supporting the digitization chain
UKB (Univerzitná knižnica v Bratislave)

Jasná, 23.09.2010
Tomáš Fiala, Stephan Tratter
Agenda

1. Why a workflow?
   The main problems & questions
1. Workflow preparations
2. ScanFlow™ parts
3. Just an example
4. Résumé
workflow

basic questions
ScanFlow™: handles the resources

- Coordination - staff
- Scan- & server-hardware
- Books - docs - logistic
- Files: images, meta-data
- Quality-Control

- Workflow-software

XML
METS
MODS
MARC
Dublin-CORE
digitization Workflow – typical questions

- Where are the books?
- Where are the images?
- Where are the backups?
- Which machine is free?
- Which jobs are at work?
- When are these books returned?
- How much time took the project?
Workflow - What is the principal aim?

- Digital object of a page/book
  - images with description files
  - meta data
    - bibliographic meta data
      - Titel, Author, Editorial house etc.
    - technical meta data
      - size, color depth, format etc.
    - content meta data
      - text, description, structure, layout, etc.

- Optimizing the use of the resources
- Overview over running projects
- Reports
ScanFlow™: Resource Management

- Main control station
- Team
- Reporting / statistic
- Archiving / publishing
- Scanning equipment
- OCR & correction desk
- QA / quality check stations
- Inspection desks
- Distributed logistic
- PDF / reprint / facsimile

Workflow-software: ScanFlow™
ScanFlow™: Ressource Management

flow-diagram representing the physical ScanCenter

divides the work in
- small tasks applied on the ScanJob
- easy decisions

What means „using a workflow“?
workflow

preparation
planning
design
Organizing resources I

- **tasks**
  - manual tasks
    - book logistic
    - scanning
    - quality check
  - automatic tasks
    - copy functions with file integrity check
    - ocr
    - cropping, deskewing etc.
    - job merging
    - repository injection
Organizing resources II

combining:
• hardware
• software
• workplaces
• infrastructure
Organizing resources III

the „staff-group-task“ table

<table>
<thead>
<tr>
<th>staff: person</th>
<th>groups</th>
<th>tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 P2 P3 P4 P5 P6 P7 P8 P9 P10</td>
<td>T1 T2 T3 T4 T5 T6 T7 T8 T9 T10</td>
<td></td>
</tr>
<tr>
<td>x x x</td>
<td>G1</td>
<td>x</td>
</tr>
<tr>
<td>x x x</td>
<td>G2</td>
<td>x</td>
</tr>
<tr>
<td>x x</td>
<td>G3</td>
<td>x</td>
</tr>
<tr>
<td>x x</td>
<td>G4</td>
<td>x</td>
</tr>
<tr>
<td>x x x x x</td>
<td>G5</td>
<td>x</td>
</tr>
<tr>
<td>x x x</td>
<td>G6</td>
<td>x</td>
</tr>
<tr>
<td>x x x</td>
<td>G7</td>
<td>x</td>
</tr>
<tr>
<td>x x x</td>
<td>G8</td>
<td>x</td>
</tr>
<tr>
<td>x x</td>
<td>G9</td>
<td>x</td>
</tr>
<tr>
<td>x x</td>
<td>G10</td>
<td>x</td>
</tr>
</tbody>
</table>

**e.g.:**
P4 = Project manager
T5 = scanning task
G5 = ScanOperator-group
designing the workflow

- flow diagram
- NOT LINEAR!
- ADVANTAGE → possibility to go back!
workflow

building blocks
ScanFlow™ – Workflow parts

- Designed workflows / dataflows
- Interfaces to services
  - cataloge, ocr, repository injection etc.
- Analysis
  - statistics
    - books, jobs, pages
  - tracking
    - books and jobs
  - monitoring
    - system tasks
    - user tasks
  - reporting
ScanFlow™ parts: workflow design

workflows / dataflows representing the digi-chain

- book supply
- scanning
  - 1. QC
  - pages preview
  - processing
  - 2. QC
  - prepare OCR
  - processing
  - 3. QC
- book release
  - copy & upload
  - inject repository

user tasks
system tasks
WF4: More workflows parallel
ScanFlow™: Statistic Analysis

- statistics
  - books, jobs, pages

Custom Filter

<table>
<thead>
<tr>
<th>All workflows cumulated</th>
<th>Single workflows</th>
</tr>
</thead>
<tbody>
<tr>
<td>current Progress of Tasks within Workflows</td>
<td>BF_4</td>
</tr>
<tr>
<td>Tasks Current state</td>
<td>DJA_1V0</td>
</tr>
<tr>
<td>Digi Service - Books prepared</td>
<td>Gaswerke_1V0</td>
</tr>
<tr>
<td>Digi Service - Books free for Check Out</td>
<td>TrvDigiDienst</td>
</tr>
<tr>
<td>Digi Service - Books in Process</td>
<td>BF_4</td>
</tr>
<tr>
<td>Digi Service - Books returned</td>
<td>DJA_1V0</td>
</tr>
<tr>
<td>Digi Service - Open tasks</td>
<td>Gaswerke_1V0</td>
</tr>
<tr>
<td>Digi Service - Pages delivered</td>
<td>TrvDigiDienst</td>
</tr>
<tr>
<td>BF4 - Pages delivered</td>
<td>BF_4</td>
</tr>
<tr>
<td>Job &lt;-&gt; Workflow Comparsion</td>
<td>DJA_1V0</td>
</tr>
<tr>
<td>Digi Service - Jobs completed (old)</td>
<td>Gaswerke_1V0</td>
</tr>
<tr>
<td>1Tag</td>
<td>TrvDigiDienst</td>
</tr>
</tbody>
</table>

TREVENTUS
ScanFlow™: Statistic Analysis

- Tasks listed in a pie chart
- to understand the bottlenecks & to see possible optimizations

© Treventus Mechatronics: project GW2009
ScanFlow™ parts: tracking

- tracking
  - books and jobs

Job Tracking

[Extended Tracking]

Track
ScanFlow™ parts: monitoring

- monitoring: system & user tasks

**SCANFLOW™
stay on Track!**

<table>
<thead>
<tr>
<th>System</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC-Name</strong></td>
<td><strong>modified</strong></td>
</tr>
<tr>
<td>ethene</td>
<td>2010-05-15 10:36:49</td>
</tr>
<tr>
<td>canoni</td>
<td>2010-09-15 10:13:00</td>
</tr>
<tr>
<td>freisa</td>
<td>2010-09-13 15:13:45</td>
</tr>
<tr>
<td>ikarus</td>
<td>2010-09-15 15:20:49</td>
</tr>
<tr>
<td>pan</td>
<td>2010-06-24 14:30:16</td>
</tr>
<tr>
<td>win</td>
<td>2010-06-23 14:58:43</td>
</tr>
<tr>
<td>theania</td>
<td>2010-05-27 10:05:42</td>
</tr>
<tr>
<td>win7test-hp</td>
<td>2010-09-29 08:45:55</td>
</tr>
</tbody>
</table>

**Legend**
- modification less or equal 30 minutes
- modification greater than 30 minutes or less or equal 1 day
- modification older than 1 day
ScanFlow™ parts: reporting

- reporting
  - job reports
  - monthly, annualy reports

Job Search

Search Results for jobid with 'Gaswerke_110':

<table>
<thead>
<tr>
<th>#</th>
<th>jobid</th>
<th>bookid</th>
<th>taskname</th>
<th>start</th>
<th>end</th>
<th>duration</th>
<th>pagenumberabs</th>
<th>executor</th>
<th>compname</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gaswerke_110</td>
<td>ScanningSR30x</td>
<td>2010-01-19 14:30:49</td>
<td>2010-01-19 14:57:45</td>
<td>0h 21m 0s</td>
<td>222</td>
<td>athene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Gaswerke_110</td>
<td>CreateJpegForQC</td>
<td>2010-01-19 22:24:55</td>
<td>2010-01-19 22:34:11</td>
<td>0h 9m 15s</td>
<td>222</td>
<td>SuperScanGate</td>
<td>athene</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Gaswerke_110</td>
<td>NamingAndQualityCheck</td>
<td>2010-01-24 16:08:58</td>
<td>2010-01-24 16:17:05</td>
<td>0h 7m 0s</td>
<td>222</td>
<td>hestia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Gaswerke_110</td>
<td>CheckCropRectAndDeskewLine</td>
<td>2010-01-26 12:09:05</td>
<td>2010-01-26 12:14:45</td>
<td>0h 4m 30s</td>
<td>222</td>
<td>SR00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Gaswerke_110</td>
<td>CreateFallbackPoint1</td>
<td>2010-01-28 01:44:23</td>
<td>2010-01-28 01:48:34</td>
<td>0h 4m 11s</td>
<td>222</td>
<td>SuperScanGate</td>
<td>SR00</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Gaswerke_110</td>
<td>CropAndDeskew</td>
<td>2010-01-28 05:16:10</td>
<td>2010-01-28 05:21:33</td>
<td>0h 5m 22s</td>
<td>222</td>
<td>SuperScanGate</td>
<td>SR00</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Gaswerke_110</td>
<td>QualityCheck</td>
<td>2010-02-03 10:40:28</td>
<td>2010-02-03 10:47:59</td>
<td>0h 7m 30s</td>
<td>222</td>
<td>pann</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Gaswerke_110</td>
<td>OCR</td>
<td>2010-02-07 16:54:06</td>
<td>2010-02-07 18:09:49</td>
<td>0h 12m 0s</td>
<td>221</td>
<td>Christoph Bacher</td>
<td>KARUS</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Gaswerke_110</td>
<td>CopyTiffsToCustomerDisk</td>
<td>2010-02-07 22:00:04</td>
<td>2010-02-07 22:04:00</td>
<td>0h 3m 56s</td>
<td>221</td>
<td>SuperScanGate</td>
<td>SR00</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Gaswerke_110</td>
<td>CopyOcrResultToCustomerDisk</td>
<td>2010-02-07 22:07:09</td>
<td>2010-02-07 22:07:27</td>
<td>0h 0m 18s</td>
<td>221</td>
<td>SuperScanGate</td>
<td>SR00</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Gaswerke_110</td>
<td>UploadToInternalBackup</td>
<td>2010-02-07 22:07:55</td>
<td>2010-02-07 22:12:55</td>
<td>0h 5m 0s</td>
<td>221</td>
<td>SuperScanGate</td>
<td>SR00</td>
<td></td>
</tr>
</tbody>
</table>
workflow

just an example
Example 1

1. QC scanning
   - pages preview
   - processing
   - 1. QC

2. QC prepare OCR
   - processing
   - prepare OCR

3. QC copy & upload inject repository
   - book release
Example 1

- new job
Example 1

- barcode
Example 1

- assign barcode (job) \(\rightarrow\) book
Example 1

ScanOperator accepts his task
Example 1

- book logistic output (ready for return)
- book logistic return (who did return the book?)
ScanFlow™: Statistic Analysis I
ScanFlow™: Statistic Analysis I

TASK: Quality Check 1
ScanFlow™: Statistic Analysis I

TASK: renaming the pages
ScanFlow™: Statistic Analysis I

TASK: Quality Check 2
ScanFlow™: Statistic Analysis I

TASK: check crop & deskew
ScanFlow™: Statistic Analysis I

**TASK:** create fallback point
ScanFlow™: Statistic Analysis I

TASK: adjust crop & deskew
ScanFlow™: Statistic Analysis I

TASK: Quality Check 3
ScanFlow™: Statistic Analysis I
ScanFlow™: Statistic Analysis I
ScanFlow™: Statistic Analysis I

TASK: server tasks (i.e. copy)
workflow

résumé
A good workflow

- What makes a workflow a good workflow?

example → lunch
ScanFlow™: Example

Operator 1 login
List of all jobs
1 job selected
deskew the files
release the job
1 job finished

OCR-Server

Operator 2 login
List of all jobs
1 job selected
deskew the files
release the job
1 job finished
**Conclusio: What means “workflow”?**

The workflow-software:

- organizing & coordinating
- monitoring & overview

→ to handle the resources of a scan-center
...the tip of the iceberg
Thanks to
• our reference customer UKB for support and input
• our partner on stage, Alena kulikova (Microform)