Research partnerships, user participation, extended outreach – some of ETH Library’s steps beyond digitization
From «boutique» to mass digitization

Small scale digital reproductions on demand

Dual strategy «Digitale Bibliothek ETH Zürich»: e-publications and digitization

External scanning for early online platforms

DigiCenter mass production
- rare books: 0.8 mio
- journals: 0.3 mio
- photographs: 0.1 mio
- archival material: 0.5 mio
Total of ~ 1.7 mio. scans p.a.

DigiCenter established as an internal service unit,
~ 600,000 scans p.a.
(Cooperative) online platforms for digitized content

- **Rare books**
  - **e-rara**
  - ~57,600 books
    (share ETH Library: ~14,300 books)
  - Public Domain Mark

- **Images**
  - **e-pics Image Archive**
  - ~385,000 images
    - Mostly Public Domain Mark and CC BY-SA

- **Archival material**
  - **e-manuscripta**
  - ~54,800 sources
    (share ETH Library: ~13,400 sources)
  - Mostly Public Domain Mark

- **Journals**
  - **e-periodica**
  - ~300 journals
    (~600,000 articles)
How can we realise the full potential of digitized content?

- Research partnerships
- User participation
- Outreach
Potential I
Research partnerships
Providing a machine readable corpus: Enhancement of OCR'd text

- **Available material:** digitized journals as PDFs with OCR in e-periodica
- **Additional service (aim):**
  - Provide access to the texts in a machine readable format
  - Text enrichment (named entity recognition)
- **Test case:**
  - Two sample journals (~ 380,000 pages)
  - Tokenized text-files with OCR results
  - Aim I: Automatic correction of OCR errors
  - Aim II: Explore possibilities of automatic text enrichment (recognition and linking of the named entities [persons] and [place names])
- **Research partner:**
  - Professor Martin Volk (University of Zurich, Institute of Computational Linguistics)
Providing a machine readable corpus: Enhancement of OCRed text

DIE VIELEN WAHRHEITEN DER STADT

«Multiple City» und «New Urbanity»: Blick auf Planungskonzepte der letzten 100 Jahre und auf die Stadtwirklichkeit von heute in München, gute Beispiele aktueller Stadtplanungspraxis in Europa in Frankfurt, so lässt sich zusammenfassen, was die Ankündigungen versprechen. Die beiden Ausstellungen scheinen sich, obwohl nicht aufeinander abgestimmt, ideal zu ergänzen.

Building on existing solutions: The “Thomas Mann personal library” project

- **Access to Thomas Mann’s markings in his personal library**
  - Library is part of the Thomas Mann Archive, ETH Library
  - Integral digitization of 2,650 books containing handwritten traces
  - Transcription of the marginalia
  - Front-end for searching and displaying markings (building on e-periodica components)
  - Project duration: April 2016 to March 2019

- **Research partners**
  - Professor Andreas Kilcher (ETH Zurich, Chair of Literary and Cultural Studies)
  - Funded by Swiss National Science Foundation
Providing primary data on demand:
Data extraction from analogue long-term measurement series

- **Seismograms for seismological research**
  - Series of paper seismograms from 1927 to 1955
  - Archival description, partial digitization and publication on e-manuscripa.ch
  - Partial data extraction by the Swiss Seismological Service at ETH Zurich

- **Sunspot drawings for solar research**
  - Sun observation at ETH Zurich from 1884 to 1995
  - Large series of sunspot drawings, data sheets, and sunspot statistics at ETH Zurich University Archives
  - Archival description, digitization and publication on e-manuscripa.ch
  - Data extraction (years 1884–1949) by international solar researchers

Sunspot drawing, 5 April 1894, ETH University Archives, Hs 1304:1572
Potential II
User participation
Online crowdsourcing via E-pics Image Archive Online
Aim: fill the gaps in contemporary descriptions
Step 1: Cooperation with Swissair retirees
- Project duration: end of 2009 until end of 2013
- Closed group of experts
- 135 interested volunteers, including:
  - 44 regular users and ½ dozen experts
- 40,000 images processed

Lessons learned
- Importance of personal contacts
- Effort involved heavily underestimated
- Added value gained: priceless!
Image Archive: Crowdsourcing to fill patchy image descriptions

- **Step 2: All volunteers welcome**
  - New comment function online since 9 Dec 2015
  - Huge reader and media response (newspapers, radio, TV)

- **Statistics (1 July 2017)**
  - 22,600 pointers received
  - 20,543 description of images improved
  - 798 volunteers (90% men), very active top ten

- **Ongoing community management**
  - Weblog [ETH Bibliothek Crowdsourcing](#)
  - [Videos](#) and events

---

Original title: Untitled, 17.8.1932
New: Flawil-Gossau, reservoir in Glatt-Tobel near the ruins of Helfenberg Castle, 17.8.1932
Archival sources: Crowdsourcing to transcribe manuscripts

- **Project “e-manuscripta full text”**
  - Add transcription module to [e-manuscripta.ch](http://e-manuscripta.ch)
  - Project duration: September 2016 to November 2017
  - Cooperative project, lead: Zentralbibliothek Zürich
  - Funded by swissuniversities

- **Functionalities**
  - Extension of existing Visual Library online editor used by [Franckesche Stiftungen](http://franckesche-stiftungen.ch)
  - Workflow: registration, transcription, publication
  - Open to the public – simple transcription

---

Verehrter Herr Professor und verehrte Frau Professor Stein!

Vor zwei Tagen wurde ich (hallelujah!) an das Polytechnikum in Zürich berufen und habe hier schon meinen 1. k. Abschied angemeldet. Darob bei uns Alten und beiden Bürchen grosse Freude, sodass ich nicht unhin kann, es Ihnen mitzuteilen. Im Sommer wird schon gezügelt.

Beste Grüße an Sie und Ihre Kinder

Aufhohes Wiedersehen

Ihr A. Einstein u. Frau

---

1. Herr Professor Alfred Stem (1846-1936). Historiker
Multichannel distribution of metadata and content: Go where the users are

- **Spreading metadata**
  - locally: ETH Library’s [Knowledge Portal](#) also includes data from archival information systems (e.g. Image Archive, ETH Zurich Archives) and other sources
  - internationally: Finding aids of ETH Zurich Archives harvested by [Archives Portal Europe](#) (250 mio. archival units from serveral hundred archives)

- **Spreading content**
  - Ongoing upload of digitized images to [Wikimedia Commons](#) (~ 40,000 images so far)
  - Images used to illustrate Wikipedia articles (e.g. «[Mount Kilimanjaro](#)» article with ~ 100,000 visits per month)
  - Specialized metrics tool in development (Wikimedia CH)
Presenting content in a new way: Explora – the new storytelling platform

- Presenting content in story form
  - Research combined with storytelling
  - Visually attractive, using multimedia content
  - Change of perspective
- E.g. scientific plant images in “Mallow, Iris, Orchid”
  - Link the past (historical holdings) and the present (modern research, teaching, scientific visualization)
  - From Renaissance herbal books to 3D-modelling
  - Combining material from several collections and archives and
  - existing scans with new video testimonials
- New stories are published regularly
## Summary and outlook

<table>
<thead>
<tr>
<th>Potential</th>
<th>Summary</th>
<th>Outlook</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research partnerships</td>
<td>Solid basis:</td>
<td>• Further standardization regarding bulk access (e.g. RDF, IIIF)</td>
</tr>
<tr>
<td></td>
<td>• Existing platforms with their underlying processes (indexing, scanning,</td>
<td>• Growing demand for data driven research</td>
</tr>
<tr>
<td></td>
<td>presentation, archiving)</td>
<td>• Individual projects demand individual know-how</td>
</tr>
<tr>
<td></td>
<td>• Open data policy with clearly labelled rights statements</td>
<td>• More active role of the library</td>
</tr>
<tr>
<td>User participation</td>
<td>Crowdsourcing works, but ongoing community management is key</td>
<td>Use «collective intelligence» for further and more complex tasks (e.g.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to georeference images)</td>
</tr>
<tr>
<td>Outreach</td>
<td>• Targeted spreading of metadata and content increases visability</td>
<td>• Integration of data into link hubs (e.g. Wikidata)</td>
</tr>
<tr>
<td></td>
<td>• Visually attractive platforms reach out to new users</td>
<td>• Intensify cooperation with experts to produce content</td>
</tr>
</tbody>
</table>
Thank you very much for your attention!

Michael Gasser
Head Archives, ETH Library
Rämistrasse 101, CH-8092 Zürich
Tel. +41 44 632 21 82
michael.gasser@library.ethz.ch

You can find us on:
- [www.library.ethz.ch/en](http://www.library.ethz.ch/en)
- Facebook
- Twitter
- Google Plus