

Reflections on Design Principles for a Digital Repository in a Low Resource Environment

Hussein Suleman <hussein@cs.uct.ac.za> Digital Libraries Lab, Department of Computer Science, School of IT University of Cape Town 12 September 2019

Digital Libraries

- A Digital Repository/Archive stores and provides access to digital objects/metadata.
 - Institutional repositories, heritage collections, etc.
- Digital Libraries is the meta-discipline that studies the development of digital analogues to non-digital libraries.
 - Usually through the creation and management of digital repositories/archives.



Typical Services of Digital Repositories

- Store documents and metadata
- Search and Browse
- Submission of documents and metadata
- Moderation of content
- Access through Web interface
- Compliance with standards
- Enrichment: comments, reviews, etc.
- Linking with other systems



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?



African Problems 1/3 Skills and Education



- Typical archivists are not as highly skilled as counterparts elsewhere.
- 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 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 everywhere else.
- Preservation projects designed for high bandwidth are not suitable.
- All online solutions must be bandwidthfriendly.



What is the net effect?

What is a Digital Object Repository?



Southern A	frican	Freedom	Stru

ome	Browse the collection	
ticle title search uthor search eyword search ull text search dvanced search	List of journals:	
	Abasebenzi	
	Afro Newsletter	
	Africanist News and Views	
	Amandia-Matia	
	Apdusa Views	
icronyms hesaurus eed help?	Ariset Vukanit	
	Black Review	
	Clarion Call	
	Congress Resister	
	Contact	
	Cosatu News	
	Crisis News	

UNIVERSITY OF CAPE TOWN





What we need ...

Preservation

- Maintaining digital collections in perpetuity.
- Active curation of software/hardware/data.
- Hard? Is digital really durable?
- But maintain access/engagement
 Connecting people with resources.
 Use and contribution/enrichment.



How can we stop this trend?

Clearly, custom solutions are not working.

OSS is not working 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, |uma, !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 of Low Resource repositories 1/4

Minimalism

- Everything must be as simple as possible.
- Imposition on users
 - Do not force users to re-structure or re-number resources or fit a fixed repository model.
- No Internet
 - Internet is often expensive or unavailable.



Principles of Low Resource repositories 2/4

Simple Preservation

Just copy files, and be rescue-ready.

Any Objects

As long as the browser supports it.

Flexibility

Services should be generic so search/browse can apply to any data.



Principles of Low Resource repositories 3/4

Superimposed Information

- Layer complex services on top of simple lower layers.
- Hierarchical organisation
 - Allow the use of hierarchical information organisation.
- Platform agnostic
 - It must work on any system.



Principles of Low Resource repositories 4/4

Collection Building

Collections must be pre-processed, indexed and generated, so slow processing work happens in advance.



Technology Elements

XML/XSLT

- All data stored in simple files in hierarchical directories – no database.
- Static generation
 - Site is pre-created so software is not a long-term risk.
- Offline search/browse
 - Provides services through browser-based Javascript.



Approach 1: XML + XSLT \rightarrow XHTML

- Encode all descriptive information using XML files for objects, users, etc.
- Use XSL transformations to convert the XML metadata into corresponding HTML page views.
- Also, pre-process XML to create search/browse indices.



Approach 2: in-Browser Services





department of computer science

Approach 3: Selected Online Services

- Use online Web applications to add items, comments, annotation, etc.
- Web applications must modify XML and regenerate portions of static site.
 - No dynamic page generation.



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:

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.



department of computer science

GO

FHYA Prototype 2/3



Metadata

Title	Beadwork
	[Source of title : Nessa Leibhammer using JAG materials]
Material Designation	Object
	Textual record
Repository	Johannesburg Art Gallery (JAG)
Identifier	JL-U-8
Arrangement	[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



department of computer science

FHYA Prototype 3/3



About FHYA | Using the FHYA site | Contributors | Search | Contact Us

Search Results

Query	
zulu beadwork	1
repository	
AI	D
subcollection	
JAG/BRENTHURST	0
curationactor	
A	Ð
custodyactor	
Al	D

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
- Beadwork JAG/BRENTHURST/JL-U-14.xml
- 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



Reflections 1/2

- Databases one size does not fit all!
- We are not all online or online all the time.
- Some simple things are more difficult and vice versa.
 - Preservation is simpler.
 - New services are possibly more complex.





Scalability

- Not as good as SOLR, but fast responses for 100k items so many people will not care.
 paper at ICADL 2019)
- Thinking about resource limits suggests radically different software architectures.
 - This will ultimately benefit everyone!



that's all folks!

