ScanFlow™

Effective workflow for book digitization

Vienna, 04.02.2010

by

Christoph Bacher & Stephan Tratter
What are we doing?

• automatic book scanner: ScanRobot®
• scanning software: ScanGate™
  (scanning, image treatment & OCR)
• workflow software: ScanFlow™
  (automized ressource management)
• digitization service: ScanService
  (best practice inhouse)
ScanRobot® SR301 → simple solution

- opening angle: 60°
- gentle air flows for page turning
- minimum illumination!!!
- only 1 moving part!!!
- results in realtime onscreen → WYSIWYG (what you see is what you get)
- NO image distorsions
- NO clambs in the result images!!!
- NO permanent calibration needed
ScanGate™

- image treatment (deskewing, cropping, converting, transforming, etc.)
- batch processing
- quality check in realtime

- easy interface!
- multilanguage (de, cz, ru, ro, pl, it, en)

Utorok, 4. máj 2010
ScanRobot® - opening angle

Öffnungswinkel: 60°

Öffnungswinkel: 180°
ScanRobot® - opening angle

Öffnungswinkel: 60°

Öffnungswinkel: 180°
opening angle: $90^\circ$ vs. $60^\circ$
opening angle: 90° vs. 60°
opening angle: 90° vs. 60°
opening angle: 90° vs. 60°

missing text in the binding
opening angle: 90° vs. 60°
opening angle: 90° vs. 60°
ScanRobot® - the new book cradle

**Opening angle:** 60 degree (→ Treventus book cradle)
ScanRobot® - Europa

31 ScanRobot® sold world wide
(→ reliable technology | last update: 01.12.2009)
Treventus digitization service

best practice in house

Utorok, 4. máj 2010
DS: **Digitization Service**

- logistics
- scanning books & newspapers
- file handling
- post processing
  - renaming
  - cropping
  - quality check
  - deskew, etc.
- text recognition (OCR)
ScanFlow™

a brief overview
Workflow - What is the principal aim?
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.
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.

• Tasks & resources management
  – tasks → groups (1:n)
  – groups → operators (n:m)
  – storage (central & distributed)
ScanFlow™ – handels the ressources

coordination - staff

books - docs - logistic

scan- & server-hardware

Quality-Control

files: images, meta-data

XML, METS, MODS, MARC, Dublin-CORE

workflow-software

ScanFlow™

Utorok, 4. máj 2010
The heart of **ScanFlow™**?

- **flow diagram**
  - organized work
  - structure of the environment
  - tasks and decisions
- **statistic**
  - understanding and optimization
- **monitoring**
  - Is all ready, up and running?
ScanFlow™: Ressource Management

main control station

scanning equipment

OCR & correction desk

QA / quality check stations

inspection desks

PDF / reprint / faksimile

workflow-software ScanFlow™

reporting / statistic

Archiving / publishing

distributed logistic

team
ScanFlow™: organizing resources - strategy

archivation server
rep.ulib.sk

doc. deliv. server
mvs.ulib.sk

workflow server
TREVENTUS

OCR server
TREVENTUS

archive DVD service
TREVENTUS

administrator
UKB

work storage

Utorok, 4. máj 2010
ScanFlow™: organizing resources - strategy
ScanFlow™: Management of resources

Workflow - Software

- Head of digi-center
- process visualization
- IMPORT files/images
- Interface OPAC, ALEPH, etc.
- OCR-Server
- long time preservation
- Repository
- meta data
- files
- monitoring
- files
- meta data
- OCR-Server
- LTP
- Interface
- OPAC, ALEPH, etc.
- IMPORT files/images
- monitoring
- Head of digi-center
- Workflow - Software
- Long time preservation
- Interface
- OPAC, ALEPH, etc.
- IMPORT files/images
- monitoring
- Head of digi-center
- Workflow - Software
ScanFlow™: Management of resources
ScanFlow™: Management of resources

WORKFLOW
Head of
digi-center

process visualization

monitoring

flow-diagram
representing the
physical ScanCenter

OCR-Server

meta data

LTP

long time

preservation

Interface

OPAC, ALEPH, etc.

SOFTWARE

Workflow - Software

Import

files/images

file-diagram

head of
digi-center

Utorok, 4. máj 2010
ScanFlow™: Management of resources

flow-diagram representing the physical ScanCenter
**ScanFlow™: Management of resources**

**WORKFLOW**
Head of digi-center

**monitoring**

**process visualization**

**SOFTWARE**

**OCR-Server**

**Repository**

**meta data**

**Reporting**

**LTP**
long time preservation

**Interface**
OPAC, ALEPH, etc.

**flow-diagram**
representing the physical ScanCenter

**WORKFLOW**

**Head of digi-center**

**monitoring**

**process visualization**

**SOFTWARE**

**OCR-Server**

**Repository**

**LTP**
long time preservation

**Interface**
OPAC, ALEPH, etc.

Utorok, 4. máj 2010
ScanFlow™: Management of resources

WORKFLOW
Head of digi-center

process visualization

monitoring

Software

flow-diagram representing the physical ScanCenter

import files/images

flow-diagram representing the physical ScanCenter

divides the work in

• small tasks applied on the ScanJob
• easy decisions

Utorok, 4. máj 2010
ScanFlow™: Management of resources

- Head of digi-center
- Process visualization
- Monitoring
- Workflow
- OCR-Server
- Reporting
- LTP
- Long time preservation
- Interface
- OPAC, ALEPH, etc.

**Flow-diagram**

- Head of digi-center
- Workflow
- OCR-Server
- Reporting
- LTP
- Long time preservation
- Interface
- OPAC, ALEPH, etc.

**Software**

- **Task**

**Monitoring**

- Divides the work in
  - Small tasks applied on the ScanJob
  - Easy decisions

Utorok, 4. máj 2010
ScanFlow™: Management of resources

flow-diagram representing the physical ScanCenter

divides the work in

- small tasks applied on the ScanJob
- easy decisions
ScanFlow™: Statistic Analysis

- Tasks listed in a pie chart
- to understand the **bottlenecks** & to see possible **optimizations**
ScanFlow™: Statistic Analysis I

TASK: Scanning
ScanFlow™: Statistic Analysis I

TASK: Quality Check 1

Utorok, 4. máj 2010
ScanFlow™: Statistic Analysis I

TASK: renaming the pages
ScanFlow™: Statistic Analysis I

TASK: Quality Check 2
ScanFlow™: Statistic Analysis I

TASK: check crop & deskew
**Task:** create fallback point
ScanFlow™: Statistic Analysis I

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

TASK: Quality Check 3

Utorok, 4. máj 2010
ScanFlow™: Statistic Analysis I

TASK: OCR or recover FBP1?
ScanFlow™: Statistic Analysis I

TASK: OCR is running... (auto)
ScanFlow™: Statistic Analysis I

TASK: OCR Quality Check

Utorok, 4. máj 2010
scanFlow™: Statistic Analysis I

 TASK: server tasks (i.e. copy)
ScanFlow™: Job administration

- Job-Management
- **ITM**: Inter-Task message (virtual POST-IT)
- Priority
- Task-List
- Every station uses the same software and the same interface
ScanFlow™: Job administration

- Job-Management
- **ITM**: Inter-Task message (virtual POST-IT)
- Priority
- Task-List
- Every station uses the same software and the same interface

**ITM** (inter-task message)

i.e.:
- all pages already deskewd
- please insert book cover
- folded map A3 IMPORT

Utorok, 4. máj 2010
ScanFlow™: Example
ScanFlow™: Example

Operator 1 login

ScanFlow™

OCR-Server

......
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example

OCR-Server

List of all jobs
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example

OCR-Server

1 job selected
ScanFlow™: Example

1 job selected
deskew the files

OCR-Server
ScanFlow™: Example

OCR-Server

......

1 job selected

ScanFlow™
ScanFlow™: Example

OCR-Server

1 job selected
ScanFlow™: Example

1 job selected
release the job

OCR-Server
ScanFlow™: Example
ScanFlow™: Example

1 job selected

OCR-Server

ScanFlow™

Utorok, 4. máj 2010
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example

OCR-Server

List of all jobs
ScanFlow™: Example
ScanFlow™: Example
ScanFlow™: Example

OCR-Server

1 job selected
ScanFlow™: Example

OCR-Server

Utorok, 4. máj 2010
ScanFlow™: Example

1 job selected
deskew the files
ScanFlow™: Example

1 job selected
ScanFlow™: Example

OCR-Server

1 job selected

Utorok, 4. máj 2010
ScanFlow™: Example

OCR-Server

1 job selected
release the job

Utorok, 4. máj 2010
ScanFlow™: Example

OCR-Server

1 job selected
ScanFlow™: Example
ScanFlow™: Example

OCR-Server

1 job finished

Utorok, 4. máj 2010
ScanFlow™: Example
ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>ScanRobot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode</td>
<td></td>
</tr>
<tr>
<td>Metadata import</td>
<td></td>
</tr>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
</tr>
<tr>
<td>Operator 3</td>
<td>Import Files</td>
</tr>
<tr>
<td>Operator 4</td>
<td>crop &amp; deskew</td>
</tr>
<tr>
<td>[Automatic]</td>
<td>MERGE B &amp; C</td>
</tr>
<tr>
<td>[Automatic]</td>
<td>OCR-Server</td>
</tr>
<tr>
<td>Operator 5</td>
<td>Release Jobs</td>
</tr>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
</tr>
<tr>
<td>[Automatic]</td>
<td>copy2archive</td>
</tr>
</tbody>
</table>

ScanFlow™

- organizes parallel tasks
- storage request *(distributed & centralized)* and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™**: Jobs in the timetable

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>ScanRobot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode</td>
<td>Metadata import</td>
</tr>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
</tr>
<tr>
<td>Operator 3</td>
<td>Import Files</td>
</tr>
<tr>
<td>Operator 4</td>
<td>crop &amp; deskew</td>
</tr>
<tr>
<td>[Automatic]</td>
<td>MERGE B &amp; C</td>
</tr>
<tr>
<td>[Automatic]</td>
<td>OCR-Server</td>
</tr>
<tr>
<td>Operator 5</td>
<td>Release Jobs</td>
</tr>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
</tr>
<tr>
<td>[Automatic]</td>
<td>copy2archive</td>
</tr>
</tbody>
</table>

**ScanFlow™**

- organizes parallel tasks
- storage request *(distributed & centralized)* and secure copy of digitized jobs

---

Utorok, 4. máj 2010
**ScanFlow™: Jobs in the timetable**

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>ScanRobot</th>
<th>book 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode</td>
<td>Metadata import</td>
<td>book 1</td>
</tr>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
<td></td>
</tr>
<tr>
<td>Operator 3</td>
<td>Import Files</td>
<td></td>
</tr>
<tr>
<td>Operator 4</td>
<td>crop &amp; deskew</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ Automatic ] MERGE B &amp; C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ Automatic ] OCR-Server</td>
</tr>
<tr>
<td>Operator 5</td>
<td>Release Jobs</td>
<td></td>
</tr>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
<td>MF 01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ Automatic ] copy2archive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ScanFlow™**

- organizes parallel tasks
- storage request (**distributed & centralized**) and secure copy of digitized jobs

Utorok, 4. máj 2010
## ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th>Operator</th>
<th>Task</th>
<th>Book(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ScanRobot</td>
<td>book 1</td>
</tr>
<tr>
<td></td>
<td>Barcode</td>
<td>book 1</td>
</tr>
<tr>
<td></td>
<td>Metadata import</td>
<td>book 1</td>
</tr>
<tr>
<td>2</td>
<td>Bookeye</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Import Files</td>
<td>book 1</td>
</tr>
<tr>
<td>4</td>
<td>crop &amp; deskew</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Automatic] MERGE B &amp; C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Automatic] OCR-Server</td>
<td>book 1</td>
</tr>
<tr>
<td>5</td>
<td>Release Jobs</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Microfilm</td>
<td>MF 01</td>
</tr>
<tr>
<td></td>
<td>[Automatic] copy2archive</td>
<td>MF 02</td>
</tr>
</tbody>
</table>

**ScanFlow™**

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™**: Jobs in the timetable

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>ScanRobot</th>
<th>book 1</th>
<th>book 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode Metadata import</td>
<td>book 1</td>
<td>book 2</td>
<td></td>
</tr>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
<td>book 1</td>
<td></td>
</tr>
<tr>
<td>Operator 3</td>
<td>Import Files</td>
<td>book 1</td>
<td></td>
</tr>
<tr>
<td>Operator 4</td>
<td>Crop &amp; deskew</td>
<td>book 1</td>
<td></td>
</tr>
<tr>
<td>[Automatic] MERGE B &amp; C</td>
<td>book 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Automatic] OCR-Server</td>
<td>book 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 5</td>
<td>Release Jobs</td>
<td>book 1</td>
<td></td>
</tr>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
<td>MF 01</td>
<td>MF 02</td>
</tr>
<tr>
<td>[Automatic] copy2archive</td>
<td>book 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ScanFlow™**

- organizes parallel tasks
- storage request *(distributed & centralized)* and secure copy of digitized jobs

Utorok, 4. máj 2010
### ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>ScanRobot</th>
<th>book 1</th>
<th>book 2</th>
<th>book 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 3</td>
<td>Import Files</td>
<td></td>
<td>book 1</td>
<td>book 2</td>
</tr>
<tr>
<td>Operator 4</td>
<td>crop &amp; deskew</td>
<td></td>
<td></td>
<td>book 1</td>
</tr>
<tr>
<td>MERGE B &amp; C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCR-Server</td>
<td></td>
<td>book 1</td>
<td>book 2</td>
<td></td>
</tr>
<tr>
<td>Operator 5</td>
<td>Release Jobs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
<td>MF 01</td>
<td>MF 02</td>
<td>MF 03</td>
</tr>
<tr>
<td>copy2archive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **ScanFlow™**
  - organizes parallel tasks
  - storage request *(distributed & centralized)* and secure copy of digitized jobs
### ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th>Operator 1 ScanRobot</th>
<th>book 1</th>
<th>book 2</th>
<th>book 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator 2 Bookeye</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 3 Import Files</td>
<td>book I1</td>
<td>book I2</td>
<td></td>
</tr>
<tr>
<td>Operator 4 crop &amp; deskew</td>
<td></td>
<td>book 1</td>
<td></td>
</tr>
<tr>
<td>[Automatic] MERGE B &amp; C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 5 Release Jobs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 6 Microfilm</td>
<td>MF 01</td>
<td>MF 02</td>
<td>MF 03</td>
</tr>
<tr>
<td>[Automatic] copy2archive</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ScanFlow™**
- organizes parallel tasks
- storage request *(distributed & centralized)* and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ScanRobot</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata import</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 2</th>
<th>book 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bookeye</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 3</th>
<th>book 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Import Files</td>
<td>book I1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 4</th>
<th>book 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>crop &amp; deskew</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 5</th>
<th>book 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Release Jobs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 6</th>
<th>MF 01</th>
<th>MF 02</th>
<th>MF 03</th>
<th>MF 04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microfilm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>copy2archive</td>
<td></td>
</tr>
</tbody>
</table>

ScanFlow™

- organizes parallel tasks
- storage request (**distributed & centralized**) and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™: Jobs in the timetable**

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>ScanRobot</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>book 1</strong></td>
<td><strong>book 2</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Barcode Metadata import</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>book 1</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 2</th>
<th>Bookeye</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>book 1</strong></td>
<td><strong>book 2</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 3</th>
<th>Import Files</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>book I1</strong></td>
<td><strong>book I2</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 4</th>
<th>crop &amp; deskew</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>book 1</strong></td>
<td><strong>book 2</strong></td>
</tr>
</tbody>
</table>

**[Automatic]**

<table>
<thead>
<tr>
<th>Operator 5</th>
<th>Release Jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>book 1</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator 6</th>
<th>Microfilm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MF 01</strong></td>
<td><strong>MF 02</strong></td>
</tr>
</tbody>
</table>

**[Automatic]**

<table>
<thead>
<tr>
<th>copy2archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>book 1</td>
</tr>
</tbody>
</table>

---

**ScanFlow™**

- organizes parallel tasks
- storage request (*distributed* & *centralized*) and secure copy of digitized jobs

---

Utorok, 4. máj 2010
**ScanFlow™**: Jobs in the timetable

- **Operator 1 (ScanRobot):**
  - book 1
  - book 2
  - book 3
  - book 4
  - book 5

- **Barcode Metadata import:**
  - book 1
  - book 2
  - book 3
  - book 4

- **Operator 2 (Bookeye):**
  - book 1
  - book 2
  - extra pages

- **Operator 3 (Import Files):**
  - book 1
  - book 2

- **Operator 4 (crop & deskew):**
  - book 1
  - book 2
  - book 3

- **[Automatic] MERGE B & C:**
  - book 1

- **[Automatic] OCR-Server:**
  - book 1
  - book 2
  - book 3

- **Operator 5 (Release Jobs):**
  - book 1

- **Operator 6 (Microfilm):**
  - MF 01
  - MF 02
  - MF 03
  - MF 04

- **[Automatic] copy2archive:**

---

**ScanFlow™**

- organizes parallel tasks
- storage request (**distributed & centralized**) and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
<td>MF 01</td>
<td>MF 02</td>
<td>MF 03</td>
<td>MF 04</td>
<td>MF 05</td>
<td></td>
</tr>
</tbody>
</table>

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™: Jobs in the timetable**

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>barcode Metadata import</th>
<th>Operator 2</th>
<th>Bookeye</th>
<th>Operator 3</th>
<th>Import Files</th>
<th>Operator 4</th>
<th>crop &amp; deskew</th>
<th>Operator 5</th>
<th>Release Jobs</th>
<th>Operator 6</th>
<th>Microfilm</th>
<th>Operator 7</th>
<th>copy2archive</th>
</tr>
</thead>
</table>

**ScanFlow™**

- organizes parallel tasks
- storage request *(distributed & centralized)* and secure copy of digitized jobs
ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Import Files</td>
<td></td>
<td>book l1</td>
<td>book l2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Automatic]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MERGE B &amp; C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Automatic]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCR-Server</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Release Jobs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microfilm</td>
<td></td>
<td>MF 01</td>
<td>MF 02</td>
<td>MF 03</td>
<td>MF 04</td>
<td>MF 05</td>
</tr>
<tr>
<td>[Automatic]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>copy2archive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ScanFlow™

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™: Jobs in the timetable**

<table>
<thead>
<tr>
<th>Operator 1</th>
<th>ScanRobot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode</td>
<td>book 1</td>
</tr>
<tr>
<td>Metadata import</td>
<td>book 2</td>
</tr>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
</tr>
<tr>
<td>Import Files</td>
<td>book 3</td>
</tr>
<tr>
<td>Operator 3</td>
<td>Import Files</td>
</tr>
<tr>
<td>Operator 4</td>
<td>crop &amp; deskew</td>
</tr>
<tr>
<td>[Automatic] MERGE B &amp; C</td>
<td></td>
</tr>
<tr>
<td>[Automatic] OCR-Server</td>
<td></td>
</tr>
<tr>
<td>Operator 5</td>
<td>Release Jobs</td>
</tr>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
</tr>
<tr>
<td>[Automatic] copy2archive</td>
<td></td>
</tr>
</tbody>
</table>

**ScanFlow™**

- organizes parallel tasks
- storage request (**distributed & centralized**) and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™: Jobs in the timetable**

<table>
<thead>
<tr>
<th>Task</th>
<th>Operator 1</th>
<th>Operator 2</th>
<th>Operator 3</th>
<th>Operator 4</th>
<th>Operator 5</th>
<th>Operator 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ScanRobot</td>
<td>Bookeye</td>
<td>Import Files</td>
<td>Bookeye</td>
<td>Release Jobs</td>
<td>Microfilm</td>
</tr>
</tbody>
</table>

- **ScanFlow™**
  - organizes parallel tasks
  - storage request *(distributed & centralized)* and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (**distributed & centralized**) and secure copy of digitized jobs
ScanFlow™: Jobs in the timetable

ScanFlow™

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs
# ScanFlow™: Jobs in the timetable

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
<td>MF 01</td>
<td>MF 02</td>
<td>MF 03</td>
<td>MF 04</td>
<td>MF 05</td>
</tr>
<tr>
<td>[Automatic]</td>
<td>copy2archive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### ScanFlow™

- organizes parallel tasks
- storage request (**distributed & centralized**) and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™: Jobs in the timetable**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata import</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 2</td>
<td>Bookeye</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 3</td>
<td>Import Files</td>
<td>book l1</td>
<td>book l2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 4</td>
<td>crop &amp; deskew</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Automatic]</td>
<td>MERGE B &amp; C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Automatic]</td>
<td>OCR-Server</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 5</td>
<td>Release Jobs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 6</td>
<td>Microfilm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Automatic]</td>
<td>copy2archive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- ScanFlow™ organizes parallel tasks
- Storage request *(distributed & centralized)* and secure copy of digitized jobs

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs
**ScanFlow™**

- organizes parallel tasks
- storage request *(distributed & centralized)* and secure copy of digitized jobs

---

*Utorok, 4. máj 2010*
**ScanFlow™**: Jobs in the timetable

- Organizes parallel tasks
- Storage request (distributed & centralized) and secure copy of digitized jobs

---

Utorok, 4. máj 2010
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs
**ScanFlow™: Jobs in the timetable**

- **Operator 1 (ScanRobot)**: book 1
- **Barcode Metadata import**: book 1
- **Operator 2 (Bookeye)**: book 1
- **Operator 3 (Import Files)**: book 1, book 2
- **Operator 4 (crop & deskew)**: book 1
- **[Automatic] MERGE B & C**: book 1
- **[Automatic] OCR-Server**: book 1
- **Operator 5 (Release Jobs)**: book 1
- **Operator 6 (Microfilm)**: MF 01, MF 02, MF 03, MF 04, MF 05
- **[Automatic] copy2archive**: book 1

---

**ScanFlow™**

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

**Utorok, 4. máj 2010**
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs
ScanFlow™: Jobs in the timetable

- organizes parallel tasks
- storage request (distributed & centralized) and secure copy of digitized jobs

Utorok, 4. máj 2010
**ScanFlow™**: Jobs in the timetable

- Organizes parallel tasks
- Storage request (distributed & centralized) and secure copy of digitized jobs
ScanFlow™: Examples

• renaming example
  – impossible to automize all the tasks
  – good preparation for manual tasks

• deskew and cropping issue
  – prevent wrong results, by “live preview” of what will be after the i.e. cropping
  – to avoid information lost
**ScanFlow™: Modules**

- Monitoring server
- Statistic server
- Report server (indiv.)
- OCR server
- Search engine
- Pages viewer
- File import/export
- Metadata editor
- *and others...*

The modules are listed with status indicators showing their current phase:
- Individually in test phases
- Stable release
ScanFlow™: in test phases & development

- search engine & serachability
- visualization
  - of the workflow itself
  - of the results
- bibliografic meta data import (ALEPH...)
- meta data / cataloge enrichment
- archiving solutions
- LTP (long time preservation)
- layout analysis and structurization
Conclusio: What means “workflow”?

The workflow-software:

• organizing & coordinating
• monitoring & overview

→ to handle the resources of a scan-center