Putting 600,000 books online

The large-scale digitisation partnership between the Austrian National Library and Google

Max Kaiser
Head R&D, Austrian National Library

40th LIBER Annual Conference
Barcelona 29 June – 2 July, 2011
Austrian Books Online

www.onb.ac.at/ev/austrianbooksonline/
Digitisation of the entire historical book holdings of the Austrian National Library
Largest Austrian Public Private Partnership in the cultural sector
History back to the 14th century
One of the world’s most significant historical collections
„Legal Deposit“
Digitisation
Tagesübersicht: 4. Januar 1871
→ Digitisation
→ Online Publications
→ Web Archiving
6 years
600,000 volumes
180 Mio pages
Libraries in Europe

National Libraries

- Italy
- Austria
- The Netherlands
- Czech Republic
- Great Britain
"Stimulating the flow of private funds for the digitisation of cultural assets through equitable public private partnerships appears as a viable and sustainable way of tackling the pressing question of making Europe’s cultural wealth accessible online and preserving it for future generations."
“The key question is not whether public-private partnerships for digitisation should be encouraged, but how, and under which conditions.”
Cornerstones:

→ Respect for rights holders
  → Only public-domain works digitised

→ Transparency
  → Very detailed FAQs online

→ Access
  → Free-of-charge access worldwide
Cornerstones:

→ Digital copies for library
  → Identical with copies used by Google

→ Re-use
  → All files available for non-commercial use
  → Access via platforms like Europeana
  → Provision to research partners

→ Non-exclusivity
  → Free to digitise material with other partners
Additional key elements:

→ Selection of books by library
→ Institute for Restoration involved
→ Termination
Who is paying for what?
Costs

→ Full text-digitisation: very expensive

→ Report by Collections Trust for Comité des Sages

70–100 Euro / book

→ Typical costs for projects with external service providers
Google:

- Transport
- Insurance
- Scanning
- OCR
- Image processing
- Quality control
- Google Books
Austrian National Library:

→ Selection
→ Internal logistics
→ Conservational assessment
→ Metadata
→ Barcoding
→ Data download and control
→ Data storage & digital preservation
→ Digital Library
Which books?
200,000 volumes
State Hall
Entire historical book holdings 16th – 19th century
Department of Manuscripts and Rare Books
Map Department
Theatre Museum
Conservation

Preservation

7

Work Packages

- Book logistics
- Metadata / Catalogues
- Conservation / Restoration
- Data download / Quality control
- Access
- IT infrastructure
- Project management
Preparatory project
Mid - end 2010

→ Integration with organisational processes
→ Personnel resources
→ Logistics workflows
Internal communication

→ Change processes
→ Re-evaluation of workflows
→ Availability of internal resources
Consultation with other Google partners
70+ staff members
20+ exclusively for project

→ Book logistics
→ Metadata adaptation
→ Cataloguing
→ Conservation / restoration
→ Quality control
→ Software implementation
→ Project management
End of 2010
Test shipment & Start operational project

Spring 2011
Start of digitisation
No individual selection ...
Condition
Preparation
Conservational evaluation

<1% excluded for conservational reasons
Logistics in the State Hall
Logistics in the State Hall
Logistics in the State Hall
Challenges...
Challenges...
Challenges...
Challenges...
Challenges...
Logistics in the "Aurum" Depot
Logistics in the "Aurum" Depot
Preparation for Digitisation
Barcoding
Adaptation of metadata
4 minutes / volume
40.000 hours
5.128 working days
24,4 person years
Complex cases ...
Adligata ...
Adligata ...
„Slim“ volumes ...
Conservational protection
Conservational protection
Cataloguing of the Fidei Commiss Library
Cataloguing of the Fidei Commiss Library
Ready for Digitisation ...
Digitisation

→ Scanning Center in Germany
→ Procedures agreed
→ Austrian Federal Office for Monuments involved
→ Each volume checked after return
→ Books unavailable to users for ~ 3 months
30 Mio. digitised items / year
143,000 digitised items / day
Quality control

→ Automated jobs
→ Representative samples
→ IT assisted discovery of error clusters
→ Error candidates checked manually

→ Crowdsourcing?
Data Management & Access

- Data storage: inhouse
- JPEG-2000 master files stored redundantly
- Access copies generated on-the-fly
- URN resolver for permanent identification
Outlook

→ Full-Text: new possibilities for research
→ Data enrichment
→ Named Entity Recognition
→ Linked Data
→ New data centric research in the Humanities & Social Sciences

http://www.diggingintodata.org/
Huy! und Pfuy! Der Welt Huy Oder Aufrischung Zu allen schönen Tugenden: Pfuy ... By Abraham a Sancta Clara
Rede wider das Lesen böser Bücher. Hrsg. von einem Priester aus dem ...
More information

www.onb.ac.at/ev/austrianbooksonline
www.onb.ac.at/ev/austrianbooksonline/faq.htm
twitter.com/abooksonline
Happy Birthday!
Thank you!

max.kaiser@onb.ac.at
www.onb.ac.at
www.linkedin.com/in/maxkaiser

twitter.com/maxkaiser