

The Oxford Google Digitization Project



OXFORD
UNIVERSITY
LIBRARY
SERVICES

Frances Boyle



Overview

- Why - reasons
- What - content
- When - timelines
- How - outline process/activities
- Then what....



Why - 1

- This project is primarily about access not conservation
- An opportunity to expose content to a far wider community than is possible at the moment or in the near future
- Builds on the work of the Oxford Digital Library (ODL) and OULS strategic initiatives e.g. ELISO project
- Portfolio of the ‘born digital’ growing so the project affords us the opportunity to build up collection of digitized copies of locally held material



Why - 2

- The synergy between missions:
 - Bodley’s ‘Republic of Letters’
 - Google’s ‘to organize the world’s information and make it universally accessible and useful’
- Bring more Oxford content into the digital landscape making it available for scholarly and public benefit
- Also facilitating access to the material for University members



What - 1

- Direct discussions with Google since 2003
- The extensive collection of out of print (and out of copyright) material identified
- Oxford differs from other partners in this aspect of our agreement
- Decision made to begin with the 19th century material
- Looking at approximately 1 million items
- Win/win situation for both parties



What - 2

- The scanning operation to take place on OULS premises, at Osney Mead
- Virtually all costs met by Google
- Oxford costs: local staff time & expertise
- The deliverable being two digital copies
- English Law if there are issues relating to contract
- Contracts differ for each of the partner institutions



When

- Timelines:
 - planning work began at the beginning of 2005
 - looking towards the Autumn to commence scanning operations
 - initially a 3 year contract
- There will be a ramping up phase to allow testing of the workflows/output



How - 1

- Selection of material is done by OULS
- Scanning techniques are non-invasive
- Appraisal stage is part of the workflow
- Any book that would be released to reading rooms likely to be in a fit state to be scanned
- ‘Slow track’ process for fragile material



How - 2

- OULS staff will work closely with Google staff e.g. training on how to handle the material
- Each component of the workflow must be comfortable for both parties
- A large complex logistic operation that will not compromise the service to our users



How - 3

- Current activities:
 - Facilities – refurbishment/fitting out of space
 - Business perspective:
 - Devising & testing workflows for each of the components
 - Collection profiling work
 - Completed a paper exercise of an envisaged workflow
 - This workflow put to the test over a 3 day trial in May
 - Involved activity & cooperation from many different OULS teams



How - 4

- Evaluating the outcomes of the trial - proved a very useful exercise
- Compiling a matrix of libraries/collections to inform the business decisions e.g. distribution of material, access issues etc
- Steering Group overseeing the project with various working parties looking at the specific issues e.g. facilities, legal etc
- Oxford specific challenges – uncut material, access, disparate collections



How - 5

- Large raw images from scan stations
- Derivatives are delivered in standardised formats
- What is returned per volume:
 - JPEG2000 (probably)
 - OCR
- Audit of production process
- There are quality control processes at Google & Oxford
- Linked to OLIS catalogue record
- Ongoing software/hardware developments to improve the process



Then what...

- Future challenges include building the infrastructure of the ODL
- Develop access routes to the content to customise the Oxford experience
- Cognisant of the changes in the information landscape



Useful links

- <http://www.bodley.ox.ac.uk/google/>
- <http://print.google.com/googleprint/library.html>