The Digital Library and Researchers: from platforms to partnerships

Huw Jones
Head of Digital Library Unit and Digital Humanities Coordinator
Cambridge Digital Library

- A unified platform for digital content
- Launched in 2011 with the Isaac Newton Papers
- 35,000 items, 500,000 images
- Video, 3D, specialist imaging
- Representations and research
- Access, impact and sustainability
- Funding model – based on research
Hosting and creating data
Digitisation for access and impact

Search for “Isaac Newton Notebook”
(or Book of Cerne, Book of Deer etc.)
Hosting research outputs

Reading through research

https://cudl.lib.cam.ac.uk/view/MS-LL-00001-00010/1
Hosting research outputs

Text and translation

https://cudl.lib.cam.ac.uk/view/MS-PETERHOUSE-00075-00001/148
Library as a research partner
Investigative Imaging

3D photogrammetry

https://cudl.lib.cam.ac.uk/view/MS-CUL-00001-00155/1
Investigative Imaging

High resolution imaging of large items

https://cudl.lib.cam.ac.uk/view/MS-MAPS-MS-PLANS-R-C-00003/1
Specialist imaging and text

Mingana-Lewis

Palimpsest

https://cudl.lib.cam.ac.uk/view/MS-OR-01287-SMALL/3
Creation and use of textual corpora
<table>
<thead>
<tr>
<th>Left</th>
<th>Node</th>
<th>Right</th>
<th>Book</th>
<th>In bk.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. room was beautiful with them from New Year's Day to Christmas</td>
<td>1. In those six years I had never been away except</td>
<td></td>
<td>BH</td>
<td></td>
</tr>
<tr>
<td>2. yard, always troubled with a class-grievance (probably</td>
<td>2. may be reminiscent of that summer morning wrongly</td>
<td></td>
<td>BH</td>
<td></td>
</tr>
<tr>
<td>3. are called family affairs, which are only fit for plebeian</td>
<td>3. days, and have no manner of business with people of</td>
<td></td>
<td>BR</td>
<td></td>
</tr>
<tr>
<td>4. as to the feestings and christenings, and revelings at</td>
<td>4. and celebrations of birthdays, wedding-days, and all ma BB</td>
<td></td>
<td>BR</td>
<td></td>
</tr>
<tr>
<td>5. 'what, my dear,' said my aunt, one morning in the Christmas</td>
<td>5. season when I left school; as this knotty point is</td>
<td></td>
<td>BG</td>
<td></td>
</tr>
<tr>
<td>6. great girl in the world. ¶ They expected me home before</td>
<td>6. but had no idea of my returning so soon. ¶</td>
<td></td>
<td>DC</td>
<td></td>
</tr>
<tr>
<td>7. before their time. Mental green-paws were produced at Christmas</td>
<td>7. and intellectual asparagus all the year round. Mathematical QS</td>
<td></td>
<td>BG</td>
<td></td>
</tr>
<tr>
<td>8. gone away only this morning. He will be back at Christmas</td>
<td>8. 'Nothing could happen better. You will, on his return a ED</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>9. Nothing could happen better. You will, on his return at</td>
<td>9. arrange all matters of detail with him: you will then</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>10. I ask you, most kindly to come to me at Christmas, if I had</td>
<td>10. any particular thing to say to you? ¶ 'Why</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>11. slips together, when Mr. Edwin Drood comes back here at</td>
<td>11. and then you and I will step in, and put</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>12. you settled with her that you would come back at Christmas</td>
<td>12. observed Jasper. ¶ I see Mr. Weggles, as you quite fail ED</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>13. cue from you. I accept it. I understand that at Christmas</td>
<td>13. they will complete their preparations for May, and that</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>14. here, dear old boy. Ask Mr. Landless to dinner on Christmas</td>
<td>14. Eve (the better the day the better the deed), and</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>15. turn back; and the last day of the week is</td>
<td>15. Eve. ¶ 'You may be certain of me, sir.' ¶ 'The echoes</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>16. dog gives when a person's dead. That was MY last Christmas</td>
<td>16. Eve. ¶ 'What do you mean? is the very abrupt, and</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>17. establishment was about to undergo a serene hush.</td>
<td>17. recess was at hand. What had once, and at no</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
<tr>
<td>18. dine with the dear fellow to-morrow and next day-- Christmas</td>
<td>18. Eve and Christmas Day-- but it would never do to</td>
<td></td>
<td>ED</td>
<td></td>
</tr>
</tbody>
</table>

[https://clic.bham.ac.uk/concordance](https://clic.bham.ac.uk/concordance)
https://casebooks.lib.cam.ac.uk/identified-entities/PERSO12608
Manuscript science
• Parchment analysis
• Genetic analysis
• Ink analysis
• Epidemiology
• Economic patterns
• Agricultural history
Beasts to Craft

https://sites.google.com/palaeome.org/ercb2c/home?authuser=0
The importance of TEI

- Major metadata standard for digital humanities
- Can hold both library/archive data and research output
- Interoperable, reusable, sustainable
- The basis for collaborative projects between researchers and the Library
- Workshops for researchers
- Training embedded into postgraduate courses
The library’s role in the creation of digital resources

- Sustainability of data and platforms in a way that projects can’t
- Expertise and experience in standardisation and interoperability
- Knowledge of licensing, copyright and accessibility
- Understanding of the importance of identifiers and documentation
- Platforms as a driver of standardisation
- Importance of documenting methods and making datasets available
- Libraries are having a big effect on the use of e.g. TEI, IIIF in research contexts
Making data available to researchers
Images and IIIF

• 2015 – “The dynamic digital library – a IIIF scoping project”
• Mellon funded – also to investigate JPEG2K as a preservation format
• Now – fully IIIF compliant
• Over 500K images available via IIIF
• Drops the cost of image-based research – so encourages innovation and experimentation
Great skill is required to achieve a convincing result, but the beautiful gradations of colour in this work have led to its reputation as "perhaps the most beautiful set of prints ever made".

The work is divided into eight categories: birds, plums, orchids, bamboos, fruit, stones, ink drawings (round fans) and miscellany. Each category is divided into two fascicules. The leaves are printed on one side only, folded in half and glued together along the outer fold (the so-called 'butterfly' binding). With the exception of one category, every image is followed by an accompanying text, in most cases a poem.

This copy has been identified by the leading scholar of this work as the finest and only extant complete copy in the original binding of what he describes as the 'second superstate' of the first edition.
IIIF and comparison
IIIF as an aggregator

https://iiif.bibliissima.fr/collections/
IIIF – images as a dataset

Colour analysis
Pattern matching
Visual search

http://zeus.robots.ox.ac.uk/15cillustration/
4804
C. thickish, white. odourless.

265. A large tuft in a forest road.

P. conical, mid brown, very fibrillose surface, cretaceous, radially, dry, dry, margin wavy.

S. cylindrical, mid brown, whitish below, with dense small, whitish, upstanding fibrils, fleshy, slightly hollow, shiny, naked.

L. thin, crowded, few incomplete, adnate, size small, half brown, white edged.

E. thin, nearly white, 1+ odourless. Odour recalling

4805
Chlorophyllum

Inocybe fuscata

Sp. mid brown, smooth.

3. Suillus grevillei

4. Polyporus

x 1/2

x 1
IIIF – the cultural effect

- Open, collaborative, communal
- Inclusive – reaching out to STEM, to pedagogy, to non-technical people
- Based on shared development – low cost, low risk
- Culture is having a broader effect in the sector
- Shared datasets, shared development, shared infrastructure
Datasets – from download ...
... to GitHub
Mapping Manuscript Migrations

http://mappingmanuscriptmigrations.org/
Future developments
Technical

• IIIF for discovery
• IIIF for audio, video, 3D
• ITIF? IIIF-like mechanism for text
• Automated analysis/Al for heritage datasets
• Collaboration around scarce resources – skills and infrastructure
Cultural

- Data should be available, comprehensible, identifiable
- Culture of innovation encouraged by low cost, low risk
- Specialisation in a broader DH landscape
- Data over interfaces
- Standards for access over standards for storage
- Comprehensibility over metadata standards
- Rise of identifier schemes
Library as a research leader?

- Blurring of lines between research and library activities
- Funding bids where library is the lead partner
- Identifying collections, research questions and methodologies
- Emphasis on datasets and methodology as a research outcome
• Cambridge Digital Library https://cudl.lib.cam.ac.uk/

• Cambridge Digital Humanities https://www.cdh.cam.ac.uk/

• @CamDigLib