From DigiTool to Aleph and Back Again:

Metadata Synchronization

Matthias Gross
Bavarian State Library (Munich, Germany)
Overview

• The Bavarian Library Network: Who we are
• Metadata Synchronization: ... What?
• ... Why?
• ... How?
• Regional library network for Bavaria, Germany
• Head Office is department of Bavarian State Library
• Network Catalog (Aleph 500):
  - cooperative cataloging,
  - copy cataloging from common data pool,
  - authority files (persons, corporate bodies, subject headings)
• Local library systems get bibliographic records automatically (Z39.50-based; new records, updates, deletions)
The Bavarian Library Network

- 2005: Decision for DigiTool as a common repository
- important: connection from catalog to repository (end user navigation; making digitization visible)
- DigiTool installation will be open to all network libraries (reduces administrative overhead)
- since 2006: implementation, pilot projects with Bavarian State Library (project for long-term preservation of digital objects) and University Library Regensburg
Aleph

DigiTool (from a cataloger’s point of view)

(Images from Wikipedia)
What?
(slightly more technical)

Aleph

Catalog record

001 BV1234567
100 Baggins, Bilbo
245 Concerning Hobbits

*Link to object delivery*
*URN*

DigiTool

Object

PID 4711

MID 9876

001 BV1234567

*Metadata update*

Preservation MD

URN
What?

End user perspective
Why?

- using existing bibliographic records = saving of labor (digitization projects)
- Aleph is and will at least for some time stay the primary cataloging environment (authority files). If „low quality“ is sufficient, it is also possible to have metadata only in DigiTool.
- updates of bib. records are replicated to DigiTool
- Access to digital objects from the catalog (on multiple levels: network / library / worldwide)
How?
(even more technical)

Aleph (V. 16)

Catalog record
SYS
001 BV1234567
100 Baggins, Bilbo
245 Concerning Hobbits

URN

z403 record
SYS
label
link to object delivery

Publishing
ue_21

OAI data provider

DigiTool (V. 3 SP 16)

Object (modified)
PID 4711

MID 9876
001 BV1234567
Metadata update

Preservation MD
URN

Harvesting
ue_13

Replication

OAI data provider

Synchronization
<table>
<thead>
<tr>
<th>Terminology</th>
<th>DigiTool</th>
<th>Aleph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Updating OAI data provider</td>
<td>Replication</td>
<td>Publishing ue_21</td>
</tr>
<tr>
<td>(Repository)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OAI service provider</td>
<td>Synchronization</td>
<td>Harvesting ue_13</td>
</tr>
</tbody>
</table>
tempus fugit

OAI Service Provider (Harvester)

correct

wrong

recall-period / scope

OAI Data Provider (Repository)
Aleph side:

- recall-period 20 minutes, wait time 10 minutes (applies when the last request produced no results – otherwise next request immediately → adaptive harvesting strategy)
- after error situation: recall-period 90 minutes, wait time 10 sec. („catch-up“)

Important: choose recall-period > wait time (to avoid degeneration)

- Publishing is triggered by indexing process – otherwise constraints for publishing can’t be evaluated!
DigiTool side:

• Synchronization: every 10 minutes (not adaptive), recall-period 4 hours (relic, but still works)
  This interval mustn’t be too short, otherwise the former synchronization process may not have finished
• Replication: every 10 minutes
• change of parameters → jboss restart!
Exciting things 1

1. Deleting an object in DigiTool
   - deletion is communicated to Aleph
   - z403 record is deleted
   - but URN remains in bib. record

2. Changing the label in DigiTool
   - is communicated to, but not processed by Aleph
   - Aleph-Fix announced
3. Linking 1 bib. record with a second object
   - when Aleph is going to publish the bib. record, it may note that there was no change to it (new link is no part of publishing format) and ignore it
   - Aleph-Fix announced
   - Recommendation: use shared MD record in DT

4. Linking 1 object with 2 bib. records
   - not possible due to uniqueness of PID in Aleph index z403_id1 (which avoids infinite DTL-Aleph-Sync loop...)
   - whenever you want to do such a thing: rethink your data model...
5. correcting wrong 001 number in DT
   - Metadata update may succeed (cf. 3.)
   - Aleph notes that there is already a z403 record for this PID and is satisfied...
   - ...but you are probably not. You can
     -- remove „wrong“ z403 record (ADAM-enabled version of Aleph GUI client)
     before correction
     -- delete and recreate object in DT
1. When you have digital objects like tables of contents, reviews, errata etc. which can be used to enrich the catalog
   • you want to have a link from the bib. record
   • but when these objects have an URN in DigiTool, it is transported into the bib. record → this should only happen for „full“ representations of an item
   • DigiTool stores URN with no relation to MD records
   • so the short-term „solution“ is: no URNs (:-[)

2. It has been quite a long quest: implementation, test, frustration, re-implementation... but now it works for most scenarios (:-})
Thank you very much for your attention!

The road goes ever on and on...

matthias.gross@bsb-muenchen.de