



Top 5 technologies for budding digital archivists: Here comes the future

Lucie Burgess

Associate Director for Digital Libraries

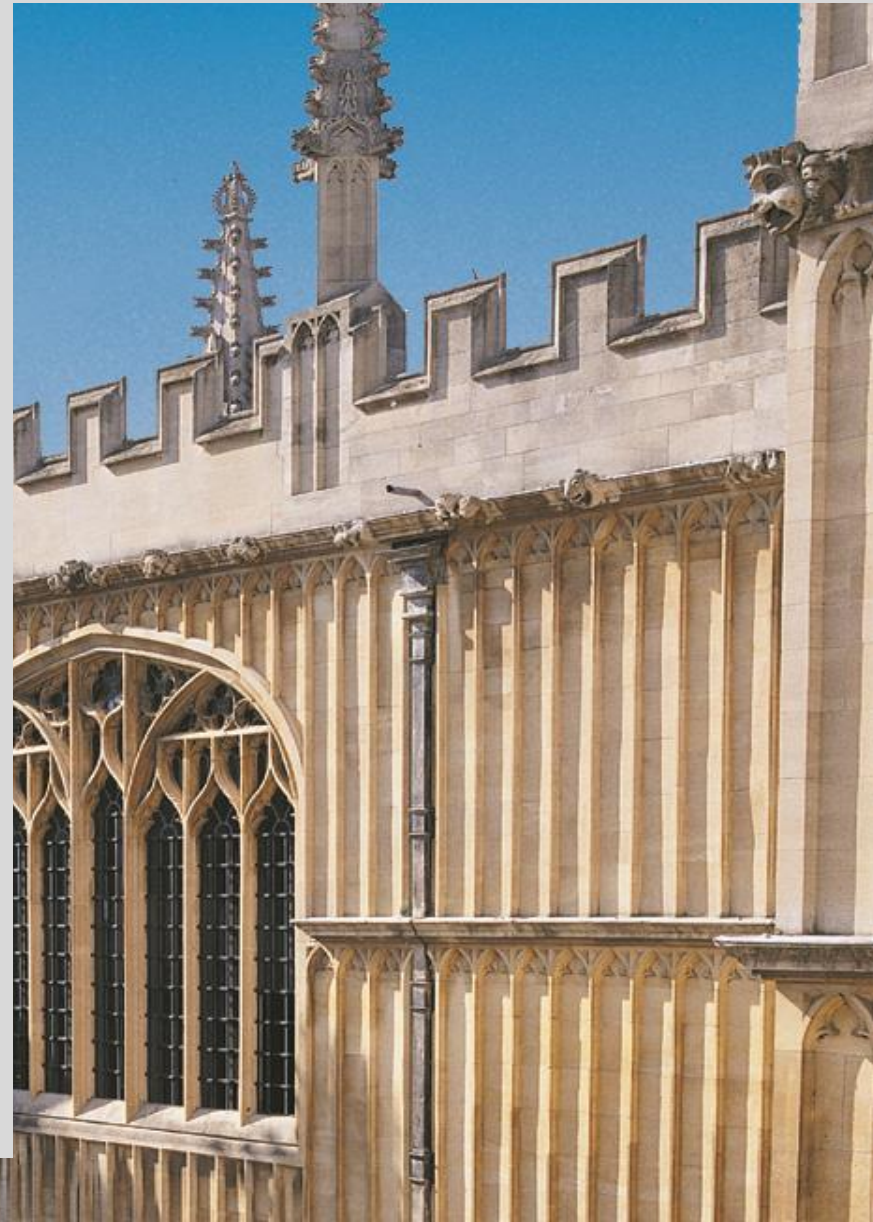
Bodleian Libraries, University of Oxford



The Bodleian Libraries

**Bodleian Library founded in 1598,
opened 1602**

- ▶ Largest academic library system in Europe, and one of the largest in the world
- ▶ 70,000 current registered users (University or Library card)
- ▶ Over 11 million printed items (increasing by 350,000 items pa)
- ▶ 560 FTE staff
- ▶ Consistently ranked highest in the UK in the national student satisfaction survey
- ▶ 100 libraries, roughly!



Integral part of Oxford University

Oldest University in English-speaking world with 800 year old tradition of discovery & invention

Supporting excellence in research, teaching and learning, public engagement (Oxford came top in 2014 REF based on quality)

Multi disciplinary: Humanities, Social Sciences, Mathematical, Physical & Life Sciences, Medical Sciences, Continuing Education

First University Library founded 1320 in Church of St Mary the Virgin



The Bodleian Digital Library

Major projects including:

- ▶ Digitisation – Google Books, Vatican, Electronic Enlightenment
- ▶ Open Access archive – ORA + 2020 REF
- ▶ Data archiving – ORA-Data pilot
- ▶ Text editing services – EEBO-TCP
- ▶ Cultures of Knowledge
- ▶ digital.bodleian (images)
- ▶ Electronic Archives project
- ▶ Electronic legal deposit



The Weston Library

- ▶ Refurbishment of 1930s 'New Bodleian Library' building holding special collections
- ▶ To include Centre for Digital Scholarship





Technology no. 1: IIIF

- ▶ IIIF = International Image Interoperability Framework
- ▶ Used to enable sharing of very high resolution digitised images without requiring interoperable infrastructures
- ▶ Image API
- ▶ Presentation (metadata) API
- ▶ Being driven by major cultural heritage institutions including the Bodleian/Oxford, Stanford, British Library, Princeton, BNF
- ▶ IIIF demo: <http://openseadragon.github.io>



IIIF transforms digital image delivery from ugly into beautiful

Slide: Tom Cramer, Stanford



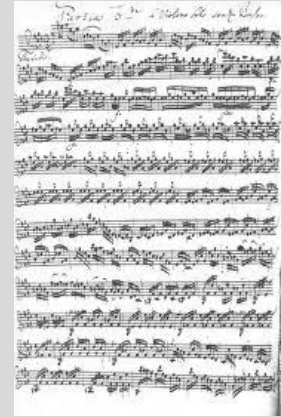
Books



Manuscripts



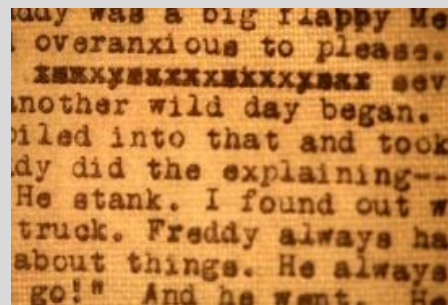
Newspapers



(Sheet) Music



Artworks



Archival Materials



Maps



STEM Imagery



Scrolls



Architecture



Image + Presentation = Object

A screenshot of a digital manuscript viewer interface. The interface includes a top navigation bar with 'Bookmark', 'Options', and 'Switch Workspace' buttons. Below this is a breadcrumb trail: 'x || MS Richardson 7. Houghton Library, Harvard University. Heures d...'. The main content area is divided into a left sidebar and a central image viewer. The sidebar, outlined in red, contains a tree view of the manuscript's structure, with a red box highlighting the 'Presentation API' section. The central image viewer, outlined in blue, displays a manuscript page with a large illuminated initial 'D' and text in Gothic script. A blue box is overlaid on the image, listing 'Image API' features. At the bottom, a red-outlined strip shows a sequence of thumbnail images from the manuscript, labeled with folio numbers.

Bookmark Options Switch Workspace

x || MS Richardson 7. Houghton Library, Harvard University. Heures d...

Hours of the Cross (ff. 21-35)

- Matins (ff. 21-23v)
- Hours of the Cross (ff. 21-35)
- Hours of the Holy Spirit (ff. 37-46v)

Presentation API

- Title Label
- Structure
- Sequence
- Licence
- Description
- Labels/ annotations

Image API

- Image Data
- Pixels
- Location on canvas

f. 24v. f. 25: Flag... f. 25v. f. 26. f. 26v. f. 27

Slide: Courtesy of Tom Cramer, Stanford

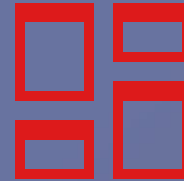
IIIF compatible software



Image Apps



Wellcome Player



Mirador



Internet Archive
Book Reader

Image Clients



IIP Moo Viewer



FSI Viewer

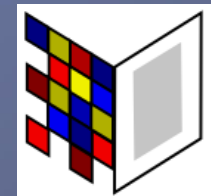
Image Servers



IIP Image



FSI Server



digilib

Technology no. 2: Fedora/ Hydra



- ▶ Fedora is open-source repository software used for ORA and ORA-Data
- ▶ Enables digital objects to be used in web services via well-defined APIs
- ▶ Data exchange via OAI-PMH, integrates linked data and RDF, and supports digital preservation activities
- ▶ Hydra is open-source repository software based on Fedora that allows distinct workflows/ digital asset management tasks
- ▶ Hydra 'heads' can be developed for distinct types of research objects



Fedora/ Hydra in ORA and ORA-Data

My ORA

The publications or datasets you have previously submitted to ORA are listed below. From this page you can complete unfinished ORA contributions that you previously saved for later, or start to deposit a new publication or dataset.

This is a beta version, and we would [welcome your feedback](#) or questions so we can improve the service in our next stage of development work.

Publications

 Add a publication

Journal article

GO >

You have no documents to display. Contribute some of your publications!

[View all publications >](#)

Data

 Add a dataset

GO >

You have no datasets to display. Contribute some of your datasets!

[View all data >](#)

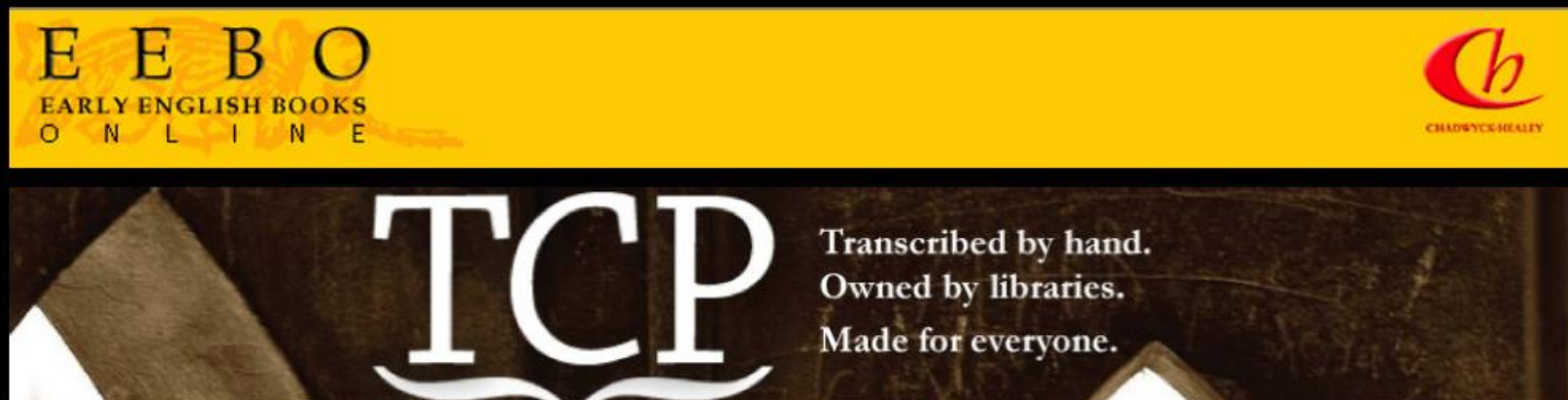
Technology no. 3: XML

- ▶ XML is ... extensible markup language
- ▶ Of which TEI, EAD and METS are but dialects
- ▶ TEI – (Text encoding initiative) used to describe, interpret and improve discovery of texts and create critical editions
- ▶ METS – is an XML schema (set of rules/ vocabulary) for encoding descriptive, administrative, and structural metadata EAD – is an XML standard for encoding catalogues or finding aids for archival collections
- ▶ <http://www.w3.org/TR/REC-xml/>
- ▶ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>



What is a (digital) edition? — Transcription

- Transcribed
 - Early English Books Online Text Creation Partnership

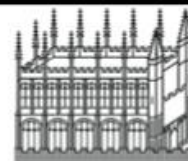


- 130,305 records; 129,541 image sets; 44,314 full texts
- Vendor-transcribed, partially proofed, encoded texts
- Subscription (ProQuest, JISC Historic Text)

<http://eebo.chadwyck.com/>

<http://www.textcreationpartnership.org/>

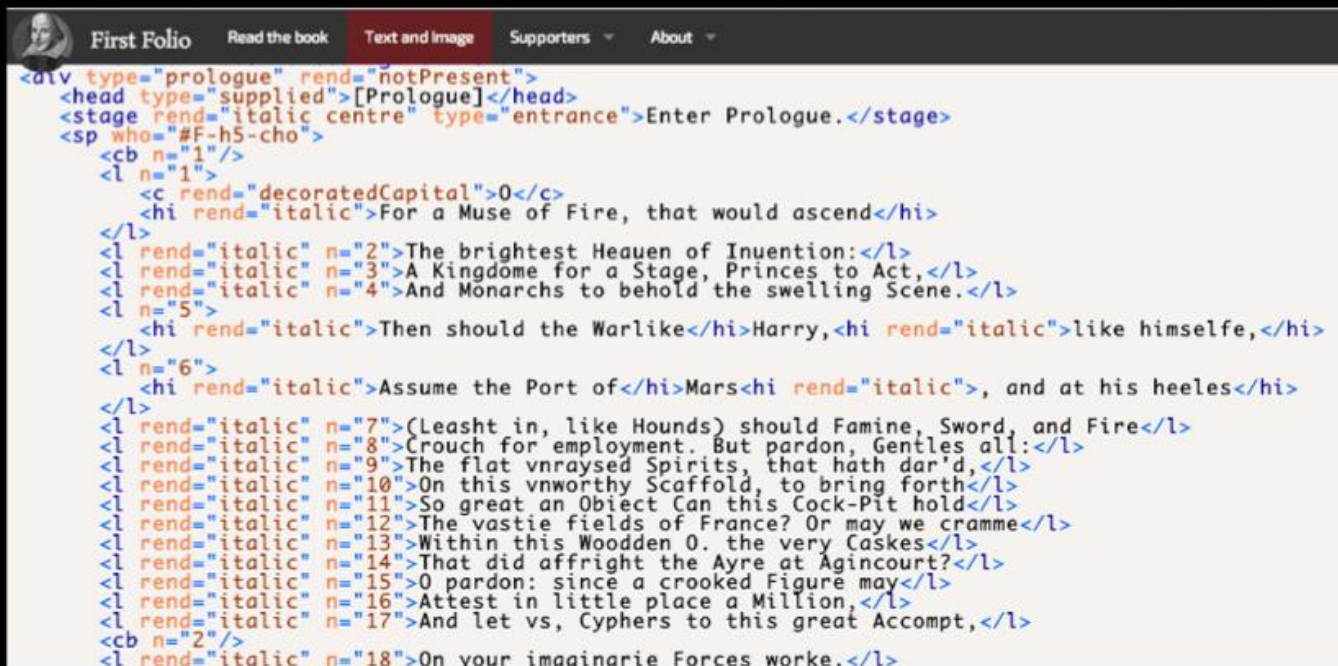
Slide: Pip Willcox, University of Oxford



Bodleian Libraries
UNIVERSITY OF OXFORD

Dialogue: text shall speak peace unto text

• Nobody has ever answered “yes” to
“Let me show you my XML”

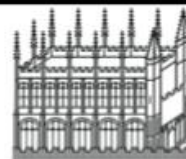


The screenshot shows the 'Text and Image' tab of the First Folio website. It displays XML code for the Prologue of Hamlet. The code uses various attributes like 'type', 'rend', 'n', and 'cb' to format the text. The rendered text is as follows:

```
<div type="prologue" rend="notPresent">
<head type="supplied">[Prologue]</head>
<stage rend="italic" centre" type="entrance">Enter Prologue.</stage>
<sp who="#F-h5-cho">
  <cb n="1"/>
  <l n="1">
    <c rend="decoratedCapital">O</c>
    <hi rend="italic">For a Muse of Fire, that would ascend</hi>
  </l>
  <l rend="italic" n="2">The brightest Heauen of Inuention:</l>
  <l rend="italic" n="3">A Kingdome for a Stage, Princes to Act,</l>
  <l rend="italic" n="4">And Monarchs to behold the swelling Scene.</l>
  <l n="5">
    <hi rend="italic">Then should the Warlike</hi>Harry,<hi rend="italic">like himselfe,</hi>
  </l>
  <l n="6">
    <hi rend="italic">Assume the Port of</hi>Mars<hi rend="italic"> , and at his heeles</hi>
  </l>
  <l rend="italic" n="7">(Leasht in, like Hounds) should Famine, Sword, and Fire</l>
  <l rend="italic" n="8">Crouch for employment. But pardon, Gentles all:</l>
  <l rend="italic" n="9">The flat vnrayed Spirits, that hath dar'd,</l>
  <l rend="italic" n="10">On this vnworthy Scaffold, to bring forth</l>
  <l rend="italic" n="11">So great an Obiect Can this Cock-Pit hold</l>
  <l rend="italic" n="12">The vastie fields of France? Or may we cramme</l>
  <l rend="italic" n="13">Within this Woodden O. the very Caskes</l>
  <l rend="italic" n="14">That did affright the Ayre at Agincourt?</l>
  <l rend="italic" n="15">O pardon: since a crooked Figure may</l>
  <l rend="italic" n="16">Attest in little place a Million,</l>
  <l rend="italic" n="17">And let vs, Cyphers to this great Accompt,</l>
  <cb n="2"/>
  <l rend="italic" n="18">On your imaginarie Forces worke.</l>
</sp>
```

• ...except a computer

<http://firstfolio.bodleian.ox.ac.uk/>



Technology no. 4: BitCurator

- ▶ BitCurator is an open source environment/ suite of tools and workflows
- ▶ Supported by the Andrew W. Mellon Foundation
- ▶ Supports digital curation through the integration of digital forensics tools into workflows
- ▶ Allows imaging of a hard disk, data triage, automated identification of personal/ sensitive information, file system analysing and reporting, and metadata capture
- ▶ www.bitcurator.net



Oxfam Archive at the Bodleian

"Unless Europe can be saved . . .
 . . . civilisation is lost."
"We must come to them . . . as physicians and surgeons at work to heal this wounded and festering continent, the mother of Christendom." Prof. Gilbert Murray, O.M.
 (one of our Hon. Trustees)

THE MAYOR OF VIENNA says his city has reached a peak of misery unknown even in the famine after the first world war.

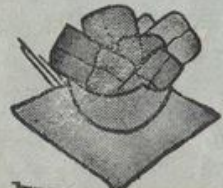
50,000 SCHOOL CHILDREN have been medically examined in Vienna and 70% were suffering from malnutrition.

IN POLAND shoes are almost impossible to get. **IN KIEL** 9,100 children are entirely without shoes.



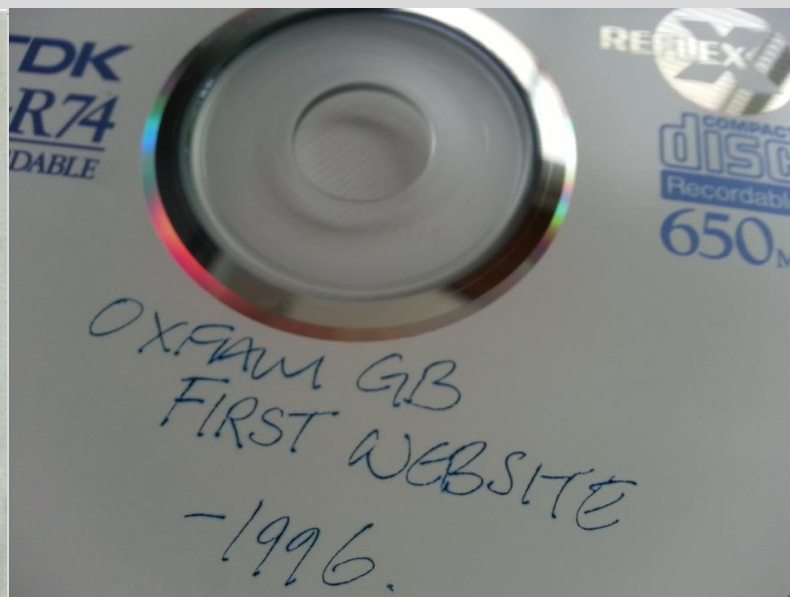
OXFORD RELIEF is entrusted to Friends Relief Service whose voluntary workers* act with care, economy and a personal touch regardless of frontiers, race, politics, class or creed. Many countries have benefited and now the peoples of Austria, Poland and the British Zone of Germany are getting most attention. F.R.S. personnel is less than 40% Quaker, 20% C. of E., many Free Church etc., and some Roman Catholics, united in service.

WE THANK ALL WHO HAVE helped us raise £20,000 in 14 months, plus clothing, etc., worth £10,000, in six months, for European Relief.



Gifts as below will be welcome
 (if to local body new collecting for Europe):—
MONEY FOR FOOD, MEDICAL SUPPLIES,
 Etc., for children, the aged, and special cases.
 (This does not reduce our rations.) Please make cheques payable to "Grady Relief," cross and post to Barclays Old Bank, High Street, Oxford. Sir Alan Pim will acknowledge.
CLOTHING, FOOTWEAR, UNIFORMS, TOWELS, BEDDING, BLACKOUTS, MENDING MATERIALS, Etc. (including clean, mending articles for repair by Displaced Persons sadly needing work), to:—"Grady," 42, Hythe Bridge Street, Oxford. Lady Franks will acknowledge (if address supplied).
BOOKS AND PERIODICALS (any language; for Reading Rooms) to "GRADELY," c/o C.O.B.S.R.A., 75 Victoria St., London S.W.1. (M.G.W., "Lancet," etc., 1939-47, also needed.)

OXFORD FOR FAMINE RELIEF
COMMITTEE (Registered War Charities Act, 1940) **FOR THE FRIENDLESS**
 N.B.—We cannot send clothing parcels to individuals.



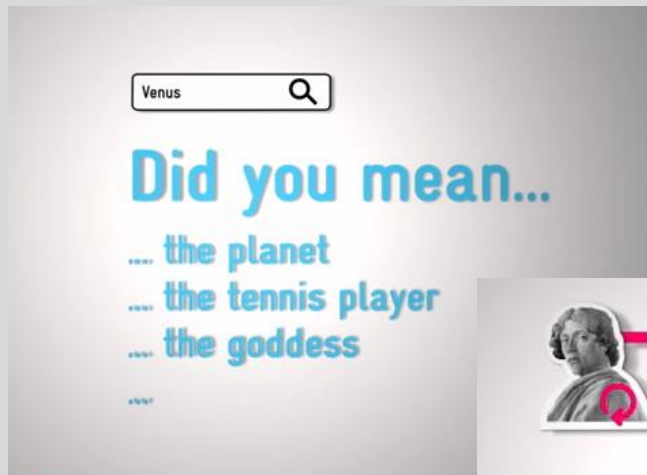
14/04/2010

Technology no. 5: Linked Data

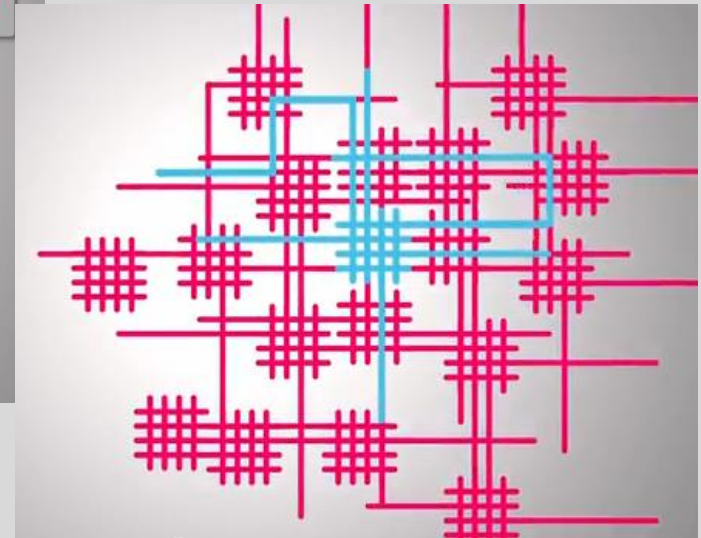
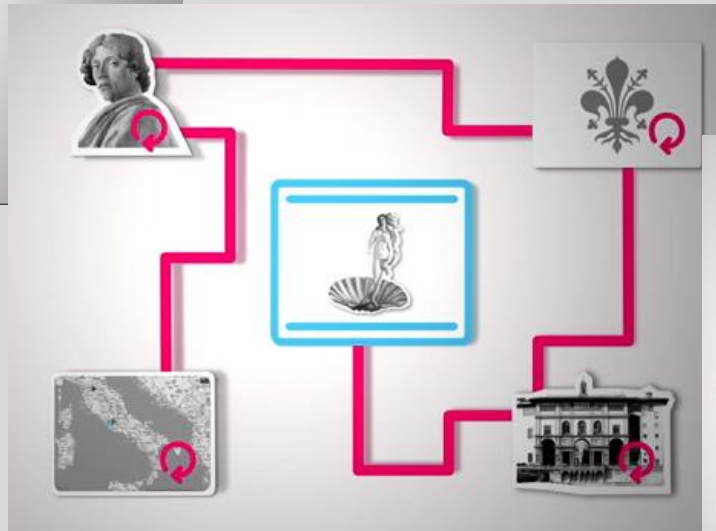
- ▶ Linked data is technology which is fundamental to the semantic web
 - ▶ The semantic web enables human-readable context to be derived by computers from machine-readable information about things (anything!)
 - ▶ To link data, publish it as RDF 'triples' and give each RDF element a URI (unique reference on the web)
 - ▶ Use a web vocabulary (known as an ontology) to describe the data and relationships between the data (e.g. geonames)
 - ▶ Use a web query language to query the data (e.g. SPARQL)
 - ▶ Derive new insight and context from the data!
 - ▶ <http://www.geonames.org/ontology/documentation.html>
 - ▶ <http://www.w3.org/standards/semanticweb/data>
 - ▶ <http://www.w3.org/DesignIssues/LinkedData.html>
-



Which Venus did you mean?



1. Make raw data about content available
2. Make it openly licensed
3. Put it in RDF format – make it linkable ‘triples’
4. Put in on web using http



Linked Data in Cultures of Knowledge



Networking the Republic of Letters, 1550-1750

[HOME](#) [ABOUT](#) [EMLO](#) [PROJECTS](#) [EVENTS](#) [RESOURCES](#) [BLOG](#) [TWITTER](#)

Early Modern Letters Online

Early Modern Letters Online – or EMLO for short – is our growing union catalogue of sixteenth-, seventeenth-, and eighteenth-century correspondence. It is at the heart of what we do. Between now and the end of 2014, we are adding **New Metadata**, **New Features**, and **New Collaborators**. Find out more below, and **Get Involved!**

**Wallis, John (Dr),
1616-1703**

1086



**Dury, John, 1596-
1680**

991



**Aubrey, John,
1626-1697**

842



**Vossius, Gerardus
Joannes, 1577-1649**

823



3	Letter	12 August 1562	Cecil, William, 1520-1598	Smallbridge Hall, Bures St Mary, Suffolk, England	Parker, Ma
4	Letter	21 October 1562	Zuichemus, Viglius, 1507- 1577	Brussels, Brussels Capital Region, (Spanish Netherlands) Belgium	Montzima, 1556-74)
5	Letter	19 January 1563	Zuichemus, Viglius, 1507- 1577	Brussels, Brussels Capital Region, (Spanish Netherlands) Belgium	Montzima, 1556-74)

Technology no. 6: Research objects

- ▶ Research objects are packages of formal and informal research outputs
- ▶ Articles, data, resources, workflows, software/ code, methodologies
- ▶ Leads to need for code/workflow repositories and citation of data, software, workflows, resources
- ▶ Research outputs beyond the pdf!
- ▶ <http://www.myexperiment.org>
- ▶ <http://openresearchsoftware.metajnl.com>
- ▶ <http://www.researchobject.org>



First time visitor? Try these videos:

 [Project Introduction](#)

 [Bioinformatics Case Study](#)

Use myExperiment to...

 [Find Workflows](#)

 [Share Your Workflows and Files](#)

 [Create and Find Packs of Items](#)

 [Find People and Make Friends](#)

 [Create and Join Groups](#)

 [Build your Profile and Reputation](#)

 [Tag and Rate things](#)

 [Write Reviews and Comments](#)

Why reproducibility is key to science

The Telegraph

[Home](#) [Video](#) [News](#) [World](#) [Sport](#) [Finance](#) [Comment](#) [Culture](#) [Travel](#) [Life](#) [Women](#) [Fashion](#)
[Blogs](#) [Columnists](#) [Personal View](#) [Telegraph View](#) [Letters](#) [Cartoon Archive](#) [My Telegraph](#) [Politics](#)

[HOME](#) » [COMMENT](#) » [COLUMNISTS](#) » [CHRISTOPHER BOOKER](#)

Climate change: this is the worst scientific scandal of our generation

Our hopelessly compromised scientific establishment cannot be allowed to get away with the Climategate whitewash, says Christopher Booker.



9K



284



1



7



10K



Email



CO2 emissions will be on top of the agenda at the Copenhagen summit in December Photo: Getty

Print this article

Christopher Booker

[Earth Debates »](#)
[Green Politics »](#)
[Comment »](#)
[Columnists »](#)
[Personal View »](#)

In Christopher Booker



The climate change scare is dying, but do our MPs notice?

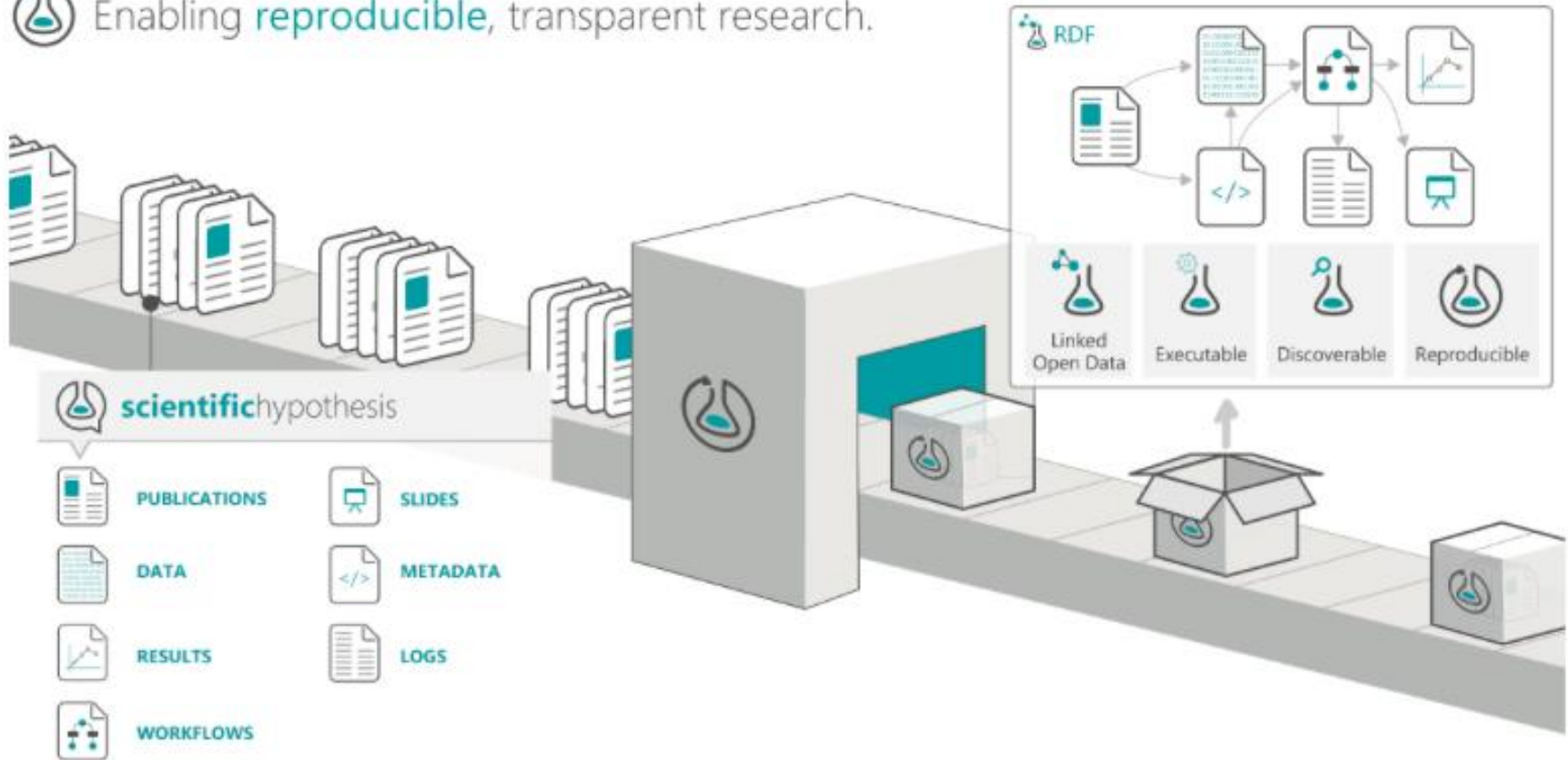


By Christopher Booker

6:10PM GMT 28 Nov 2009

Beyond the PDF

Enabling **reproducible**, transparent research.



My technology rules to live by

- ▶ 1. Always ask questions
- ▶ 2. Look and feel, make and play
- ▶ 3. Be open and collaborative



A final word from Thomas Bodley

*Quod feliciter vortat academici
Oxoniens bibliothecam hanc vobis
reipublicaeque literatorum T.B.P.*

*Thomas Bodley has built this
library for you and for the Republic
of the Learned. May the gift turn
out well.*



Thank you for listening
Any questions?



Lucie.Burgess@bodleian.ox.ac.uk
Twitter @LucieCBurgess

