Digital Libraries and Development

Hussein Suleman
hussein@cs.uct.ac.za

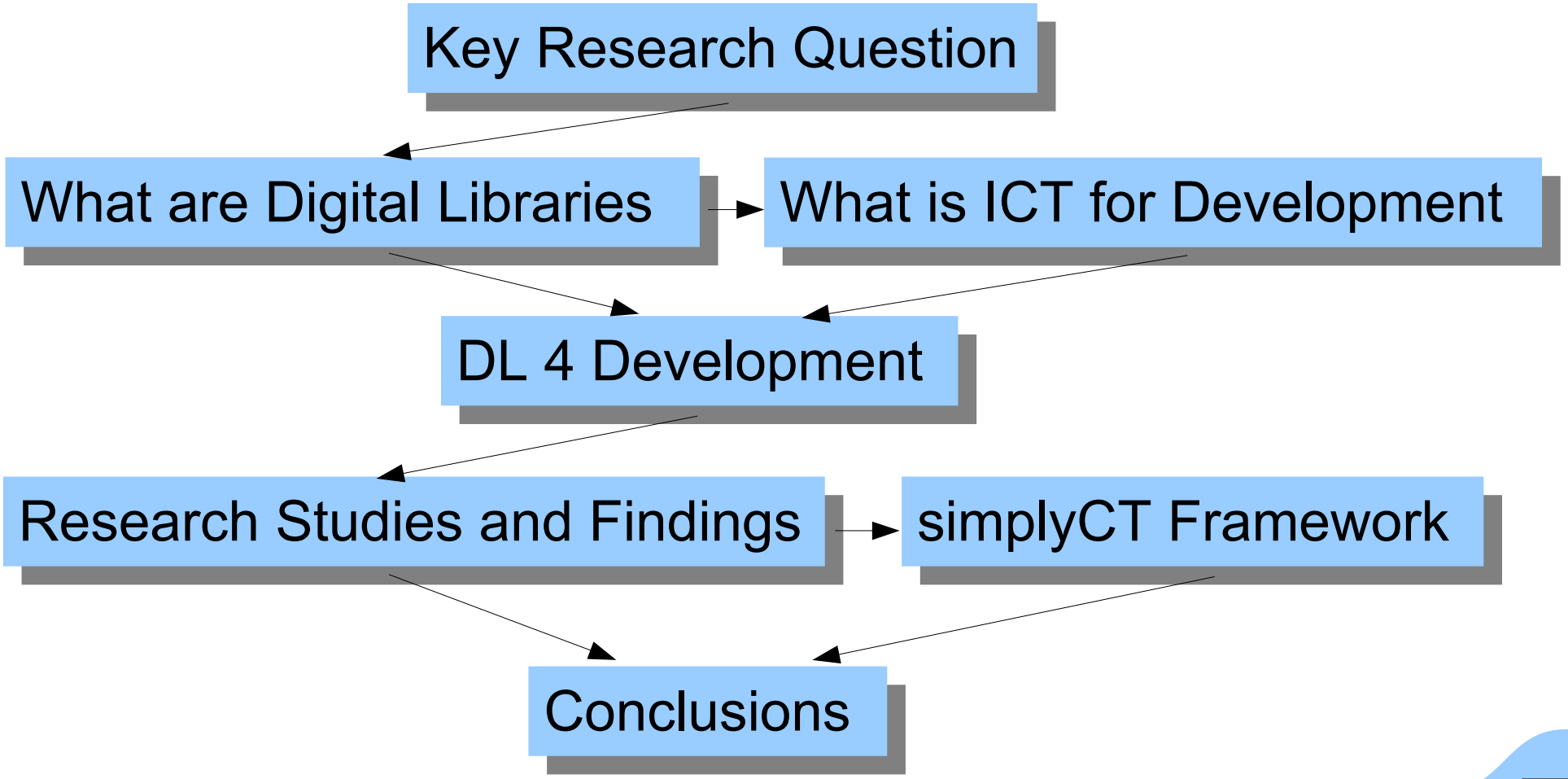
University of Cape Town
Department of Computer Science
Centre for ICT4D / Digital Libraries Laboratory

August 2017

Key Research Question

How do we build **Digital Libraries** to support **Development** in Africa?

Outline of Talk

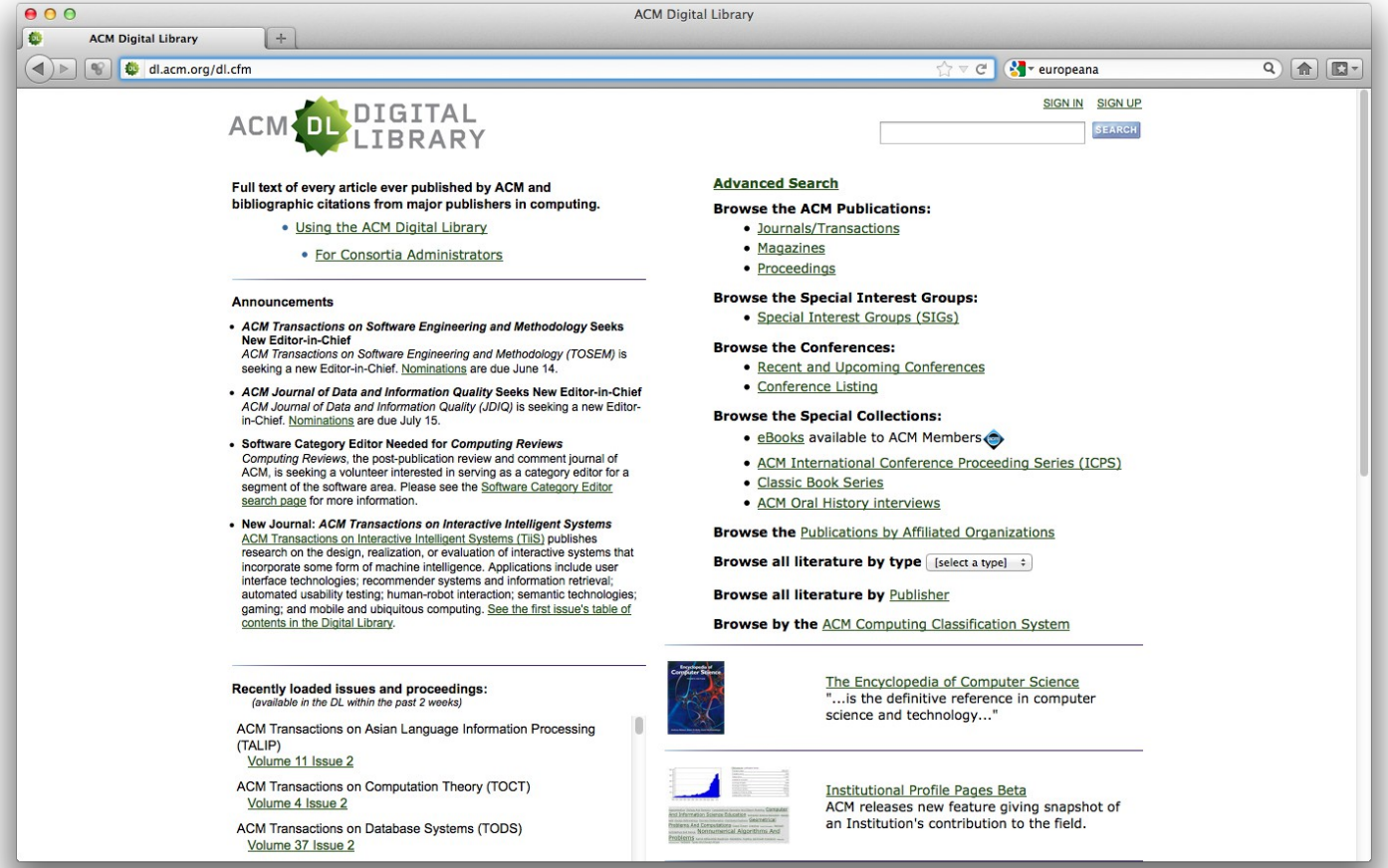


What are Digital Libraries

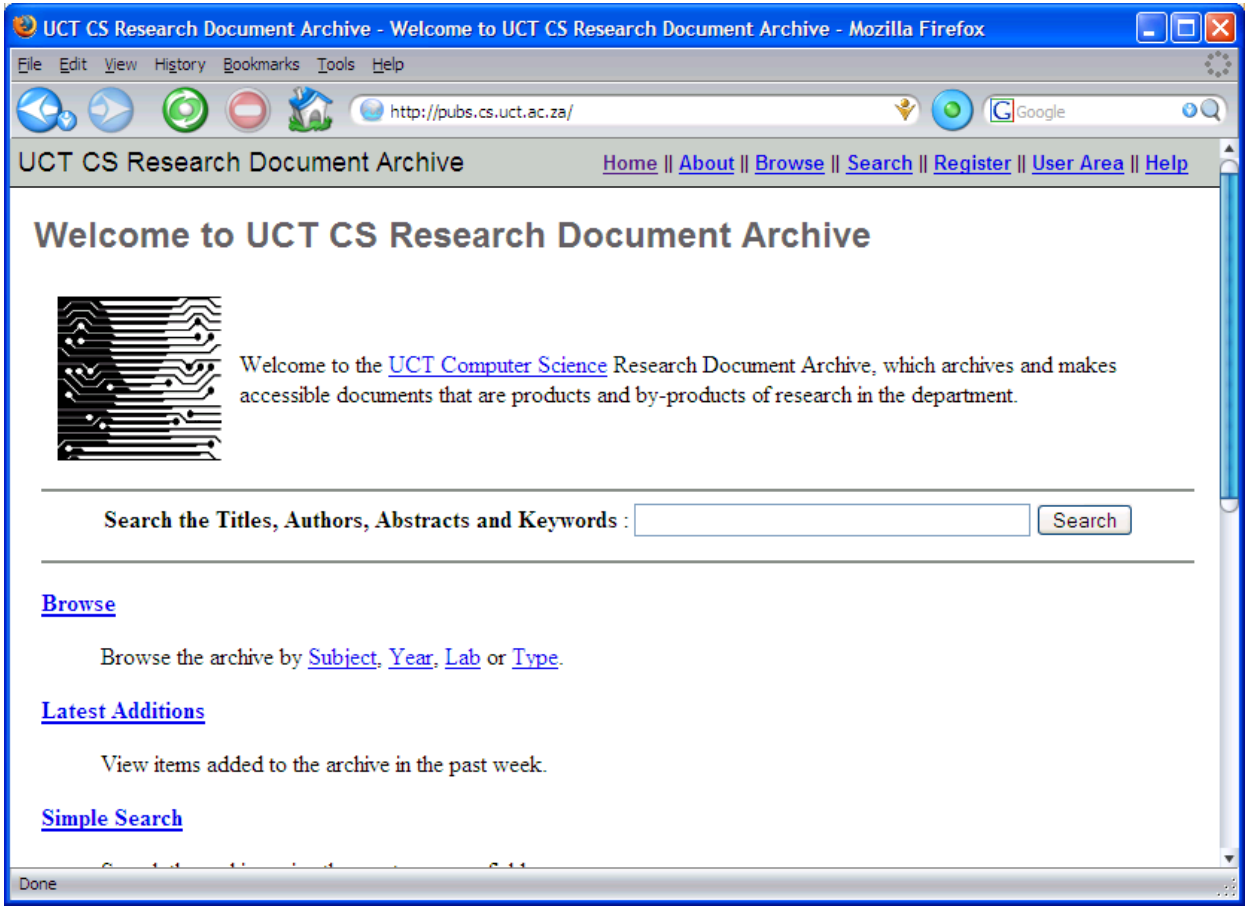
What is a Digital Library: Example 1/5

What is a Digital Library: Example 2/5

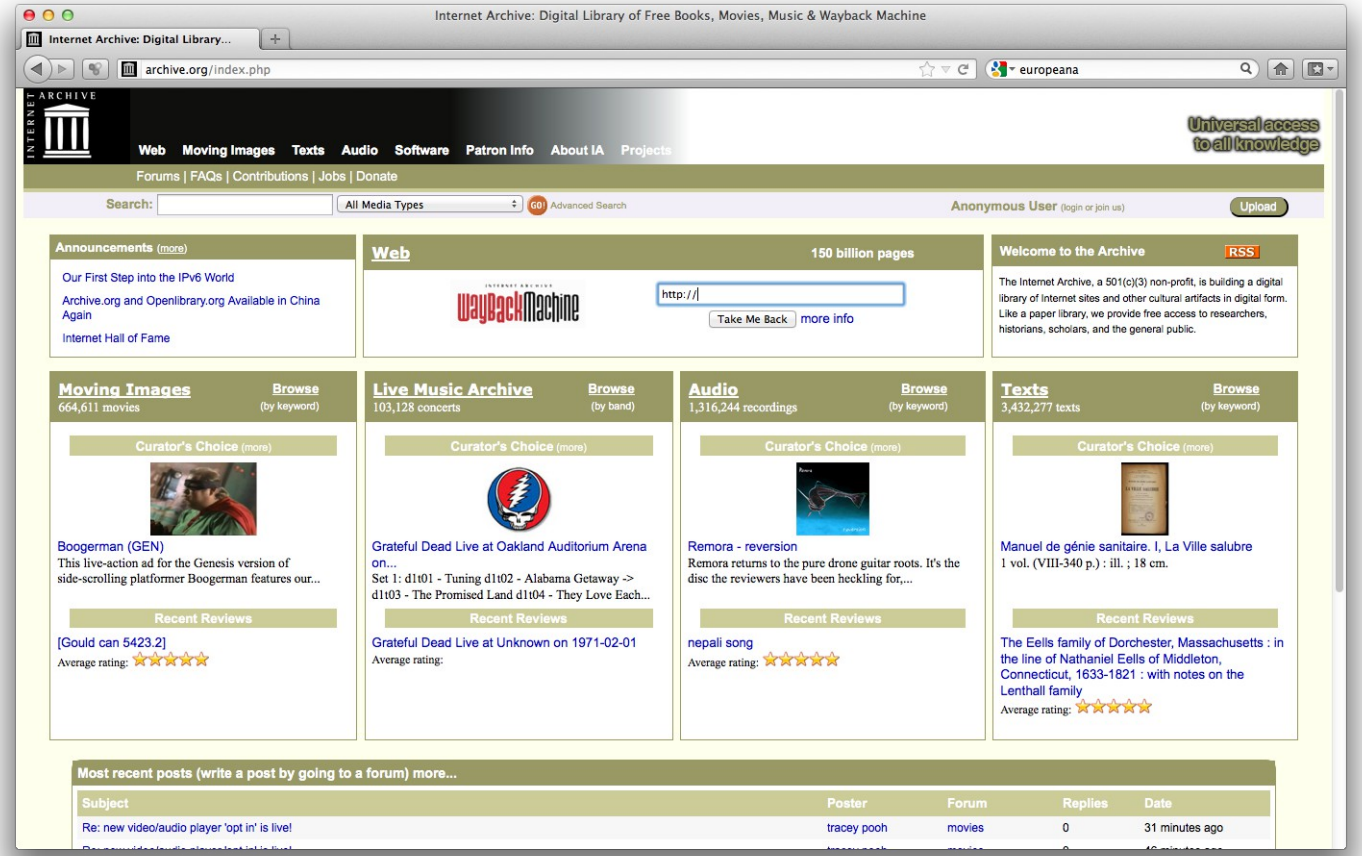
What is a Digital Library: Example 3/5



What is a Digital Library: Example 4/5



What is a Digital Library: Example 5/5



Defining Digital Libraries 1/3

- "Digital libraries are complex data/information/knowlege (hereafter information) systems that help: satisfy the information needs of users (societies), provide information services (scenarios), organize information in usable ways (structures), manage the location of information (spaces), and communicate information with users and their agents (streams)."
  - Fox, Edward A. (1999), DL Self-Study: definitions. Available http://ei.cs.vt.edu/~dlib/def.htm

Defining Digital Libraries 2/3

- "Systems providing a community of users with coherent access to a large, organized repository of information and knowledge."
  - Lynch, Clifford and Hector Garcia-Molina (1995), "Interoperability, scaling, and the digital libraries research agenda: A report on the May 18-19, 1995 IITA Digital Libraries Workshop".

- "The virtual or digital library is not an oxymoron-it is redundant. ... Since we did not bother to qualify our libraries by calling them clay libraries or papyrus scroll libraries, why now do we have to call them digital libraries?"
  - Braude, Robert and Samuel J. Wood (1999), "Virtual or actual: The term library is enough", Bulletin of the Medical Library Association, p. 87.

Defining Digital Libraries 3/3

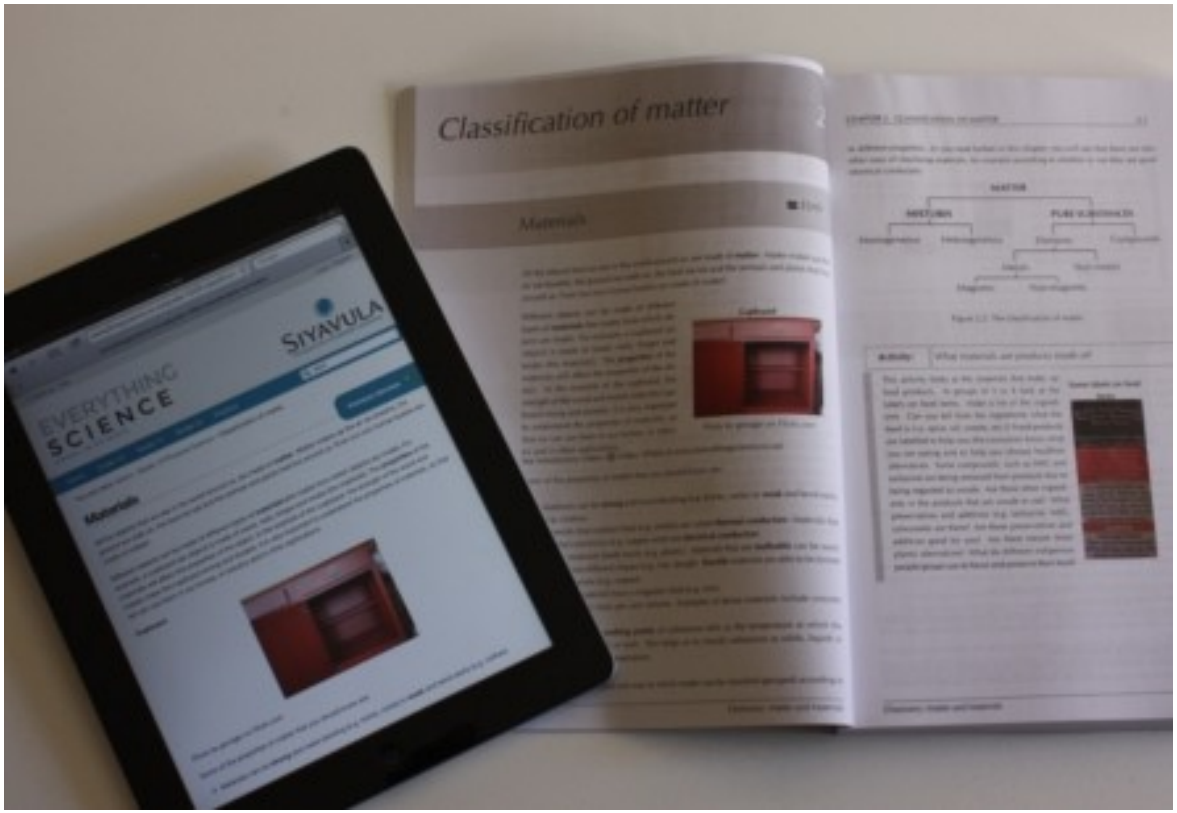
- A digital library is "a world of literature, history, photographs, movies and maps open, free of charge, to any curious mind that wants to meander through the electronic equivalent of library stacks."
  - Lipkin, Richard (1995), "The library that isn't there: Digital libraries transform books, photos, and videos into bits and bytes", Science News, Vol. 147, No. 22, pp. 344-346.

Typical DL Services

- User Management: accounts, auth, profile
- Searching: info retrieval, Google, indexing
- Browsing: categories, classification, subsets
- Submission: explicit/harvested/crawled
- Review: quality, workflow
- Annotation: reviews, ratings, discussions
- Recommendation: suggestions, collab filtering

What is ICT for Development

What is ICT4D: Example 1/3



What is ICT4D: Example 2/3

PATIENT'S BLOOD GROUP

KNOWN ALLERGIES

VACCINATION RECORDS

Vaccine	Date Received	Given by (signature)	Effective for approx	Booster/ Revaccination due date
HUMAN NORMAL IMMUNO-GLOBULIN For Hepatitis A 2ml			1-3m	
5ml			3-5m	
HEPATITIS A				
Dose 1	30.1.95		boost 6-12m	
Booster	21098		≤10y	2015 2005
HEPATITIS B**	11.8.00	KDWm		
Dose 1			N/A	
Dose 2	14/2/11	KDWm	N/A	
Dose 3			≤5y	
Dose 4 (IF REQUIRED)				
COMBINED HEPATITIS A & B				
Dose 1			N/A	
Dose 2			N/A	
Dose 3			≤10y hepatitis A	
			≤5y hepatitis B	

Vaccine	Date Received	Given by (signature)	Effective for approx	Booster/ Revaccination due date*
MENINGITIS (A+C) W Y	21.2.95		5y	
	14/2/11	KDWm	3-5yrs	
POLIO'	1995 & Yellow fever		10y	
RABIES				
Dose 1			N/A	
Dose 2			N/A	
Dose 3			2-3y	
DIPTHERIA/ TETANUS BOOSTER	22/2/11		10y	
	23.1.95		10y	
TETANUS' BOOSTER				
TYPHOID	21098		3y	22/2/11
	21/1/05	UP	3yr	(3yr)
YELLOW FEVER	1995		10y	
	14/2/11	KDWm	10 yrs	2021

As booster recommendations may change from time to time, check with your doctor or nurse before travelling that your immunisations are up-to-date.
*Some people may require a booster before date suggested. Check with your GP.
'Children under the age of 2 years: effective up to 2 years.
'Primary course must have been completed previously
'Hepatitis B schedule 0,1,6.

What is ICT4D: Example 3/3

Defining Development 1/3

- Development entails a modern infrastructure (both physical and institutional), and a move away from low value added sectors such as agriculture and natural resource extraction.
- Developed countries usually have economic systems based on continuous, self-sustaining economic growth and high standards of living.

Aronson (2006) Development: Definitions and Assumptions

Defining Development 2/3

"...development is a complex multi-level ongoing process for all individuals, groups, organizations and societies. It is not something possessed by one group of countries as distinct from another. I therefore find the linguistic distinction of 'developed' and 'developing' countries rather unfortunate and in some ways offensive. It is patronising to the latter group, and an inaccurate description of the former."

Walsham (2005) Development, global future and IS research: a polemic, J Strategic Info Sys.

Defining Development 3/3

- "Development requires the removal of major sources of unfreedoms: poverty as well as tyranny, poor economic opportunities as well as intolerance or overactivity of repressive states."

Sen (1999) Development as freedom, Oxford Univ Press.

Digital Libraries for Development

The African Context/Question

- Do we as Africa need to innovate:
  - more?
  - differently?
  - faster?
  - more urgently?

- Do we build digital libraries differently in Africa?
- How do we build DLs for Development?

Examples: What We Do Not Need

- applications that waste bandwidth
  - e.g., video where text will work fine
- solutions that require lots of money+staff
  - e.g., outsourced digital archives
- expensive technology
  - e.g., replicated NAS/SAN storage
- solutions that lock us in
  - e.g., iPad-specific solutions

Principles of DL4D

- Efficient bandwidth use
- Advanced technology
- Appropriate technology
- Local relevance
- Modernization instead of Africanization
- Global applicability of solutions
- Minimalism of staff/money
- Multicultural/multilingual inclusivity

Implementing DLs for Development

recent research at UCT

- cloud computing for archives
- bandwidth-sensitive information applications
- multilingual IR
- heritage preservation
- rock art sites navigation
- mobile IR in African languages
- social media analytics for democracy

Experiments and Findings

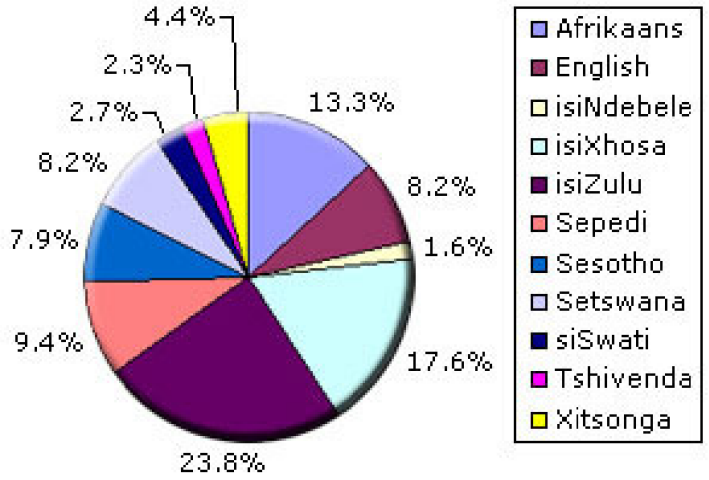
Cloud-based Archives

"It was much nicer before people started storing all their personal information in the cloud."

Lebeko Poulo, Lesotho
Mushashu Lumpa, Zambia

- individual services and whole archives in private clouds
- install locally
- reduces need for skilled staff
- instant archives
- shared resources
- automatic scalability
- acceptable performance, after cache priming
- user studies in progress

Multilingual Information Retrieval

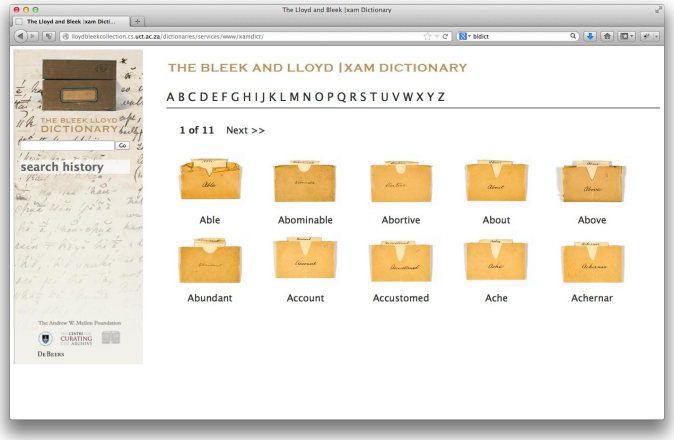


(source: SA Info)

Mohammed Mustafa Ali, Sudan

- search queries with multiple languages
- current systems biased to one language
- rerank documents by understanding query and reweighting languages/results
- better quality results found, higher up in results

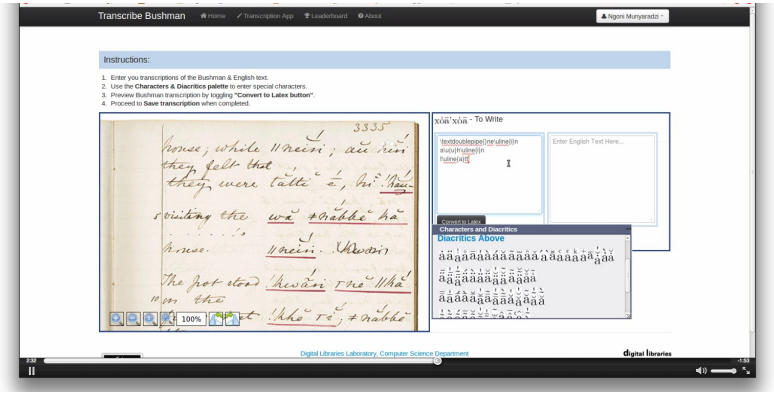
Language Preservation: Online |Xam dictionary



- visual dictionary of |Xam language
- simple archive foundation
- client-side processing as far as possible
- linked into Bleek and Lloyd

Kyle Williams, South Africa

Document Transcription: Bleek and Lloyd Stories



Ngoni Munyaradzi, Zimbabwe

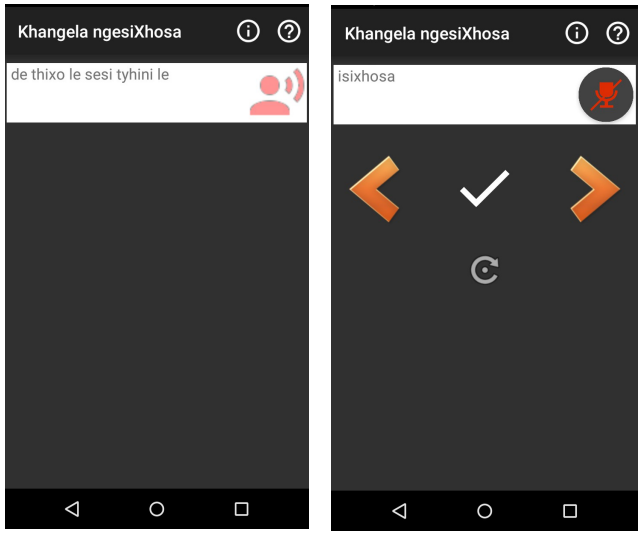
- crowdsourced transcription application
- volunteers to convert images to text
- automated algorithms to check and assess quality
- interactive Web interface for users to enter text
- 10% better than AI approaches!

Mobile Navigation in Rock Art Sites

Ayodeji Olojeda, Nigeria

- rock art sites are in remote locations, poorly marked and protected
- mobile phone with image processing and content-based IR
- detect image and provide contextual information
- tested successfully with rock art in Western Cape

Voice-based IR interface in isiXhosa



- 8 million isiXhosa speakers, low literacy levels
- Mobile application to search using voice
- trained using emerging speech corpora
- successfully evaluated for usability of voice interface

Morebodi Modise, Botswana

Social Media Analytics for Democracy

- analyse social media, such as Twitter to detect attributes:
- how democratic is a society?
- is there freedom of expression?
- are elections free and fair?
- is there equality in society?

Selvas Mwanza, Zambia

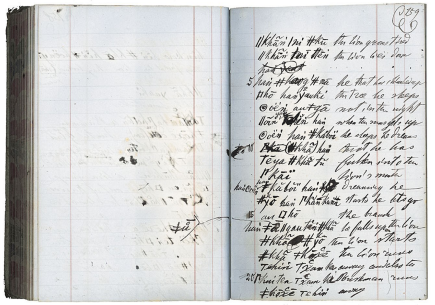
simplyCT Framework

Principled Architecture of Modern DLSs

- Minimalism
- No imposition on users
- Preservation by copying
- Web or no Web
- Generality
- Reduce abstractions

simplyCT is a framework for simple online or offline digital archives with preservation, standardisation and extensibility being major focii.

- /archive
  - file1.jpg
  - file1.jpg.metadata
  - file2.jpg
  - file2.jpg.metadata
- /index
  - search.1/ ...
- /service
  - onlinesearch.1/ ...
  - offlinesearch.1/ ...
- /static
  - file1.html



```
<oai:dc
xmlns="http://purl.org/dc/elements/1.1"
xmlns:oai="http://www.openarchives.org/OAI/2.0/oai_dc/">
  <title>...</title>
  <creator>...</creator>
  ...
</oai:dc>
```

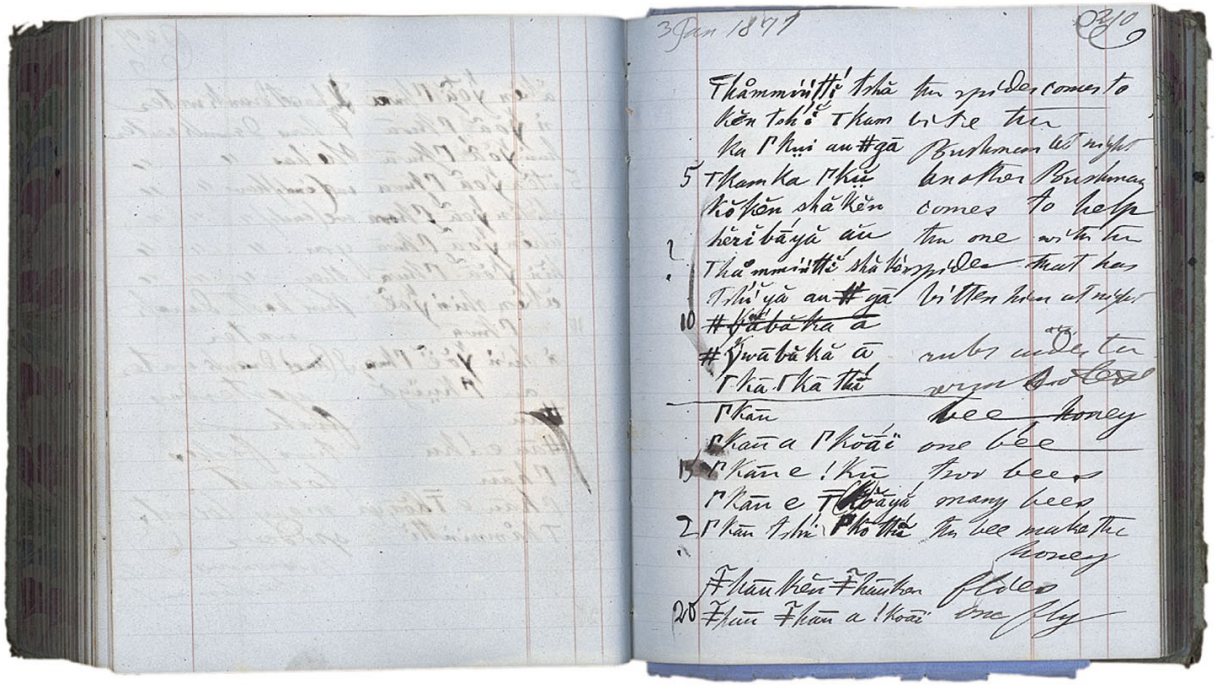
simplyCT Experiments 1/3

- Bleek and Lloyd Collection
  - Database-less custom repository
  - Proof of concept

- Caljax
  - Hybrid online-offline collection
  - Minimal network use
  - Transparent integration of information

Bleek and Lloyd Collection 1/2

- Books and drawings documenting now-extinct culture of |xam and !kun KhoiSan groups.
- Documented by Wilhelm Bleek, Lucy Lloyd and others in late 1800s in Cape Town.
- ~20000 page images

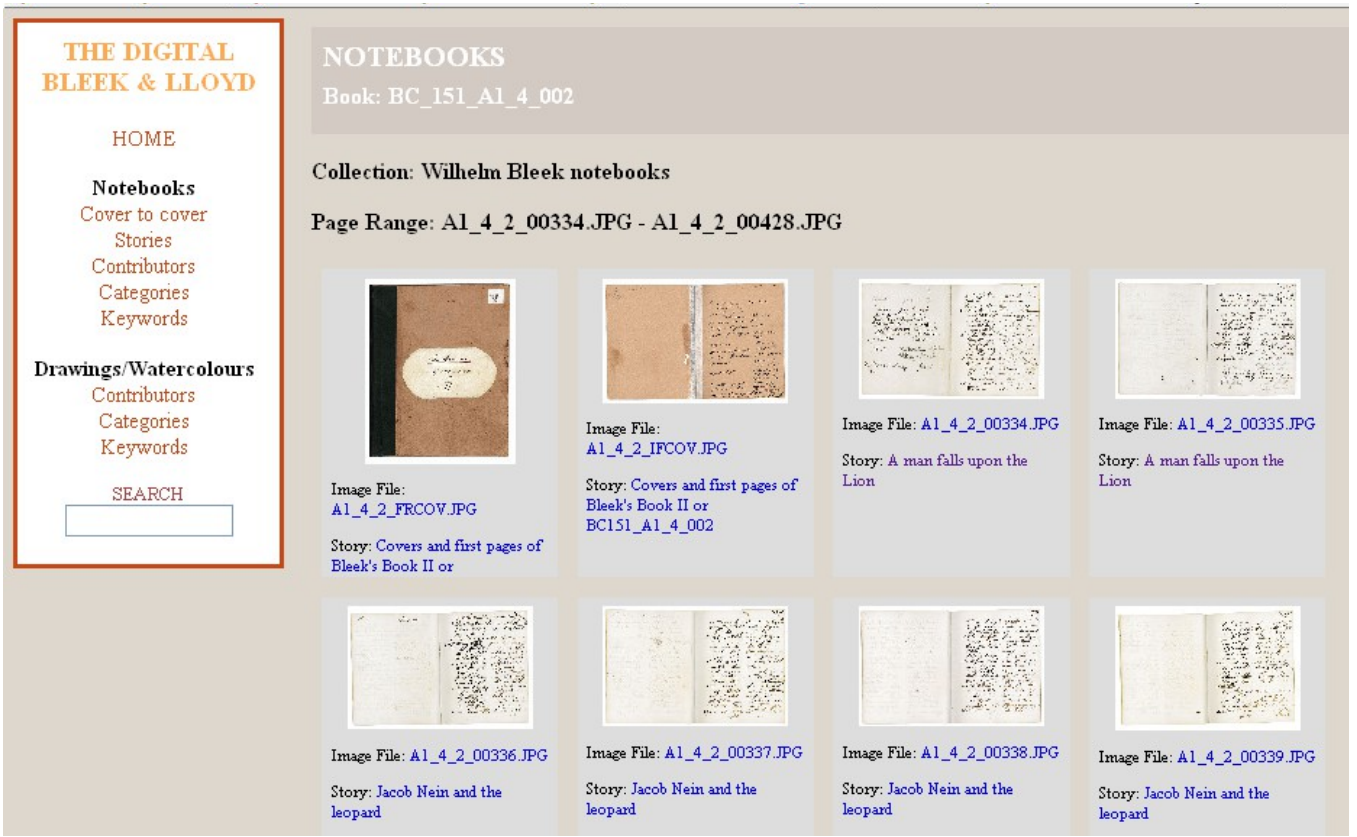


Bleek and Lloyd Core Requirements

- Make the collection accessible as widely as possible:
  - Over the Web,
  - Off a CDROM,
  - Off a network-shared drive,
  - Etc.
- Platform independence (Mac/XP/Linux/etc.).
- Low barrier to use.
- Standards-compliance.

Bleek and Lloyd Technology

- Pre-generated XML+XSLT->HTML.



Bleek and Lloyd Technology

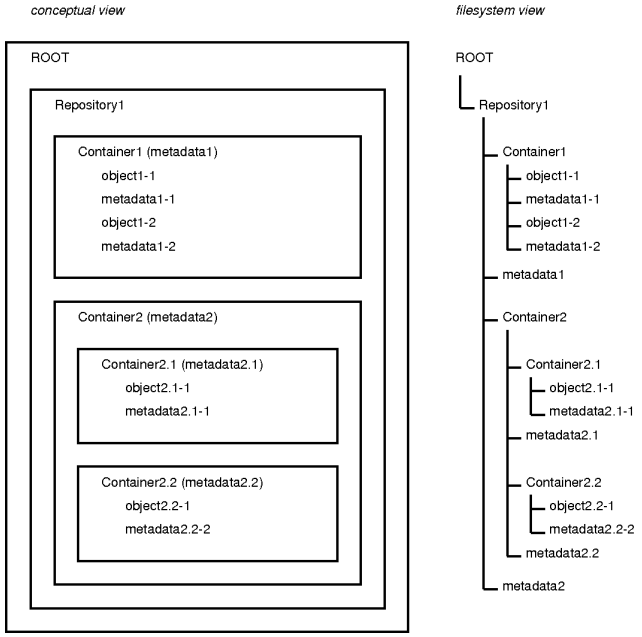
- Javascript IR engine within-browser.

simplyCT Experiments 2/3

- simplyCT IR
  - Institutional Repository software
  - Proof of concept

- Bonolo
  - Generic Web-based curation system
  - Positive user experience results
  - Some performance concerns

simplyCT Experiments 3/3



- performance
- understandability
- flexibility
- applicability

- good performance for small to medium collections
- easy to use and expand

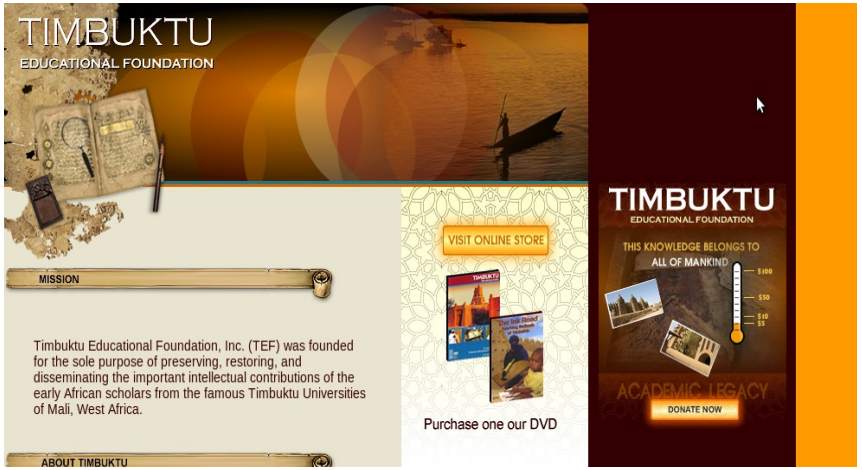
Phiri Lighton, Zambia

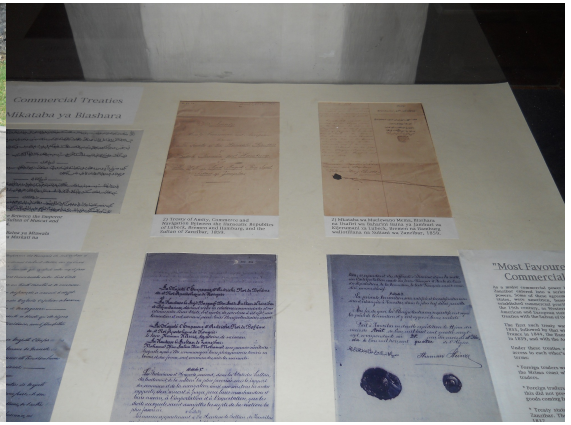
Conclusions

Where we are now

- We are not special. We produce local solutions with global applicability.
- Modern DLs can focus on different design philosophies e.g., scalability.
- African DLs can be driven by the development imperative.
  - This can and should influence the global design process.
- Still a lot to be done ...

Where to in Future





questions, comments, ...

Google "hussein suleman"

Facebook/slumou

Twitter@slumou

hussein@cs.uct.ac.za