The Turing Digital Archive b. 2000

PATRICIA MCGUIRE

ARCHIVIST since 2004

KING'S COLLEGE, CAMBRIDGE

TIMELINE

1931 – Alan Turing at King's
1935 – *Entscheidungsproblem* paper
1960 – Sara Turing gives some of her papers
1997 – First approach from Southampton Dep. of Elect. and Comp. Sci.
2000 – Pilot launched, 1/3 of the Turing papers
2004 – 'Complete' Turing Papers on-line
2008 – Server moves to King's and is redesigned (without databases?)
2021 – Website in its current form

Aims/Benefits

- **Using a small collection to set up a paradigm metadata/database driven dynamically generated pages,**
- **that make our cultural heritage available on-line,**
- **but only to view (protect copyrights),**
- **with added-value essays and guidance for various user groups.**
- **Link to or include material held elsewhere.**
- **Use OCR to harvest keywords or transcripts.**
- **Increase profile of Southampton and King's.**
- **Preservation (with less consultation of the original papers) not a priority.**

Costs

- **Develop and populate database for structured metadata**
- **Scanning – 1.7 to 4.6 minutes per image**
- **Hardware – scanner and web server (4 GB RAM and 300 GB disk space on existing server, DB2 database)**
- **Copyright – clear and/or pay licensing fees**

Pilot: ***papers that are copyright-'free'***

- **786 files/sheets/images/original scans, 11 GB**
- **Develop databases: 14 based on Dublin Core Metadata**
- **Scan the images (8-bit greyscale @600 dpi or 300 for plain photocopies, 24-bit colour 300 dpi .tif), break into tiles**
- **c. £25,000, half funded by Southampton, half IEEE and BCS**

Management (board met 3 times in 2000-2001): Southampton Comp. Sci., Turing scholars, KC archivists

Full scale: ***other 2/3***

- **Clear copyright (database has 506 entries)**
- **Update server, moved to SQL**
- **Elsevier gave c. £25,000**
- **New server at King's**
- **Now (150 dpi .pdf images only) 3287 files, 4 GB for c. 261 catalogue entries/webpages**
- **Underlying .tif: 46 GB, 3474 image files**

Ongoing costs

- **Updating hardware**
- **Updating web design**
- **Adding new accessions**
- **Domain name(s) licenses**

Lessons Learned

- **Project planning: minimum deliverables *vs* (prioritised) extras, get your budget and funding lined up**
- **You can create an archive-shaped digitised archive!**
- **Archivists need to fight for superiority of the catalogue and preservation of copyright, IT need to fight for infrastructure including updates of underlying databases, image viewers and web interfaces**
- **Choice of viewers affects ability to read the wide variety of sized documents, and to protect copyright and control of images**
- **Copyright is a bear to clear and assurances may become harder to implement**
- **Website redesign is an ongoing cost and features are lost as new ones are added**
- **New accessions are easier if the archivist can add pages but that requires a simpler infrastructure**
- **Scanning at 300/600 dpi .tifs, keep the scans handy, foliate and tell people where they are in a document**
- **You can’t do OCR on archives**
- **It’s easy to recatalogue on-line**
- **Improved web search engines may have mitigated the need to index but keywords/tags are still useful**

If we were starting from scratch

- **Might not incorporate the archival info as metadata**
- **Images might be tile-able automatically**

QUESTIONS?

Patricia McGuire
Archivist
King's College, Cambridge CB2 1ST
archivist@kings.cam.ac.uk
https://turingarchive.kings.cam.ac.uk

Original development archivists:
Ros Moad and Jacky Cox
Original tech. contact:
Kirk Martinez