



Post digitization: Challenges in managing a digital collection

Jagiellonian Library , 25th of January 2013

Jasper Faase
Project Manager Innovative Projects
Email: jasper.faase@kb.nl

Mission

- The Koninklijke Bibliotheek is the national library of the Netherlands: we bring people and information together
- Our aim is to:
 - offer everyone everywhere access to everything published in the Netherlands
 - play a central role in the (scientific) information infrastructure of the Netherlands
 - promote permanent access to digital information nationally and internationally

How do we translate these words into practice?

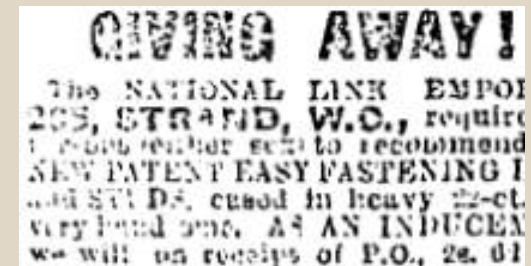
- Mass-digitization:
 - between 2009 and 2013 10% of all Dutch publications will be digitized
 - public-private partnerships (Google, ProQuest)
 - standardized workflows for digitizing of Books, Magazines and Newspapers
 - preservation by means of digitization
 - designating digitization part of the library's primary process

Results so far

Project	Timeline	Pages
Dutch Parliamentary Newspapers	2004-2010	2.500.000
Dutch Daily Newspapers	2007-2012	9.000.000
Early Dutch Books Online	2008-2010	2.000.000
Magazines	2009-2011	1.500.000
Google	2011-2014	30.000.000
ProQuest	2011-2016	6.000.000
Metamorfoze (Books, Newspapers and Magazines)	2012-2013	9.000.000
Totaal		60.000.000

Post digitization challenges

1. Improving results of Optical Character Recognition
2. The costs of running two libraries
3. Setting up a viable model for cooperation at national level



Challenge 1: Improving results of OCR

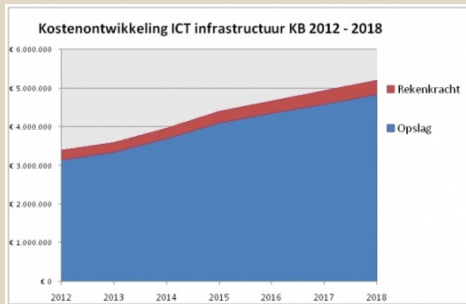
- **Problem:**
 - Target groups, especially linguists, value high quality OCR but poor recognition rates of software means pre 1850 OCR can often not be used for research
- **Approach:**
 - Use results of the European Project IMPACT (<http://www.impact-project.eu/>) within KB's digitization workflow (2012)
 - Introduce crowdsourcing to improve OCR of individual texts to a high degree of accuracy (2013)
 - Expectation management

Challenge 2: the costs of running two libraries



Challenge 2: the costs of running two libraries

- Problem:



- Approach:
 - Long-term storage planning & rethinking of storage policies
 - Reduce costs of the physical library
 - Creating a digital mindset: shift of focus from physical collection to the digital collection

Challenge 3: Cooperation

- **Problem:**
 - Fragmentation caused by lack of coordination of digitization efforts
 - Lack of standardization and different interpretations of metadata standards
- **Approach:**
 - Centralized mass-digitization financed by Metamorfoze (National Programme for Preservation of Cultural Heritage)
 - Create ‘national infrastructure’ to provide permanent access to digitized heritage collections in partnership with Dutch academic libraries
 - Promote Open Access and use of dataservices (via OAI-PMH/SRU)



Thanks for your attention
Jasper.Faase@kb.nl