not suitable.





# What is the net effect? 1/5

## What is a Digital Object Repository?



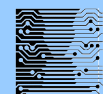
<p><a href="#">Home</a></p> <p><a href="#">Article title search</a>  <a href="#">Author search</a>  <a href="#">Keyword search</a>  <a href="#">Full text search</a></p> <p><a href="#">Advanced search</a></p> <p><a href="#">Acronyms</a>  <a href="#">Thesaurus</a></p> <p><a href="#">Need help?</a></p>	<p><b>Browse the collection</b></p> <p>List of journals:</p> <p><a href="#">Abasebenzi</a></p> <p><a href="#">Afro Newsletter</a></p> <p><a href="#">Africanist News and Views</a></p> <p><a href="#">Amandla-Matla</a></p> <p><a href="#">Amdusa Views</a></p> <p><a href="#">Arise! Yukani!</a></p> <p><a href="#">Black Review</a></p> <p><a href="#">Clarion Call</a></p> <p><a href="#">Congress Resister</a></p> <p><a href="#">Contact</a></p> <p><a href="#">Cosatu News</a></p> <p><a href="#">Crisis News</a></p>
--	---

source: DISA, Univ. of KZN  
<http://disa.ukzn.ac.za>





# What is the net effect? 2/5







# What is the net effect? 3/5

- The Digital Sub-Divide
  - Those with the resources
    - vs.
  - Those without the resources





# What is the net effect? 4/5





# What is the net effect? 5/5





# What we need ...

- Archiving
  - Creation of metadata and digital objects.
  - Management and access to collections.
- Preservation
  - Maintaining digital collections in perpetuity.
  - Active curation of software/hardware/data
- Without the drama of funding ending, staff shortages and access issues ... and collections disappearing or never starting





# How can we stop this trend?

- Clearly, custom solutions are not working.
  
- OSS is not working for us either.
  - How are repository tools designed?
  - What are the principles governing this?
  
- Can we create a more suitable architecture for low resource environments?





# Bleek and Lloyd Collection as exemplar?



## THE DIGITAL BLEEK AND LLOYD

## HOME

This digital publication is part of a Llarec project to digitise, research and publish the Bleek and Lloyd Archive. The Digital Bleek and Lloyd includes scans of every page of the 110 Lucy Lloyd !xam notebooks, 17 Lloyd (mostly) !kun notebooks and 28 Wilhelm Bleek !xam notebooks. It also includes Jemima Bleek's solitary Korana and !kun notebook and four Lloyd Korana notebooks in the Maingard collection of the Library at the University of South Africa, as well as Dorothea Bleek's 32 notebooks. All the drawings and watercolours made by !han#kass'o, Dia!kwain, Tamme, Juma, !nanni and Da are also in the digital collection. The digital archive includes a 280 000-word searchable index, cross-referenced and including notes and summaries for each of the stories listed. Notes in italics are direct quotes from the reports of Bleek and Lloyd in which they detailed the progress of their research.

Llarec (the Lucy Lloyd Archive, Resource and Exhibition Centre) is part of the Centre for Curating the Archive, a University of Cape Town research centre directed by Pippa Skotnes and located at the Michaelis School of Fine Art. The initial "Digital Bleek and Lloyd" accompanied the publication "Claim to the Country: the Archive of Wilhelm Bleek and Lucy Lloyd" by Pippa Skotnes (2007), published by Jacana Media and Ohio University Press. Subsequently Jemima Bleek's and Dorothea Bleek's notebooks have been added, as well as the Digital Stow, featuring the rock art copies of George Stow. The search index and summaries have also been extended and currently the Bleek and Lloyd dictionaries are being digitised. Please refer to the CCA website at <http://www.cca.uct.ac.za> for updates.

The project has been made possible by funding provided by the Andrew W. Mellon Foundation and De Beers; and is the result of the cooperation of the four curating institutions: University of Cape Town, Unisa, Iziko South African Museum and The National Library of South Africa.

**These scans of the documents and images that comprise the Bleek and Lloyd archive may not be used or reproduced for any purpose without permission of the copyright holders.**





# Principles and Design Goals

- ❑ Simplicity/Minimalism
- ❑ No imposition on users
- ❑ No Internet
- ❑ Simple preservation
- ❑ Flexibility in objects/services
- ❑ Superimposed and hierarchical information
- ❑ Platform agnosticism
- ❑ Pre-processing





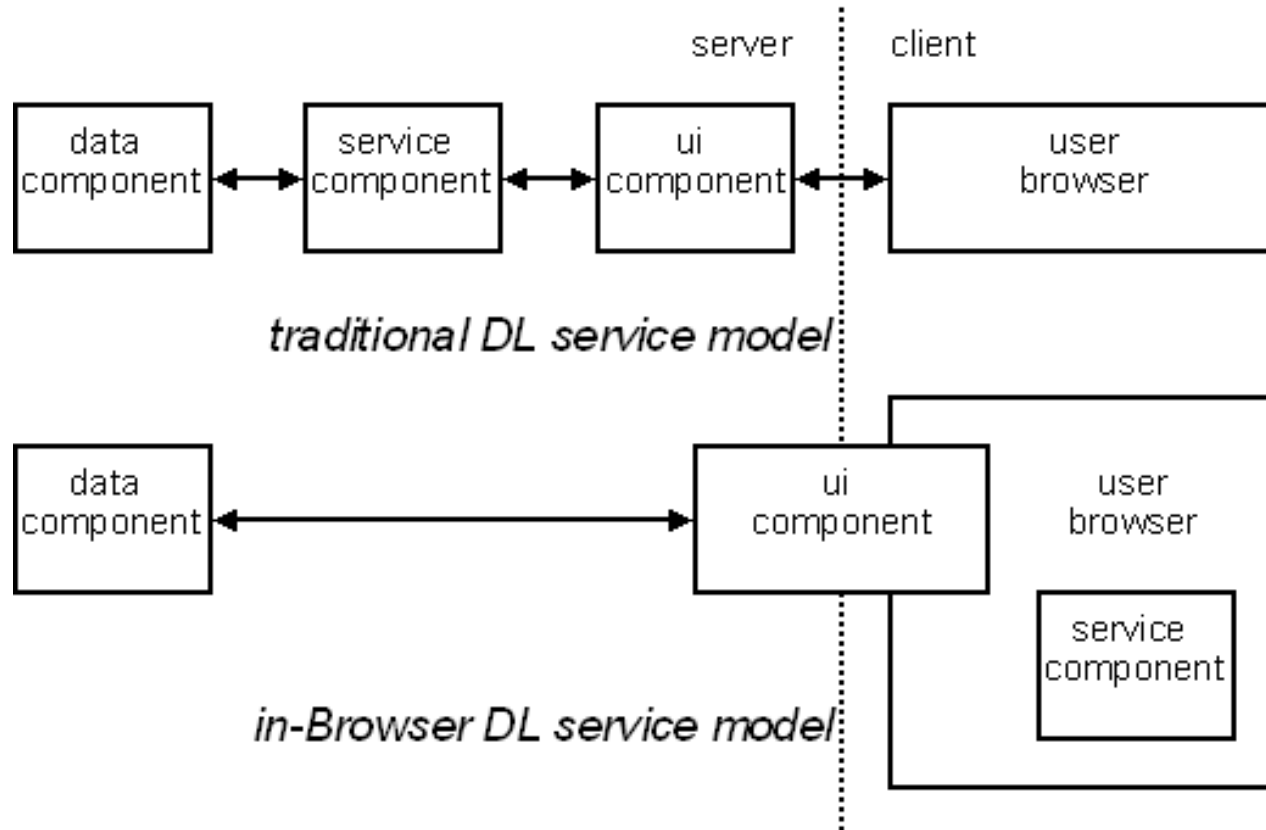
# Implementation Approaches

- XML/XSLT
  - All metadata/data stored in simple files in hierarchical directories – no database.
- Static generation
  - Site pre-generation (source->metadata->site) so software is not a long-term risk.
- Offline search/browse – edge computing
  - Provide services through browser-based Javascript.





# in-Browser Services





# FHYA Prototype 1/3



Logged in as: Hussein Suleman ([edit profile](#))



[About FHYA](#) | [Using the FHYA site](#) | [Contributors](#) | [Search](#) | [Contact Us](#)

## Search FHYA Collections

Enter your search terms:

GO

## Browse FHYA Collections

- [JAG](#)

## About FHYA

In its current form FHYA is an archival exemplar that includes a sample selection of materials pertinent to a small geographic area (KwaZulu-Natal-Swaziland region) in a limited time frame (from about 1770 to various points in the nineteenth century).

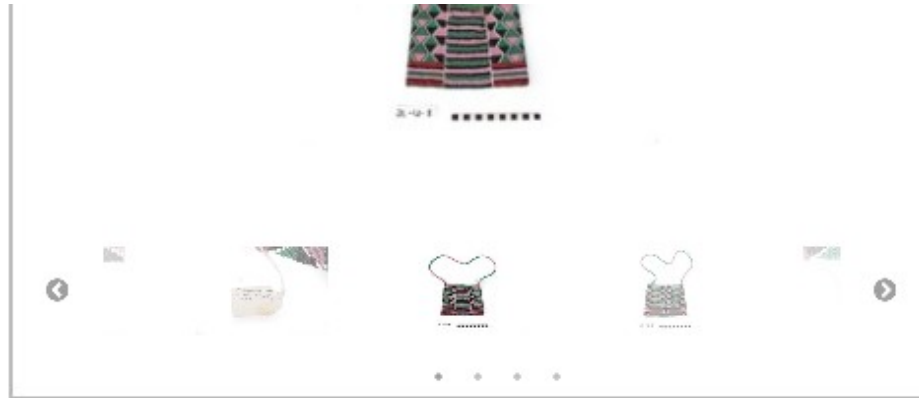
The exemplar includes a highly diverse selection of materials in different media - documents, photographs, recordings - from a few different institutions - museums, libraries, archives - as well as personal collections.

This selection demonstrates that is possible to convene online materials that have been historically separated through institutional practice.

Materials have been processed to make them readily searchable. This creates opportunities for researchers to step beyond institutional categories and make new connections between diverse materials.



# FHYA Prototype 2/3



## Metadata

<b>Title</b>	Beadwork [ Source of title : Nessa Leibhammer using JAG materials ]
<b>Material Designation</b>	Object Textual record
<b>Repository</b>	Johannesburg Art Gallery (JAG)
<b>Identifier</b>	JL-U-8
<b>Arrangement</b>	[Source - Nessa Leibhammer for FHYA, 2015: Accession numbers had already been allocated to the objects in the collection before it was sold to Harry Oppenheimer and the numbering system was retained by JAG: the initials JL stand



# FHYA Prototype 3/3

**500** FIVE  
HUNDRED  
YEAR  
ARCHIVE

About FHYA | Using the FHYA site | [Contributors](#) | [Search](#) | Contact Us

## Search Results

### Query

zulu beadwork

### repository

All

### subcollection

JAG/BRENTHURST

### curationactor

All

### custodyactor

All

### Results

1. [Beadwork](#)  
*JAG/BRENTHURST/JL-U-8.xml*
2. [Beadwork](#)  
*JAG/BRENTHURST/JL-U-166.xml*
3. [Beadwork](#)  
*JAG/BRENTHURST/JL-U-42.xml*
4. [Beadwork](#)  
*JAG/BRENTHURST/JL-U-88.xml*
5. [Beadwork](#)  
*JAG/BRENTHURST/JL-U-14.xml*
6. [Beadwork](#)  
*JAG/BRENTHURST/JL-U-17.xml*
7. [Beadwork](#)  
*JAG/BRENTHURST/JL-U-19.xml*
8. [Staff](#)  
*JAG/BRENTHURST/JL-C-23.xml*
9. [Snuff-spoon](#)  
*JAG/BRENTHURST/JL-H-1.xml*





# Concluding Reflections

- Databases – one size does not fit all!
- We are not all online or online all the time.
- This is about compromises:
  - Archiving/Preservation are simpler.
  - Scalability is not as good (if you have millions of objects).
- Thinking about resource limits suggests radically different system approaches.
  - This will ultimately benefit everyone in the world!



**that's all folks!**

