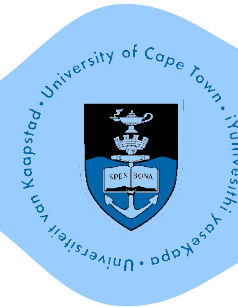
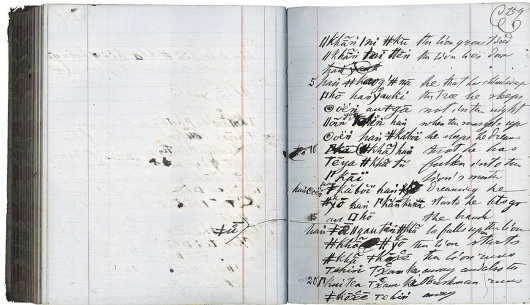


# Simple Cultural Heritage Preservation

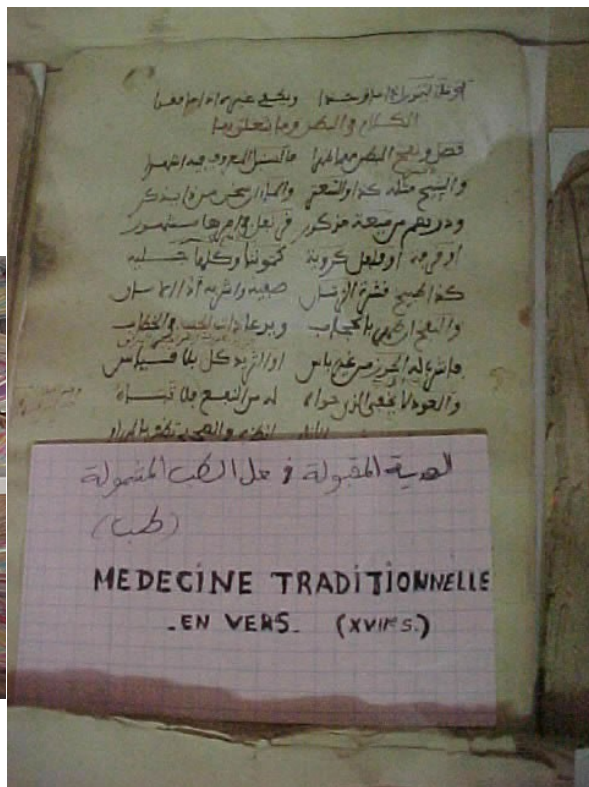


*Hussein Suleman*  
*hussein@cs.uct.ac.za*

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

*February 2013*

# Motivation



Timbuktu Manuscripts  
<http://www.timbuktufoundation.org/Manuscripts/index.htm>





# Anti-Motivation

## Budget Europeana 2011-2015

Year	2011	2012	2013	2014	2015
<b>INCOME (thousands)</b>					
→ Operating Income	4.923	4.978	4.856	5.261	5.504
* Long term subsidy European Commission	-	-	-	4.661	4.894
* Subsidy from projects	4.061	4.336	4.241	480	488
* Subsidy Ministries to support projects & overhead	857	637	610	110	112
* Other Income	5	5	5	10	10
<b>Total operating income</b>	<b>4.923</b>	<b>4.978</b>	<b>4.856</b>	<b>5.261</b>	<b>5.504</b>





# Anti-Motivation

49  
full-time and  
part-time staff

www.pro.europeana.eu/web/guest/about/staff







Europeana professional

Home About us Press & Events Projects Europeana Foundation Europeana Network Provide data

Europeana Foundation

- Executive Committee
- Board
- Funding and Sponsorship
- Friends of Europeana
- Staff
- Jobs
- Facts & Figures

### Staff

	<p><b>Ad Pollé</b> Senior Project Coordinator ad.polle@kb.nl</p>		<p><b>Albert Verhaar</b> Financial Controller albert.verhaar@kb.nl</p>
	<p><b>Alastair Dunning</b> The European Library Programme Manager alastair.dunning@kb.nl</p>		<p><b>Andrew MacLean</b> Senior Software Developer andrew.maclean@kb.nl</p>
	<p><b>Anna van den Broek</b> Marketing Assistant anna.vandenbroek@kb.nl</p>		<p><b>Annette Friberg</b> Business Development Manager annette.friberg@kb.nl</p>





# MORE With LESS ?

- Open Source Content Management Systems
  - Drupal
  - Joomla!
  - ...
- Open Source Digital Repository Systems
  - DSpace
  - EPrints
  - Greenstone
  - ...





# Have you ever configured or maintained DSpace?

## HTTP Status 500 -

---

**type** Status report

**message**

**description** The server encountered an internal error that prevented it from fulfilling this request.

---

Apache Tomcat/7.0.35



## 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.

## 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.





## Internet Bandwidth (Digital Divide)



- ❑ Non-existent in some places and poor everywhere else.
- ❑ Preservation projects such as LOCKSS are not suitable in most cases.
- ❑ All online solutions must be bandwidth-friendly.



## Is (South) Africa Special?

- ❑ Definitely NOT!
- ❑ The same problems are faced by some other communities.
- ❑ Many communities face some of the problems.
- ❑ Most communities can benefit from solutions to these problems.





# Inspiration

**Gutenberg**

search book catalog

Author:

Title Word(s):

- Book Search
- Catalog
- Bookshelves

search site

Term:

- Main Page
- Categories
- News
- Contact Info

From Project Gutenberg, the first producer of free ebooks.

[Mobile Site](#) · [Book search](#) · [Bookshelves by topic](#) · [Top downloads](#) · [Recently added](#) · [Report errors](#)

**Project Gutenberg** is the place where you can download over 33,000 [free](#) ebooks to read on your PC, [iPad](#), [Kindle](#), [Sony Reader](#), [iPhone](#), [Android](#) or other portable device.



We carry high quality ebooks: Our ebooks were previously published by *bona fide* publishers and digitized by us with the help of thousands of volunteers.

All our ebooks can be found in HTML and simple text.

No fee or registration is required. If you find them useful, we kindly ask you to help. Other ways to help include [reporting errors](#).

Over 100,000 free ebooks

“1. Electronic Texts (Etexts) created by Project Gutenberg are to be made available in the simplest, easiest to use forms available.”

*The History and Philosophy of Project Gutenberg by Michael Hart*





## Seeking Simplicity ...

- Lightweight and Reusable Systems
  - Flat files / NoSQL, XML+Javascript
  - Minimalist Archives
- Metadata management using office suite.
- Collections accessible over CD/DVD-ROM, local drives, network, etc.
  - Preservation by copying.
- Static collections rather than dynamic.
  - Preserve files instead of services.
- Minimal bandwidth use e.g., using AJAX/Comet





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





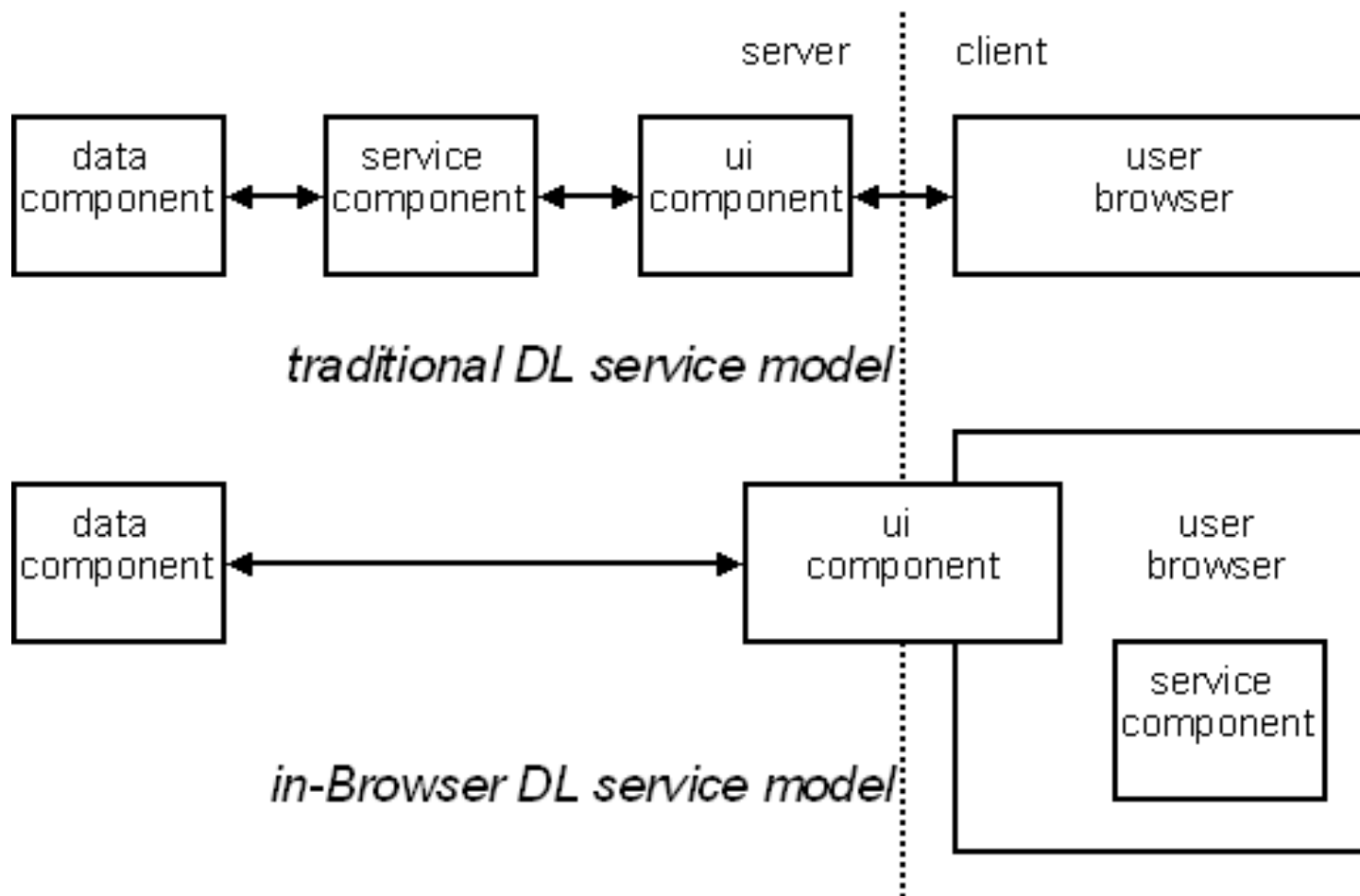


## Solution 1: XML + XSLT + XHTML

- 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.



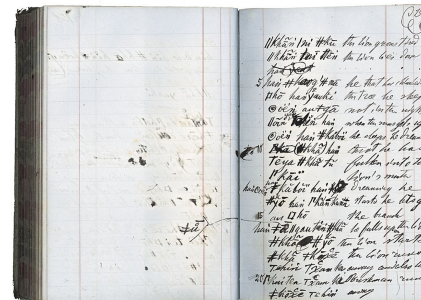
## Solution 2: in-Browser Services





## □ /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>
```



# Principled Architecture of Modern DLSs

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





# Experimental Systems 1/2

- Bleek and Lloyd Collection
  - Database-less custom repository
  - Proof of concept
  
- Caljax
  - Hybrid online-offline collection
  - Minimal network use
  - Transparent integration of information





## Experimental Systems 2/2

- simplyCT IR
  - Institutional Repository software
  - Proof of concept
  
- Bonolo
  - Generic Web-based curation system
  - Positive user experience results
  - Some performance concerns





## Status

- about 8 projects using simplyCT framework
- very different scenarios
- no stumbling blocks
- no databases in most cases
  - minimise use of software dependencies
- in progress (as of early 2013): experimental validation of efficiency
  - works quite well for medium-size collections (<1million objects)





# Future Challenges

- Standard tools to manage metadata/data?
  - Usability
  - Scalability
  - Extensibility
  - System independence
  
- More case studies in the pipeline ... but we are almost ready to field-test this with other projects.





# Future ?



**TIMBUKTU**  
 EDUCATIONAL FOUNDATION

**MISSION**

Timbuktu Educational Foundation, Inc. (TEF) was founded for the sole purpose of preserving, restoring, and disseminating the important intellectual contributions of the early African scholars from the famous Timbuktu Universities of Mali, West Africa.

**ABOUT TIMBUKTU**

VISIT ONLINE STORE

Purchase one our DVD

**TIMBUKTU**  
 EDUCATIONAL FOUNDATION

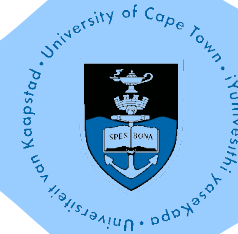
THIS KNOWLEDGE BELONGS TO ALL OF MANKIND

**ACADEMIC LEGACY**

DONATE NOW



**that's all folks!**



*questions?*

*search for "hussein suleman"  
on Google or Facebook*

*hussein@cs.uct.ac.za*

*remember to Tweet :) ... #sadi2013*