Enhancing Digitool with a new METS/ALTO viewer

Ghent 1.9.2010
www.eluxemburgensia.lu

- **Historic Newspapers**
  - 120,000 pages digitized, more will follow

- **Data format: METS**
  - Container format for TIFF, ALTO, PDF of pages
  - Layout information of articles, used to build a table of contents

- **Data format: ALTO**
  - OCR of full text
  - Coordinates of individual words

- **Went live in May 2009**
  - With Digitool’s out of the box METS/ALTO viewer
  - Since then: feedback of users about shortcomings
Changes requested

- **From user feedback**
  - More intuitive interface needed!
  - Make navigation as easy as on comparable Internet sites
  - Need translation into French + German
  - Table of contents + navigation by page at the same time
  - Directly jump to a certain page

- **Ideas from library staff**
  - Direct link to page / article
  - Zoom / drag using the mouse
  - Navigation between issues
  - Keyboard shortcuts for reading articles

- **Additional ideas during development**
  - On-the-fly translation of OCR text
Live demo of viewer

http://www.eluxemburgensia.lu
Development

- **Development by National Library**
  - 7 months, 1 programmer from project start to finish
  - Java, Javascript, ExtJS, C++, Oracle

- **Analysis phase**
  - Choice to use Digitool Webservices
  - Integration into Digitool viewer architecture
  - Need to develop own search (oracle text + alto parser)
  - Difficulty passing additional parameters to viewer

- **Programming phase**
  - Server side caching (images, structural data)
  - 6 different webservices (search, issue, article, page, calendar, toc)
  - 2 local C++ programs (search, image conversion)
Future

- Will be released under open source license
  - At the end of October probably

- Downloadable on El Commons

- Future enhancements:
  - Display license status for each object
  - Support for mobile devices (iPad etc.)
  - Adaptation to display digitized books
Questions?
www.eluxemburgensia.lu

Yves.maurer@bnl.etat.lu