The Wits Digitisation Experience: Lessons Learned and Suggestions for National Collaboration

Michele Pickover
Charl Roberts
Outline

• Background to the Wits Digitisation Centre
• What services the Centre delivers
• Sustainability of the Centre
• Integrated Resource Discovery
• Policy at Wits
• Challenges to Collaboration
• Suggestion for a way forward for a South African Digital Libraries and Archives Portal
Background to the Wits Digitisation Centre

- Wits as a hub of academic information
- 2009 – strategic needs assessment
- 2010 Fact Finding mission (visited Cornell, University of Michigan, National Archives, UK, Oxford digitisation Centre)
  - Physical infrastructure and hardware requirements
  - Workflow and digitisation processes
  - Policies
  - Human Resources
Some pictures – fact finding mission
The Wits Digitisation Centre

• Completed and went into production in 2011
• Housed in the William Cullen Library
• Houses equipment, conducts digitisation activities for online access and preservation
• Training Laboratory
• Technical Support
Picture of the completed centre
Cont.
Services of the Digitisation Centre

• Providing quality digitisation and planning for access, research and preservation
• The use of specialised technologies and resources for increased access
• Custom digitisation projects for the university and community
• Project planning and co-ordination
Services provided through:

• State of the art equipment and lab facilities
• New technologies, standards and best practices for preservation and access
• Expertise in text recognition (OCR)
• Expertise in metadata creation
• Essential instructions, processes, specific outputs and quality control procedures expected from service providers (outsourcing is a viable, sometimes essential, aspect of digitisation)
Cont.

• Expertise in digitisation workflow processes
• Expertise in budgeting and project costing
• Expertise in the handing of fragile manuscript and document handling, as well as the preparation of physical originals
• Experience in digitisation problem solving
• Expertise in content selection, preparation and description
The NRF / Carnegie Digitisation Centre

• Currently a SLA is under discussion
• centre to provide training, workflow and policy development and assist to build digitisation capacity building
• Will be hosted at the Wits digitisation centre
Examples content from WIREDSPACE and Historical Papers at the Wits Library

- WIREDSPACE – Wits History Workshop Collection

<table>
<thead>
<tr>
<th>Statistics</th>
<th># Of Visits For the Top 5 Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>Views</td>
</tr>
<tr>
<td>Problems of democratic governance in Nigeria: the past in the present</td>
<td>273</td>
</tr>
<tr>
<td>The United Democratic Front and township revolt: South Africa</td>
<td>129</td>
</tr>
<tr>
<td>Reconciliation and revenge in post-apartheid South Africa: rethinking legal pluralism and human rights</td>
<td>88</td>
</tr>
<tr>
<td>The 'Labyrinth of my history': the struggle with filiation in J.M. Coetzee's 'Dusklands'</td>
<td>80</td>
</tr>
<tr>
<td>The 1920 black mineworkers strike</td>
<td>76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Countries</th>
<th># Of Visits From The Top 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>6,000</td>
</tr>
<tr>
<td>UK</td>
<td>4,000</td>
</tr>
<tr>
<td>China</td>
<td>2,000</td>
</tr>
<tr>
<td>Canada</td>
<td>1,000</td>
</tr>
<tr>
<td>South Africa</td>
<td>1,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continents</th>
<th># Of Visits From The Top 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>5,000</td>
</tr>
<tr>
<td>Asia</td>
<td>4,000</td>
</tr>
<tr>
<td>Europe</td>
<td>2,000</td>
</tr>
<tr>
<td>Africa</td>
<td>1,000</td>
</tr>
<tr>
<td>Others</td>
<td>1,000</td>
</tr>
</tbody>
</table>
More examples

- WIREDSPACE – Architecture Archives

<table>
<thead>
<tr>
<th>Items</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>98</td>
</tr>
</tbody>
</table>

- # Of Visits From The Top 5 Countries
- # Of Visits From The Top 5 Continents
Architecture Collection examples
Architecture Collection examples
Historical Papers collection
Examples from the Historical Papers website
Historical Papers collection cont.
Historical Papers collection cont.
Sustainability plan for the Centre and Digital Content

- Supported by Wits at the highest level, support for:
  - Growth
  - Provision of capacity
- Digitisation is a major part of the Wits Library Strategy and Business Plan
- As such it is an initiative which funding is allocated for annually (depending on the financial ability of the University)
- The Library continually reviews staffing structures to allow redeployment for digitisation efforts where needed
- In partnership with CNS – Storage requirements, long term preservation of digital masters
Historical Papers Research Archive

- Conversion of Online Finding Aids
- Compliance with international standards, resources properly searchable and data harvestable
- XML/EAD (Encoded Archival Description)
- Implementing ICA-Atom - [https://www.ica-atom.org/](https://www.ica-atom.org/)
  - Web-based, scalable open source
  - Standards based archival description
  - Multilingual, multi-repository environment
  - Supports Dublin Core, MODS and METS
  - Flexible - supports descriptions of various types of cultural materials from libraries, archives, museums, galleries and special collections
Resource Discovery of Institutional Digital Content

• Dynamic, integrated and searchable resource discovery portal for institutional digital content
• Initially library and archival content
• Aggregate and open up our substantial collection of digital assets into a single object store
• Searching and browsing within discrete curated collections and across the entirety of our content
• Bring our historical digital content up to date
• Enrich our metadata where necessary
• Provide and an interface for accessing the Open Linked Data
• Roll out Linked Data framework to item level to open up searching across Wits libraries, archives, museums and galleries
Resource Discovery - Example

Digitised images from the Bodleian Libraries Special Collections

Search Digital Images

Subject Categories:
- All
- Library collections
- Asian and South Asian collections
- Children's books
- Ephemera
- Hebrew collections
- History and politics
- Incunable and printed books to 1500
- Literature
- Maps and local history
- Medieval and Renaissance collections
- Middle East and Islamic collections
- Music
- Prints and portraits
- Science, medicine, and technology
- Theology and the Bibles

Western books and manuscripts after 1600

Blockbook
- Early Printing in Europe
- Greek manuscripts
- Masterpieces of the non-Western book

Medieval and Renaissance Manuscript Illumination (from 35mm)
- A Nation of Shopkeepers: Trade Ephemera from 1654 to the 1860 in the John Johnson Collection
- Arthur Evans Archive (Knossos excavations)
- Bodleian Library/Toyota City Imaging Project

Bringing Laxton to life
- Broadsides: Ballads Online Database
- Burmese life and devotion
- Citizen Milton
Policy Development for Digitisation and Digital Content

Wits Library
- institutional Repository Policy
- Electronic Theses d Dissertations Policy
- Wits Library Digitisation Policy

Wits GLAMs Initiative
- Wits as Memory Institution
- Responsible and dynamic management of heritage collections
- Support of policymakers
- 2003 – survey/audit of its collections
- Long term strategies and integrated polices for Wits Collections
Cont.

Wits GLAMs (Galleries, Libraries, Archives and Museums) forum - 2011

• Guidance in terms of policy-formation, ethics, joint activities and digitisation
• A vision and strategy for the selection of content and digitisation of Wits’ GLAMs materials
• A Wits Digitisation and Digital Preservation Policy
• A Digital Preservation Programme
Challenges to Collaboration

• Institutions at different stages of development
• Using different specs, schemas and methodologies
• Need to develop an agreed upon approach to content selection, the aggregation of metadata and distinct sector methodologies
Content Selection

- Do institutions know what they have and what to digitise?
- Have you done an audit of your holdings for digitisation purposes?
- Have you met with your stakeholders, custodians and other role-players?
- Is preservation of content the main driver? Or is it Access? Or both?
- What about physical preservation of the original – will this be part of your digitisation plan?
Content Selection cont.

- Mass digitisation?
- Digitising entire collections?
- Only items from collections – cherry-picking?
- Will be developing a themed approach to content?
- Or all of these?
Content cont.

- What is the physical condition of selected content?
- The format?
- What is the extent of arrangement and description of collections - are there detailed inventories and finding aids or do they need to be enhanced for digitisation?
- Are there Deposit Agreements and do these cover your planned digitisation endeavours?
- Are there restrictions on use of parts of the collections?
- What are the copyright/intellectual clearance issues?
Content Cont.

Whose Agenda? Who Decides?

- Digitisation of heritage material a site of struggle
- Archives, are contested spaces
- Content selection is not a neutral act.
- Decisions about WHAT to digitise and what is digitised speaks to power, ideology and propaganda.
- What role is digitisation playing in framing research agendas, repackaging history and shaping national histories?
- Is digitisation of specific types of content reinforcing the dominant economic and political milieu and a linear nationalist narrative?
Digitisation and the Politics of Memory

- Digitisation projects should enhance public interest, service researchers in the South and promote South-South dialogue.
- The construction and creation of digital content must be interrogated and unpacked.
Distinct Sector Methodologies

• Libraries, museums and archives have different descriptive schemas, methodologies and metadata standards

• Need to unpack these differences if we are to collaborate

• Archival descriptions are not following ICA standards or using international archival community standards such as XML/Encoded Archival Description (EAD), General International Standard Archival Description (ISAD (G)), ISAAR (CPF), thesauri etc.,
Metadata and Schemas

We will need to:

• Improve quality of metadata to ensure better search results across all digitised content

• International best practice and lessons - e.g. LIBER, Europeana and the JISC RDTF Project

• Technical architecture and sustainability plan to enable discovery across all media types

• The benefits of Linked Open Data?

• Agree to provide information in a standard, open format and under an open licence

• Develop pragmatic metadata guidelines

• Metadata standards for text, pictures, audio and video for harvesting into a multi-institutional portal

• Agreement of standards for searching and semantic interoperability
Suggestions for a Way Forward

• Scope what a national resource discovery service should, and should encompass
• Undertake a pilot project – tertiary institutions?
• Set up Task Teams which would look at:
  1. A model for Collaboration and Governance
  2. Technical Architecture, including Metadata Standards
  3. Content Identification – including processes and principles.
  4. Licensing, legal, IP and copyright Issues.
  5. Lobbying, Fundraising and Long-term sustainability of the collaborative Resource
  6. Training. Training programmes, including academic curricula, must be developed to provide information professionals with the skills permitting them to implement both digitisation and preservation practices.
An entity such as SADI, in a more re-imagined form, with the inclusion of multi-stakeholders, could be the vehicle to drive this process.
Thanks!