The Wellcome Library did not want to introduce an off-the-shelf “silo” digital library product.
How do we get to this?
It is interesting to contemplate a tangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth; and to reflect that these elaborately constructed forms, so different from each other, and dependent upon each other in so complex a manner, have all been produced by laws acting around us. These laws, taken in the largest sense, being Growth with Reproduction; Inheritance which is almost implied by reproduction; Variability from the indirect and direct action of the conditions of life, and from use and disuse: a Ratio of Increase so high as to lead to a Struggle for Life, and as a consequence to Natural Selection, entailing Divergence of Character and the Extinction of less-improved forms. Thus, from the war of nature, from famine and death, the most exalted object which we can capable of conceiving, namely, the production of the higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed by the Creator into a few forms or into one; and that, whilst this planet has gone cycling according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved.
Features

- Deep Zoom
- Search within
- Bibliographic metadata
- Access control
- Thumbnails
- Standalone player, can be embedded
- Audio, video, Born-digital

- http://wellcomelibrary.org/player/b18032096
Digitisation Workflow

- cat record: b18021013
- DDS
- METS
- Digital Asset Management System

METS
b18021013.xml

Read METS

DDS

- Serve JSON data
- Request data about b18021013

Serve JPG tile generated from JP2

Request data about b18021013

Serve JPG tile generated from JP2

Digital Asset Management System

Read JP2, Audio, Video

library.wellcome.ac.uk/player/b18021013

Wellcome Library | P X

library.wellcome.ac.uk/player/b18021013

Wellcome Library
The DDS

• Takes information from METS, METS-ALTO, Catalogue to generate presentation data
• Provides Image services to generate deep zoom tiles, arbitrary crops, and PDFs
• Provides APIs for search within, authentication and authorisation and biblio data
• **Package:** [http://wellcomelibrary.org/package/b20432033](http://wellcomelibrary.org/package/b20432033) (will be IIIF eventually)

• **Tiles:**
  [http://wellcomelibrary.org/dz/b20432033/0/d7b80202-a1d7-4f88-ab57-0f6b369fe41b.jp2_files/12/2_3.jpg](http://wellcomelibrary.org/dz/b20432033/0/d7b80202-a1d7-4f88-ab57-0f6b369fe41b.jp2_files/12/2_3.jpg)
• Services:

```
[
  "band",
  "bank",
  "banish",
  "bantam",
  "bankiva",
  "bantams",
  "banished",
  "bandicootsand"
]
```

```
"y": 1874,
"w": 188,
"h": 44,
"hit": 0,
"before": "cases it can be shown that widely-different checks act on the same species in different districts. When we look at the plants and bushes clothing",
"word": "entangled",
"after": "banks, we are tempted to attribute their proportional numbers and kinds to what we call chance. But how false a view is this! Every one has heard that",
}

", "index": 222,
"rects": [
  {"x": 1841,
   "y": 304,
   "w": 283,
   "h": 45,
   "hit": 1,
   "before": "ng like long ears, they swept a branch rapidly across the object-glass of my microscope. When a branch was placed on its face, the vibracula became",
   "word": "entangled",
   "after": "and they made violent efforts to free themselves. They are supposed to serve as a defence, and may be seen, as Mr. Busk remarks, "to sweep slowly and"
  }
],
"index": 457,
```
The DDS

- A combination of existing software and bespoke application

- IIPImage for generating jpeg tiles and other images from archival JPEG2000 files

- Bespoke software for generating JSON packages and other API
Challenge: tiles at scale

- SDB (DAM) is a long way from the public facing web application
- SDB provides one way to get a JPEG2000: http://sdb.wellcome.ac.uk/sdb/rest/digitalFileContents/{0}

- IIPImage needs a local file path, not a URL – can’t have fast random access to file like that http://WT-MANILA/cgi-bin/iipsrv.fcgi?DeepZoom=/path/to/image.jp2_files/12/2_3

/path/to/image.jp2 – needs to be a local file system on the IIPImage server
Tile Request

Load Balancer selects stack

Reverse Proxy and Routing
Will serve tile if it has it

Load Balancer selects stack

Access control decision:
Biblio: b20432033
Sequence: 0
Image ID: d7b80202-a1d7....

Does file exist on 5TB cache?

If No, fetch from SDB "REST" API

Log JP2 usage

Add cache headers if appropriate

Rewrite URL

Store in cache if headers permit (100GB)

IIP Image

SDB

http://wellcomelibrary.org/dz/b20432033/0/d7b80202-a1d7-4f88-ab57-0f6b369fe41b.jp2_files/12/2_3.jpg
<rule name="Reverse Proxy to tile">
   <action type="Rewrite" 
</rule>
5TB cache

• The 5TB of JP2s that IIImage needs most are available on fast disk

• This cache is maintained by a “scavenger” process

Read logs to determine scavenge candidates
PDF Generation

b20549591, Manifestation 0
Museum Truscottanum, Start. 2014-10-31 12:40
222/222
Warnings: no warnings

b1863042x, Manifestation 0
General account book: bills received and cash disbursements
Start. 2014-10-30 13:20
293/293
Warnings: no warnings

b21557354, Manifestation 0
Photograph of Henry Wellcome's camp in Maine. 1886
Start. 2014-10-29 11:40
1/1
Ended: 2014-10-29 11:40
Warnings: no warnings

b21417428, Manifestation 0
[WMS Arabic 396]
Start. 2014-10-28 18:40
212/212
Warnings: no warnings

b20295212, Manifestation 0
[WMS Arabic 90]
Start. 2014-10-28 15:40
79/79
Ended: 2014-10-28 15:45
Warnings: no warnings

b21557299, Manifestation 0
Advertising panel portraying a pregnant

View Queue

Cache status

Number of JP2s registered in cache
523908

Storage root
\wt-nashville\thecache

Image Server URL mapped to cache
http://WT-MINSK/cgi-bin/ipsrv.cgi?FIF=/ddsficache

Database
Server=\wt-havana\Database=DDsInstrumentation;user=ddsinst, password=**********

Required free space
279 GB (300000000000)

Current free space
277 GB (297782007200)

Number of files per deletion batch
100

Cache threshold exceeded. This app will delete files if run now with the scaveng parameter.

Please be careful with the following link... it needs to look at ALL the METS files

Launch METS report

Unfinished PDFs

PDFs with warnings

Other warnings

Recent cache activity

View top 100 scaveng candidates

View LAST 100 scaveng candidates

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-10-30</td>
<td>ScavengeEnd</td>
<td>Finished scavenging, 1300 files deleted, freeBytes=300010692008, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetEnd</td>
<td>END Scavenge set 13, freeBytes=300010692008, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetEnd</td>
<td>END Scavenge set 12, freeBytes=296273994240, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetStart</td>
<td>Scavenge set 13, freeBytes=296273994240, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetStart</td>
<td>Scavenge set 11, freeBytes=296182603108, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetStart</td>
<td>Scavenge set 12, freeBytes=296182603108, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetEnd</td>
<td>END Scavenge set 11, freeBytes=296182603108, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetEnd</td>
<td>END Scavenge set 10, freeBytes=297022609742, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetEnd</td>
<td>END Scavenge set 9, freeBytes=297022609742, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetEnd</td>
<td>END Scavenge set 8, freeBytes=296224600704, freeSpaceRequiredBytes=300000000000</td>
</tr>
<tr>
<td>2014-10-30</td>
<td>ScavengeSetEnd</td>
<td>END Scavenge set 7, freeBytes=29549403206, freeSpaceRequiredBytes=300000000000</td>
</tr>
</tbody>
</table>
IIPI mage

- Keep IIPI mage boxes completely vanilla – just lighttpd and IIPI mage
- Caching everywhere
- Check for file heavily optimised
- Fetch from SDB the weak point

- Only JP2s cached this way (volatile) – everything else is permanently available
Infrastructure Review

- Caching/scavenging strategy
- Is 5TB enough? “long-tail”
- Can this problem be solved a different way?