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

## Europeana Professional - Strategy
http://www.pro.europeana.eu/web/guest/strategy

### Budget Europeana 2011-2015

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

49 full-time and part-time staff
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?
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

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.
Case Study: Bleek and Lloyd Collection

- Books and drawings documenting now-extinct culture of |xam and !kun speakers.

- Documented by Wilhelm Bleek, Lucy Lloyd and others in late 1800s in Cape Town.

- 14000+ page images, 800 images, dictionary, etc.
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

```xml
<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?
that's all folks!

questions?

search for “hussein suleman”
on Google or Facebook

hussein@cs.uct.ac.za

remember to Tweet :) ... #sadi2013