Approaches to Putting Library Collections Online

Neil Jefferies

R&D Project Manager, Bodleain Libraries

Neil.Jefferies@bodleian.ox.ac.uk

Bodleian Libraries
UNIVERSITY OF OXFORD

UNIVERSITY OF OXFORD

Introduction to BDLSS

- The Bodleian Library
  - ~1320 First University library
  - 1488 Duke Humfrey's Library
  - 1602 Bodleian Library
    - Not **THE**University Library
  - 1610 Origin of legal deposit
    - Agreement with Stationers' Company

- ~30 "Bodleian" libraries
  - 550-600 staff
  - £36M budget
  - 13 million items in the electroinic catalogue
    - Much still in paper catalogues
  - 156+ shelf miles (250 km)

- ~30 other libraries

- 46 college libraries

- And a few others!

- Bodleian Digital libraries Systems and Services (BDLSS) provides electronic services for **all**
  - ILS
  - Resource Discovery and Access
  - Digital Library
  - Data Repository (soon!)
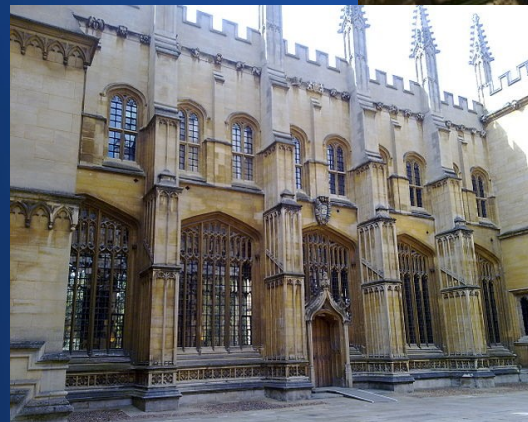
Evolving Environment

- Historically funding was often available for digitisation for whole collections
  - Metadata drawn from existing catalogues (if at all!)
    - Description rather than transcription/transliteration
    - Bias towards "easy" collections – already catalogued, in good shape
  - Collection-specific dissemination
- Ad hoc "commercial" digitisation service for individual items
  - Not strategic, many individual items and fragments, minimal metadata
- Emergence of born-digital content and electronic publishing
  - Special Collections become more important
  - Linking and metadata essential **RDF**
- Increasing volumes of material requires scalable systems
  - Standardised delivery and metadata
  - Manage heterogeneity – more content types, standards
- Increasing funding focus on research questions rather than pure content
  - Uncatalogued/unusual materials
  - Researcher rather than library driven standards and presentation
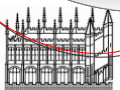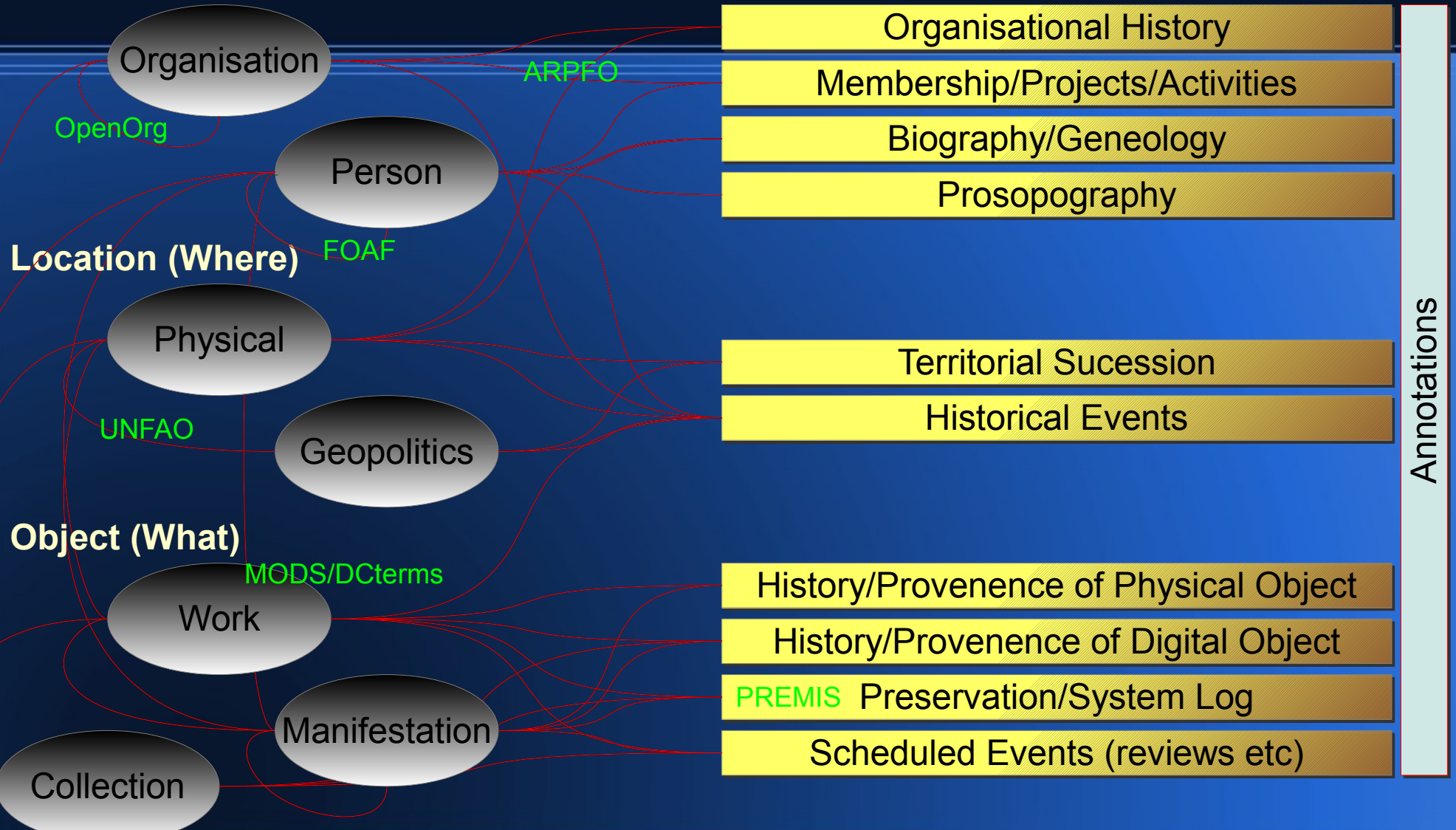 - Interactivity

Flexibility

- Legacy digitised materials must be migrated to new systems
  - We do not have the resources to create much new metadata (use automation)

- The amount of digitisable content vastly exceeds our resources and capacity
  - Do we need "archival" quality images of everything? Are we a museum or a library?
  - Retool digitsation workflow for volume
  - Research driven funding changes selection criteria
    - We still have a few big donors!
  - Significant amount of material still not electronically catalogued

- Hybrid collections
  - Integrated view of physical, digitised and born digital materials

- Common platform for online collections
  - Balance standardisation with collection-specific realities
    - Amounts/quality of metadata – Minimal/DC...MODS/MARC...EAD/TEI/Provenance
    - Amount and type of content – catalogues, text, images, zoomability, rich data
      - Online critical/annotated editions, comparative texts, analytics and visualisations
    - Differing academic requirements
      - Annotation/commentary
      - Transcription/transliteration
      - Unusual languages and scripts

Abstract Data Model

Event Streams (When)

Agents (Who)

Organisation

ARPFO

OpenOrg

Person

FOAF

Organisational History

Membership/Projects/Activities

Biography/Geneology

Prosopography

Location (Where)

Physical

UNFAO

Geopolitics

Territorial Sucession

Historical Events

Object (What)

MODS/DCterms

Work

History/Provenence of Physical Object

History/Provenence of Digital Object

PREMIS Preservation/System Log

Manifestation

Scheduled Events (reviews etc)

Collection

Annotations

Examples
http://www.bodleian.ox.ac.uk/bdlss/projects

Old Catalogues Online

- Fihrist – National Union Catalogue of Islamic Manuscripts
  - http://www.fihrist.org.uk
  - Started as Oxford and Cambridge Islamic Manuscripts Online (JISC funded)
  - Records created using TEI subset (with extensive style guide)
  - Converted to RDF on upload
  - Code subsequently re-used for other mixed RTL language catalogues (Shan, Tibetan etc.)

- Maps and Music Catalogues (Mellon funded)
  - Approx 500,000 cards and paper slips digitsed and keyed
  - Collection-specific metadata encoded in XML
  - Minimal conversion to MARC for ILS (circulation purposes)

- These make previously undiscoverable material visible
  - Starting point for further digitisation projects which can add content incrementally
    - e.g. Genizah (http://genizah.bodleian.ox.ac.uk/) is based on the Fihrist code
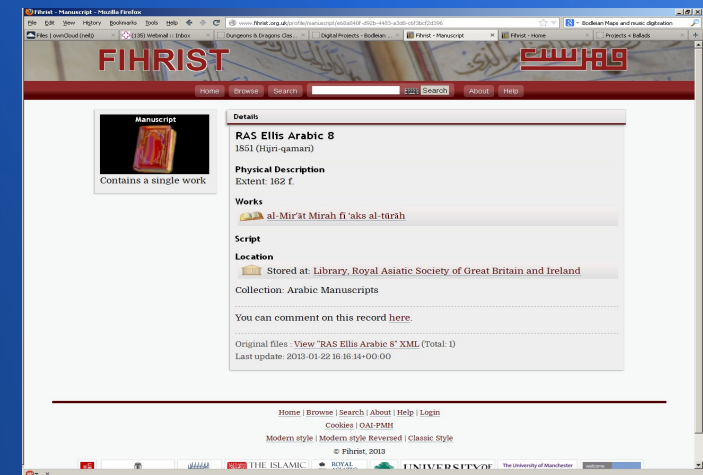
Enhanced Catalogues

- Catalogues to which additional enriching material has been added
    - Information about people, places, context, detail descriptions, abstracts or texts/transliterations
    - This model is very popular with researchers 9especially if they can contribute!)

- Integrated Broadside Ballads Archive (JISC funded)
    - http://ballads.bodleian.ox.ac.uk/projects
    - 1999 online catalogue migrated to RDF semantic model
    - Links with EBBA and Roud Index
    - Experimental image search with some new colour images

- Early Modern letters Online (Mellon funded)
    - http://emlo.bodleian.oc.ac.uk
    - Digitised card catalogue of C17 letters
    - Enhanced with transcriptions, some images, annotations, biographies and geography
    - Data imported from other sources including new catalogues and scholarship

- IMPAcT (ERC funded)
    - http://impact.orient.ox.ac.uk/
    - Re-use EMLO code for C14-15 Islamic Codices
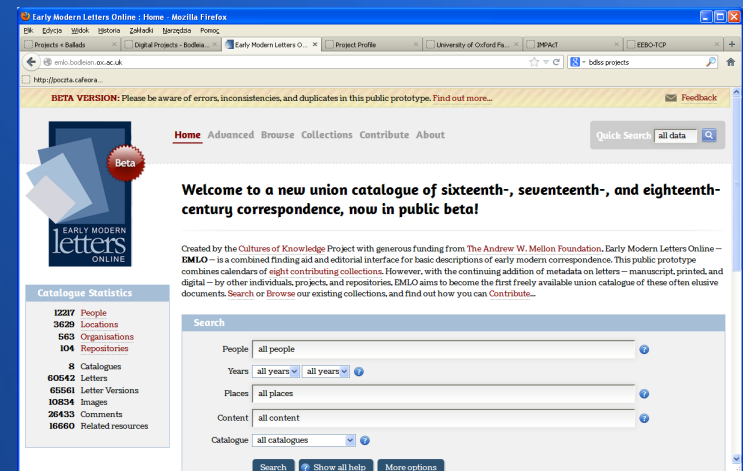
Enhanced Texts

- EEBO-TCP (JISC Funded)
  - http://www.bodleian.ox.ac.uk/eebotcp/
 - Early English Books Online (Parnership with Michigan University)
  - Marked up TEI-lite texts based on EEBO Images

- MLGB3 (Mellon Funded)
  - http://www.history.ox.ac.uk/research/project/medieval-libraries-of-great-britain-mlgb3.html
  - Online version of original book with updated content, links to resources
  - Catalogue of extant MSS with current and original location
  - Incorporates information about evidence and provenance.

- Poetic Miscellanies
  - Selected EEBO texts presented camparatively with extensive markup and annotation

- London Book Trades
  - Semantic MediaWiki resourcing mixing structured and unstructured information
  - Includes personal and geneological information

- Stationers' Register Online
  - Register of the Worshipful Company of Stationers and Newspaper Makers (1577 – 1708)

Digitisation

- Google Books
  - http://www.bodleian.ox.ac.uk/dbooks
  - Mass digitisation of approx. 350,000 books with OCR text and coordinates
  - Integrated with Library catalogue as holdings

- Polonsky Foundation Digitisation Project
  - http://blogs.bodleian.ox.ac.uk/pfdp/
  - Collaboration with Biblioteca Apostolica Vaticana
  - Target approx. 1.5M MSS

- What's the Score at the Bodleian (Google funded)
  - http://www.bodleian.ox.ac.uk/bodley/library/special/projects/whats-the-score
  - Digitisation of uncatalogued collection of Music MSS
  - Experiment with crowdsourcing metadata with public and expert clouds

- Sprint for Shakspeare
  - http://shakespeare.bodleian.ox.ac.uk/
  - Crowd-funding for digitisation of First Folio

- Digital.Bodleian
  - Migrate legacy digitisation projects to a standard platform (JPEG2000, page turner)
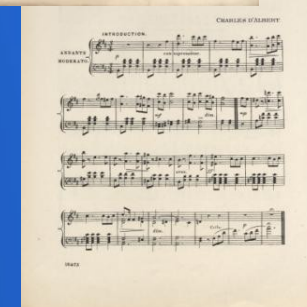
Summary

- Online collections are an increasingly diverse set of resources
  - Even considering just those based on traditional library materials
  - Don't be too strict about standards...the key is to make material available
    - Some standards are obviously required but keep this set small

- Interactvity and evolution are key to sustainability
  - New content and scholarship keeps a resource relevant
  - Getting content online generates research income
  - Metadata sharing and linking are essential
    - This also boosts your search ranking

- Resources should be designed to evolve
  - A card catalogue is a very focussed finding aid
  - Adding images and abstracts is **more** useful\ than enhancing metadata

- Legacy custom database driven Websites remain a problem

Questions?

Neil Jefferies
R&D Project Manager, Bodleain Libraries
Neil.Jefferies@bodleian.ox.ac.uk