The Google Library Project at Oxford

Ronald Milne

Fiesole Retreat
Lund 4 August 2006
Thomas Bodley’s Vision

• Bodleian founded 1602
• Universal library
• Bodley’s “Republic of Letters”
• Legal deposit privilege since 1610
• 60% of Bodleian readers not members of Oxford University
QUOD FELICITER VORTAT
ACADEMICI OXONIENS
BIBLIOTHECAM HANC
VOBIS REIPVBLCÆQVE
LITERATORVM
T. B. P.
Bodleian Library

- 400 staff
- Budget of £14m (€20m)
- Stock 8 million items
- 45,000 registered users
- 120 Miles (192km) of shelving
- 123,000 monograph items and 194,000 serial items added each year
Oxford University Library Services

- 700 staff (600 fte)
- 40 libraries, including the Bodleian, Radcliffe Science Library, Taylor Institution Library (modern languages) Sackler (classics and archaeology) and Social Science Library
- Budget: £30m (€43m)
- Total bookstock: 11 million items
- 160 linear miles of shelving
Some Oxford digitisation projects

• Toyota City Imaging Project (1993)

• Specialised Research Collections in the Humanities (NFF) and eLib projects (1995 -1998)
  – Early manuscripts in Oxford
  – Broadside Ballads
  – John Johnson Collection

• Oxford Digital Library (Mellon Foundation, 2000 - 2005)
  – Scoping study
  – Three production phases

• Digital Shikshapatri (NOF, 2001 – 2003)
Early Manuscripts at Oxford University

Digital facsimiles of complete manuscripts, scanned directly from the originals

This site provides access to over 80 early manuscripts now in institutions associated with the University of Oxford. Please read the information about using the website.

Between 1995 and 2000 the Early Manuscripts Imaging Project created high resolution digital images from manuscripts which were selected as major treasures from their respective libraries, to create wider availability for originals which may otherwise be too fragile for handling.

In 2001 interim responsibility was taken up by the Libraries Automation Service’s R&D section who have undertaken a complete redesign of the site to improve the service. The site is now in the care of the Oxford Digital Library, to whom enquiries should be addressed.

Please note: We are aware that a number of manuscripts are incomplete. Unfortunately, the omission of some images only came to light after the project had completed (during the redesign of the website). We are currently in the process of identifying the resources needed to complete the scanning of these manuscripts. The incomplete manuscripts are still accessible from this site, on the principle that a nearly complete manuscript is better than none at all.

We realize that this situation can be frustrating but currently have no further information on when we can make the missing images available.

list manuscripts by collection:
- Halliwell College
- Bodleian Library
- Corpus Christi College
- Jesus College
- Magdalen College
- Merton College
- St. John’s College

list all manuscripts
Bodleian Library Broadside Ballads

The allegro Catalogue of Ballads

The Bodleian Library has unparalleled holdings of over 30,000 ballads in several major collections. The original printed materials range from the 16th- to the 20th-Century. The Broadside Ballads project makes the digitised copies of the sheets and ballads available to the research community.

Bodleian Library, University of Oxford

© Bodleian Library 1999 - Last updated 20 Jan 2005 / TSL
A New SONG

Ho Brother Teague doft hear de Decree, Lil--li Burleró Bullen a--la, Dat
The John Johnson Collection of Printed Ephemera

The John Johnson Collection is one of the most important collections of printed ephemera in the world. It was assembled by John de Morgan Johnson between c. 1923 and 1956 and was housed at the Oxford University Press (where it was called The Constance Meade Collection of Ephemeral Printing) until its transfer to the Bodleian Library in 1968. Johnson collected retrospectively, establishing 1939 as his terminus ante quem (although there are exceptions). While the majority of material dates from the 18th, 19th and early 20th centuries, there is ephemera dating back to 1508. There are over 1 million items in the original collection. The material is principally British.

Further reading

The principal work on John Johnson and his Collection is John Johnson Collection, catalogue of an exhibition, Oxford (Bodleian Library), 1971. Both this, and the catalogue of the 2001 John Johnson exhibition, A Nation of Shopkeepers, Trade Ephemera from 1654 to the 1950s (£17.50, less discounts, + p&p) can be purchased from the Bodleian shop, email sales@bodley.ox.ac.uk
Welcome to the Bodleian Library / Toyota City Imaging Project

A collection of motoring and transport images from the John Johnson Collection of Printed Ephemera

Browse the Database
- Browse the Whole Database
- Browse by Type of Transport
- Random Browse

About the Project...
- How to Use the Database
- About the Project
- The John Johnson Collection website

Copyright Bodleian Library 1996
We acknowledge with thanks the support of Powersmile in making this project possible.
ROAD RISKS ILLUSTRATED.
MOTOR PROSPECTUS - 1926.
For only British but North British.

NORTH BRITISH

CLINCHER MOTOR TYRES

1909

SOLE MANUFACTURERS

The NORTH BRITISH RUBBER CO. LTD.

Castle Mills, EDINBURGH

PRICES QUOTED IN THIS LIST ARE SUBJECT TO 15% ADVANCE.
Digital Shikshapatri

Digital Shikshapatri provides instant online access to a treasure of the British Hindu cultural heritage that is held in Oxford’s Bodleian Library. This fragile Sanskrit manuscript, called the Shikshapati, was written by Shree Swaminarayan, founder of Swaminarayan Hinduism, and outlines moral and spiritual codes for everyday life.

This site provides a wide variety of online resources which set the Shikshapati in its historical, cultural and religious context. It is designed to help people of all backgrounds to better understand the significance of this manuscript, which receives hundreds of visitors every year.

View the manuscript

Bodleian Library

New Opportunities Fund
LOTTERY FUNDED

Centre for Hindu Studies

Refugee Studies Centre
I meditate, in my heart, upon Lord Shri Krishna, on whose left resides Radha and in whose heart resides Lakshmi, and who performs Lila with His Bhakats in Vrajawan.
Google Library Project objectives in outline

• To digitise materials from five major research libraries: Harvard, Michigan, New York Public, Oxford, Stanford

• To create OCR’d text, with indexes for search and retrieval via Google search services and, in particular, Google Book Search

• To provide online searching and access to hitherto inaccessible printed materials for the public worldwide
From the Google perspective

• Google’s mission:
  – “to organize the world’s information and make it universally accessible and useful”

• Library digitisation programme formulated in early 2003 as a means of tapping into ‘Deep Web’ materials

• Conversations with publishers and major libraries during 2003-2004

• Library programme under the umbrella of Google Book Search

• Consistent with constant enhancement programme for its search engine services
From the Oxford perspective

• Oxford’s Mission: ‘…to maintain and develop access to Oxford’s collections as a national and international research resource’

• Would be part of hybrid library project to create as much local electronic content as possible: ‘The Oxford Digital Library’

• Direct discussions with Google 2002 onwards

• Win/Win situation quickly identified
Oxford Google digitisation

• Agreement signed December 2004
• Digitisation of 19th-century out-of-copyright material: 1 - 1.5M items
• Includes material from the Bodleian Library and other major Oxford University libraries
• A multi-year project
Two frequently asked questions about the Oxford project

- **Why 19th century materials?**
  - Out of copyright and out of print
  - Oxford already a principal partner in Early English Books Online (EEBO) and Eighteenth Century Collections (ECCO) projects

- **When did the project start?**
  - Google Manager on site from 1 August 2005
  - Scan stations were delivered in mid-December; work began March 2006
Some terms and conditions

- Digitisation takes place on University premises
- Virtually all costs met by Google
- Oxford costs: seconding local project manager and selection of materials
- Two digital copies to be made: Google and Oxford free to exploit non-exclusively
- English Law if there are issues relating to contract
A TREATISE
ON THE
PHYSIOLOGY AND DISEASES
OF
THE EYE:

CONTAINING
A NEW MODE OF CURING CATARACT WITHOUT AN OPERATION;
EXPERIMENTS AND OBSERVATIONS ON VISION, ALSO ON THE
REFLECTION, REFLECTION, AND COLOURS OF LIGHT;

TOGETHER WITH
REMARKS ON THE PRESERVATION OF SIGHT, AND ON SPECTACLES,
READING-GLASSES, &c.

By JOHN HARRISON CURTIS, Esq.

LONDON:
BENJAMIN HARRISON CURTIS, PUBLISHER.

1796.
ASKJA.

ICELAND'S LARGEST VOLCANO:

WITH A DESCRIPTION OF THE

GREAT LAVA DESERT IN THE INTERIOR:

AND

CHAPTER ON THE GENESIS OF THE ISLAND.

BY

W. G. LOCK, F.R.G.S.,

TRANSLATOR OF "VOLCANS ET VOLCANOLOGIE," AND

AUTHOR OF "THE Eruption of Tungurahua in 1884-85."

Published at the Author of R. Rongga Trottin,

Copenhagen, 1891.

1897, c. 119.

© Copyright Bodleian Library, 2005
A JOURNAL OF THE DISASTERS IN AFFGHANISTAN, 1841-2, BY LADY SALE.

LONDON: JOHN MURRAY, ALBEMARLE STREET. 1842.
HISTORY
OF
THE WAR IN AFGHANISTAN.
FROM
THE UNPUBLISHED LETTERS AND JOURNALS
OF POLITICAL AND MILITARY OFFICERS EMPLOYED IN AFGHANISTAN
THROUGHOUT THE ENTIRE PERIOD OF BRITISH OCCUPATION
WITH THAT COUNTRY.

BY JOHN WILLIAM KAYE.

IN TWO VOLUMES
VOL. I.

LONDON:
RICHARD BENTLEY, NEW BURLINGTON STREET.
Publishers in Ordinary to Her Majesty.
1851.
A DICTIONARY OF THE
TELOOGOO LANGUAGE,
COMMONLY TENDED THE GENTOO,
Pertaining to the Inhabitants of the North Eastern Provinces
OF THE
INDIAN PENINSULA.

By A. D. CAMPBELL,
OF THE
MARATHA CIVIL SERVICE,
LATE MEMBER, AND SECRETARY,
OF THE
BOARD OF SUPERINTENDENCE
FOR THE
COLLEGE OF FORT ST. GEORGE.

JOHN NICHOLS
PRINTED AT THE COLLEGE PRESS,
1811.

306. 30. 42.
SIX LECTURES
ON
HARMONY.

DELIVERED AT
THE ROYAL INSTITUTION OF GREAT BRITAIN,
Before Easter, 1827.

BY
G. A. MACPHERSON.

LONDON
LONGMANS, GREEN, READER, AND DYER.
1827.
THE HISTORY OF FASHION IN FRANCE;

OR,

THE DRESS OF WOMEN FROM THE GALLO-ROMAN PERIOD TO THE PRESENT TIME.

FROM THE FRENCH OF

M. AUGUSTIN CHALLAMEL

BY MRS. CASHEL HOXY AND MR. JOHN LILLIE.

Printed by RAMPSON LOW, MARSTON, SKEALE & RIVINGTON.

CROWN BUILDINGS, 65, Fleet Street.

1838.

[Stamp: BODLEIAN LIBRARIES OXFORD]
THE ART OF MONEY-GETTING
HINTS AND HELPS TO MAKE A FORTUNE
BY P. T. BARNUM.
WARD, LOCK, & CO., LONDON & NEW YORK
© Copyright Bodleian Library, 2005
CROCKFORD'S

Clerical Directory

1860:

BEING A BIOGRAPHICAL AND STATISTICAL BOOK OF
REFERENCE

WITH

FACTS RELATING TO THE CLERGY AND THE CHURCH.

PUBLISHED ANNUALLY.

Kings:
CROCKFORD'S CLERICAL JOURNAL AND DIRECTORY OFFICE, No. 36, WELLINGTON STREET NORTH, STRAND, W.C.
THE OFFICIAL ILLUSTRATED GUIDE TO THE GREAT WESTERN RAILWAY, INCLUDING THE OXFORD, WORCESTER AND WOLVERHAMPTON, AND CHESTER AND HOLBEAD LINES, AND ISLE OF MAN, WITH DESCRIPTIONS OF THE MOST IMPORTANT MANUFACTURES IN THE SEVERAL TOWNS ON THE LINE, BY GEORGE MEASON. EMBELLISHED WITH 800 ENGRAVINGS.

LONDON: Published under the Authority of the Directors, by RICHARD GRIFFITH AND CO., STATIONERS' HALL COURT, NAPIER & WEIGHT, NEW STREET, BIRMINGHAM. ONE SHILLING.
Some issues

• Duplication among libraries, and language representation
• Conservation
• Logistics
• Workflow
• Delivery of the ‘Oxford Digital Copy’
Duplication among libraries, and language: collections analysis

- Analysis of holdings on WorldCat

- Collection overlap among the ‘Google Five’: 56% of works are held uniquely by one ‘Google Five’ library

- When comparing only two libraries out of the five, eight out of ten books are held uniquely

- On average, about 50% of ‘Google Five’ libraries’ holdings are in English

- Over 430 different languages represented

  www.dlib.org/dlib/september05/lavoie/09lavoie.html
Conservation

• Conservation: Oxford has the absolute right to say that particular items may not be digitised

• 19th-century material – acidity of paper; but Google scanning non-invasive; ‘slow track’ for fragile material

• Any book that can be read likely to be in a fit state to be digitised

• Newspapers and large format material excluded at this stage
Logistics

- Books transferred by van from the New Bodleian stack to the digitisation centre
- Space requirement – depends on number of scanning stations
- Currently working 2 shifts; 24 x 7 working – a possibility: need for a ‘buffer zone’
- ‘Selection’ of books: – digitisation *en masse* rather than ‘cherry-picking’
- Google currently digitising material Roman character set material
- Expect to be able to undertake digitisation of material in Arabic, Cyrillic and CJK character sets during 2006
Workflow

• Operating on an industrial scale

• Workflow tests/trial runs in May and December 2005

• Identification of stock, pulling the stock, appraisal (conservation checking etc), barcoding of stock, associated bibliographic work, movement to the scanning site and return

• Trial allowed us to test methodologies and configure the project profile

• The real challenge is the associated bibliographic/barcoding work (Unlike other libraries, Oxford has many unbarcoded items)

• Survey revealed that circa 20% of 19th century stock is uncut …

• Survey also revealed that circa 3% of 19th century stock is uncatalogued …
The miscellaneous works of Oliver Goldsmith [ed. by S. Rose]. (1812)
The works of Samuel Johnson [vol 5 of 12] (1823)
Truckleborough Hall [by W.P. Scargill]. (1827)
Lays of the minnesingers or German troubadours of the twelfth and thirteenth centuries
[ed. by E. Taylor.] (1825)
Le roman du renart, publ. d'après les MSS de la Bibliothèque du roi des xiiiie, xive et xve siècles par D.M. Méon (1826)
Q. Horatii Flacci opera, cum scholiis veteribus castigavit, et notis illustr. G. Baxterus. (1806)
Krakas Maal; eller, Kvad om Kong Ragnar Lodbroks Krigsbedrifter og Heltedod, med Dansk, Lat. og... by Ragnar (1826)
Tales of a voyager to the Arctic ocean, by Robert Pierce Gillies (1826)
Voyage historique et littéraire en Angleterre et en Écosse by Joseph Jean M.C. Amédée Pichot (1825)
Practical remarks on modern paper by John Murray (1829)
Reliquiae Conservatae: from the primitive materials of our present globe (1827)
Reliquiae Conservatae – plate illustration
The woodlands: or, A treatise on the preparing of ground for planting; on the planting [&c.] of... by William Cobbett (1825)
Code de commerce espagnol, décrété ... le 30 mai 1829. (1830)
Observations on the importance in purchases of land and in mercantile adventures of ascertaining... (1826)
Description des objets d'arts qui composent le cabinet de ... m. le baron V. Denon (1826)
Oxford Digital Copy

- Oxford books already available on the Web through Google Book Search
- OCR’d text
- Will be linked to Oxford catalogue record
- Oxford and other libraries’ material will be hosted on a Google server, in the first instance
- Available to anyone, anywhere, on the Web
- Google Library Project represents a step-change in the dissemination of information
- Potentially a transforming agent in learning, teaching and research
Collection development matters

• Much out-of-print, out-of-copyright material will become available on the Web

• All libraries should be able to point to Google Library Project public domain material

• Create portals for particular genres of material, or different subject categories, create ‘boutique’ sites?

• Content rather than collection strategy (cf the British Library)?

• With so much electronic material available, both in the public domain and by subscription, what defines and distinguishes any one particular library?
QVOD FELICITER VORTAT
ACADEMICI OXONIENS
BIBLIOTHECAM HANC
VOBIS REIPUBLICAEOVE
LITERATORVM
T. B. P.
Ronald Milne
Acting Director of University Library Services
& Bodley’s Librarian

ronald.milne@ouls.ox.ac.uk

+44 (0) 1865 287107