Upscaling digitisation at the Wellcome Library showcasing the Goobi workflow system

Christy Henshaw
Programme Manager
Wellcome Digital Library

3rd LIBER-EBLIDA Workshop on Digitisation of Library Material in Europe

Koninklijke Bibliotheek
6 October 2011
The Wellcome Trust

• A global charitable foundation

• Achieving extraordinary improvements in human and animal health

• Supporting the brightest minds in biomedical research and the medical humanities

• Exploring medicine in historical and cultural contexts
The Wellcome Library

• Major resource for the study of medical history

• Collections of books, manuscripts, archives, films and pictures on the history of medicine from the earliest times to the present day.

• Provide insight and information to anyone seeking to understand medicine and its role in society, past and present.

• Provide access to a growing collection of contemporary biomedical information resources relating to consumer health, popular science, biomedical ethics and the public understanding of science.
The Wellcome Library
Digitisation – the story so far

- Image library created from transparencies/prints, and on demand photography – 300,000 images
- Journal backfiles digitisation – (funder) Med. Hist., BMJ, etc. in PMC
- Wellcome Film – 500+ titles (also Wellcome Film YouTube channel)
- AIDS posters project – 3,000 posters
- Arabic manuscripts – 500 manuscripts
- 17th century recipe books – 74 manuscripts
- Contributions to Europeana via the Europeana Libraries project, and World Digital Library
The Library Transformation Strategy 2009 - 2014

To provide global access to, and expert interpretation of, a world class collection that explores medicine in its cultural contexts

- **Targeted collecting** – putting challenges in context
- **Expert interpretation** – engaging (new) audiences
- **Strategic digitisation** – online access to our collections
The Wellcome Digital Library pilot 2010-2013

Genetics and its Modern Foundations
A new online resource for everyone interested in the history of human and animal health.

Aims

- **build sustainable/expandable mechanism** – foundation stone for WDL
- **digitise key library holdings** - relating to a major Trust challenge area
- **digitise important third party content** – linked to theme
- **use innovative content and tools** – to encourage discovery and use
- **explore commercial partnerships** – enhance access to non-theme material
Archival material – 900,000 images
Wellcome Library - 600,000 images
External – 300,000 images
Books related to genetic research - 600,000 images
ProQuest, Early European Books – 5.5m images
## Digitisation strategy

<table>
<thead>
<tr>
<th>Then</th>
<th>Now</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small projects (&lt;10,000 pp)</td>
<td>Large projects (&gt;100,000 pp)</td>
</tr>
<tr>
<td>Relatively ad-hoc</td>
<td>Major strategic programme</td>
</tr>
<tr>
<td>SMT &amp; Project teams</td>
<td>Programme Board, advisors</td>
</tr>
<tr>
<td>Library-centric</td>
<td>W. Trust, external stakeholders</td>
</tr>
<tr>
<td>Entirely open access</td>
<td>Commercial partnerships</td>
</tr>
<tr>
<td>Little impact on IT systems</td>
<td>Requires major IT development</td>
</tr>
<tr>
<td><strong>Examples</strong></td>
<td><strong>Everything (within reason)</strong></td>
</tr>
</tbody>
</table>
## Digitisation processes

<table>
<thead>
<tr>
<th>Then</th>
<th>Now</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual processes</td>
<td>Automated processes</td>
</tr>
<tr>
<td>Centralised conservation</td>
<td>Distributed conservation</td>
</tr>
<tr>
<td>Low QA</td>
<td>Increased QA, error minimization</td>
</tr>
<tr>
<td>TIFF</td>
<td>JPEG 2000</td>
</tr>
<tr>
<td>Individual tracking lists</td>
<td>Centralised tracking system</td>
</tr>
<tr>
<td>Incremental storage growth</td>
<td>Completely new storage strategy</td>
</tr>
<tr>
<td><strong>Detailed, painstaking</strong></td>
<td><strong>Streamlined, pragmatic</strong></td>
</tr>
</tbody>
</table>
Streamlining digitisation

- Staff dedicated to specific projects, or streams of work
- Carry out sample workflow tests for new types of material
- The right equipment for the right job – eliminate the “fiddly bits”
  - Live-view monitors
  - Easy-clean surfaces
  - Foot-pedals
  - Custom-made supports
Streamlining digitisation

• Photographers do the photography…
• Prepare materials separately
• Leave loose pages and bindings as they are, they are easier to digitise that way!
• Use existing staff as support – moving items to and from stack
• Minimise movement
• Keep plenty of shelving, working space at hand
• Find a preferred supplier for ad hoc support
Upscaling and streamlining digitisation requires a higher level of project management
Streamlining project management
What is it?

- Web-based workflow system
- Open source (core system)
- Used by many libraries in Germany, and half a dozen other European libraries
- Intranda version developed by Intranda to meet Wellcome Library specific requirements
What does it do?

- Task-focused, customisable workflows developed by Intranda
- User-specific “dashboard”
- Import/export and store metadata
- Encode data as METS
- Display progress of tasks, statistics on activities
- Tracks projects, batches, and units (location, current activity)
- “Command central” for 3rd party systems
# User tasks

**My tasks**

Please select one of the listed tasks.

<table>
<thead>
<tr>
<th>Task</th>
<th>Process</th>
<th>Project</th>
<th>Status</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is not a test, this is rock and/or roll</td>
<td><img src="a.ocm00665772i-dnt" alt="Image" /></td>
<td>Sample_Project</td>
<td></td>
<td><img src="a.ocm00665772i-dnt" alt="Image" /></td>
</tr>
<tr>
<td><img src="a.ocm00665772ix" alt="Image" /></td>
<td></td>
<td>Sample_Project</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="_ThisisathPPNdigitalisatz" alt="Image" /></td>
<td></td>
<td>Sample_Project</td>
<td></td>
<td><img src="_ThisisathPPNdigitalisatz" alt="Image" /></td>
</tr>
</tbody>
</table>

**Possible actions**

- Customise view
This book is very fragile, requires post-digitisation preservation.
Project management tasks

### Possible actions
- Set status of process up
- Set status of process down
- Calculate number of metadata and images

### Statistical evaluation
- Volume status
- Responsible user groups
- Duration of the steps
- Production statistics
- Production throughput
- Storage calculator
- Error tracking
- Project association

### Generate excel

### Statistical evaluation: Error tracking
Define statistical question

<table>
<thead>
<tr>
<th>Period of</th>
<th>bis</th>
<th>unit</th>
<th>View</th>
<th>Volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/9/11</td>
<td></td>
<td>months</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Calculate statistics
Administrative tasks
Digital asset management

| File conversion       | • Lightroom used to convert RAW to TIFF  
|                      | • LuraWave converts TIFF to JP2K  
|                      | • Validation of JP2K conversion coming soon – *via Goobi*  |
| Ingest               | • Automated ingest workflow in the DAM (Safety Deposit Box - SDB) – *via Goobi*  
|                      | • One file serves as master and dissemination file  |
| Preservation          | • DAM is a preservation system  
|                      | • Manages all preservation actions (characterisation, format migration)  
|                      | • API to allow 3rd party systems access to content  |
| Storage              | • Master files backed up offsite to WORM storage drive  
|                      | • WORM = Write Once Read Many – permanent storage  
|                      | • Self-healing of errors on main storage system from WORM  |
In-house (RAW) → Temp → QA → Lightroom - post-processing, convert to TIFF → Hotfolder

External (TIFF) → Temp → QA → Hotfolder

External (JP2) → Temp → QA → Hotfolder

Hotfolder

LuraWave automatically converts files to JP2 and outputs to a folder

Goobi automatically triggers validation

Person triggers ingest via Goobi

SDB ingests

WORM backup
Really permanent

Pillar permanent
Thank you!

Christy Henshaw
c.henshaw@wellcome.ac.uk