

Satellite Meeting
"Conservation and preservation of library material in a cultural-heritage oriented context"

31 August - 1 September 2009
Rome, Italy

Organized by IFLA Core Activity on Preservation and Conservation
and IFLA Preservation and Conservation Section

Thanks to the support of:





robotic mass digitisation of cultural heritage library materials

by Stephan Tratter

IFLA post conference
Rome, 01.09.2009



Agenda

1. TREVENTUS

- a) activities
- b) story and market

2. mass-digitization

- a) digitization
- b) mass (with robotic support)

3. Tools

- a) scanner: manual & automatic
- b) workflow-software

4. TREVENTUS → solutions for mass digitization

1.) TreVentus Mechatronics – Vienna



1.a) What are we doing?

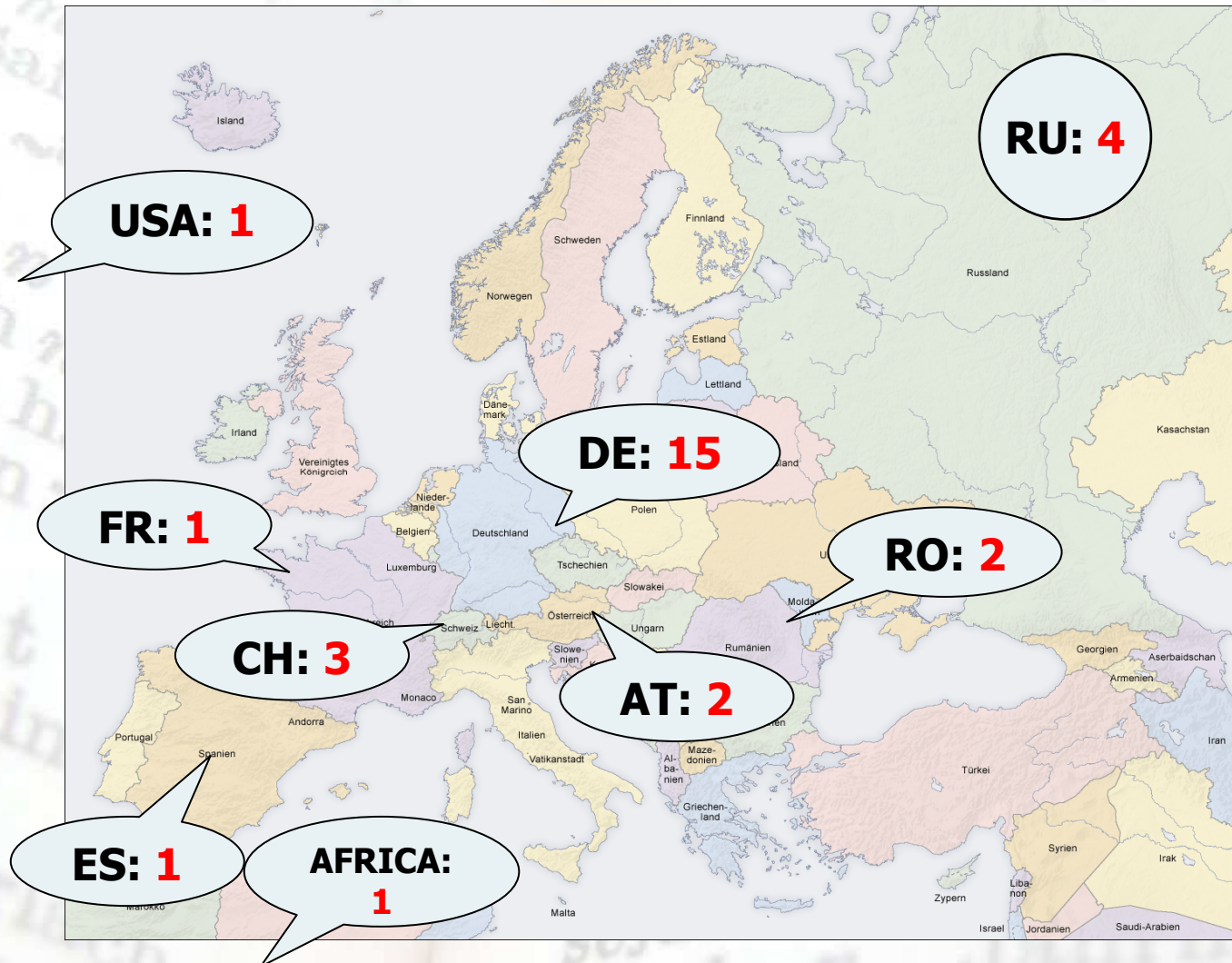
- automatic book scanner: **ScanRobot**®
- scanning software:
(scanning & image treatment) **ScanGate**™
- workflow software: **ScanFlow**™
- digitization service
(best practice inhouse)

1.b) Treventus 2004 - 2009

- 2004:** KickOFF to the pilot project „ScanRobot®“
- 2005:** proof of concept / first prototype & papers for the Patent
- 2006:** TU Wien SpinOFF: TREVENTUS Mechatronics
- 2007:** „ICT Grand Prize“ of the EU (CeBIT)
200.000 € prize money
- 2007:** product ScanRobot® roll-out & strategic partnerships
- 2008:** start: marketing, sales & production
start: digitization service
- 2009:** start: development ScanFlow™
→ status quo: 30 ScanRobots® sold worldwide



1.b) ScanRobot® - worldwide



30 ScanRobot® sold worldwide
(→ reliable technology | last update: 01.09.2009)

2.) What is mass digitization?

mass-digitization

workflow

logistic

organization

scanning

image treatment

text recognition

meta data

2.a) What means “digitization”?

- **book**
 - localized content
 - not easy accessible & transportable
- **scanned image**
 - de-localized content = a digital copy
 - easy to transport and to handle
(with no need to re-handle the original)
accessible from everywhere



digitizing = delocalization & accessibility
to the cultural heritage (content)

2.b) What means “mass”?

digitization

1 book	→ 400 pages	8 GByte
10 books	→ 4.000 pages	80 GByte
100 books	→ 40.000 pages	800 GByte
1.000 books	→ 400.000 pages	8 TByte
10.000 books	→ 4 Mio pages	80 TByte

mass digitization

Scale from “digitization” to “mass digitization” means “the digital industrial revolution”... = split the whole work in small pieces (tasks) and organize them (→ Taylorism).

2.b) challenges/efforts we encountered

- effective organization → higher throughput
- books are not homogeneous
- conservatorial aspects
- quality control → real time view
- staff motivation → intuitive user interfaces

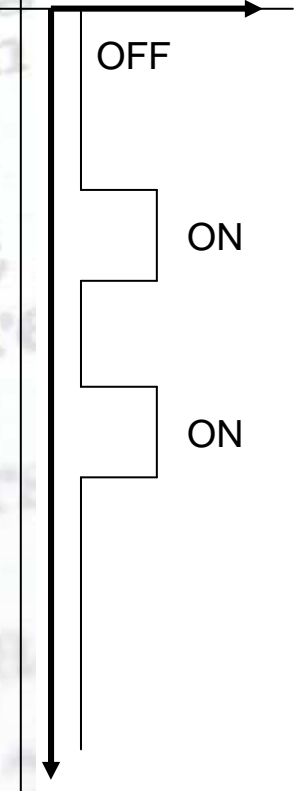
2.b) What means “effective”?

Digitization → one operator

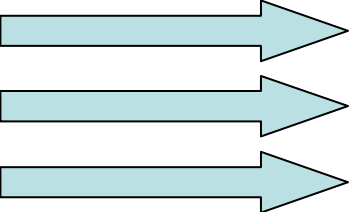




- picks up the book
- checks for meta data
- scans the book
- checks the pages – QC
- evtl. rescans pages
- makes image treatment
- saves the images on server
- bring back the book

ScanRobot

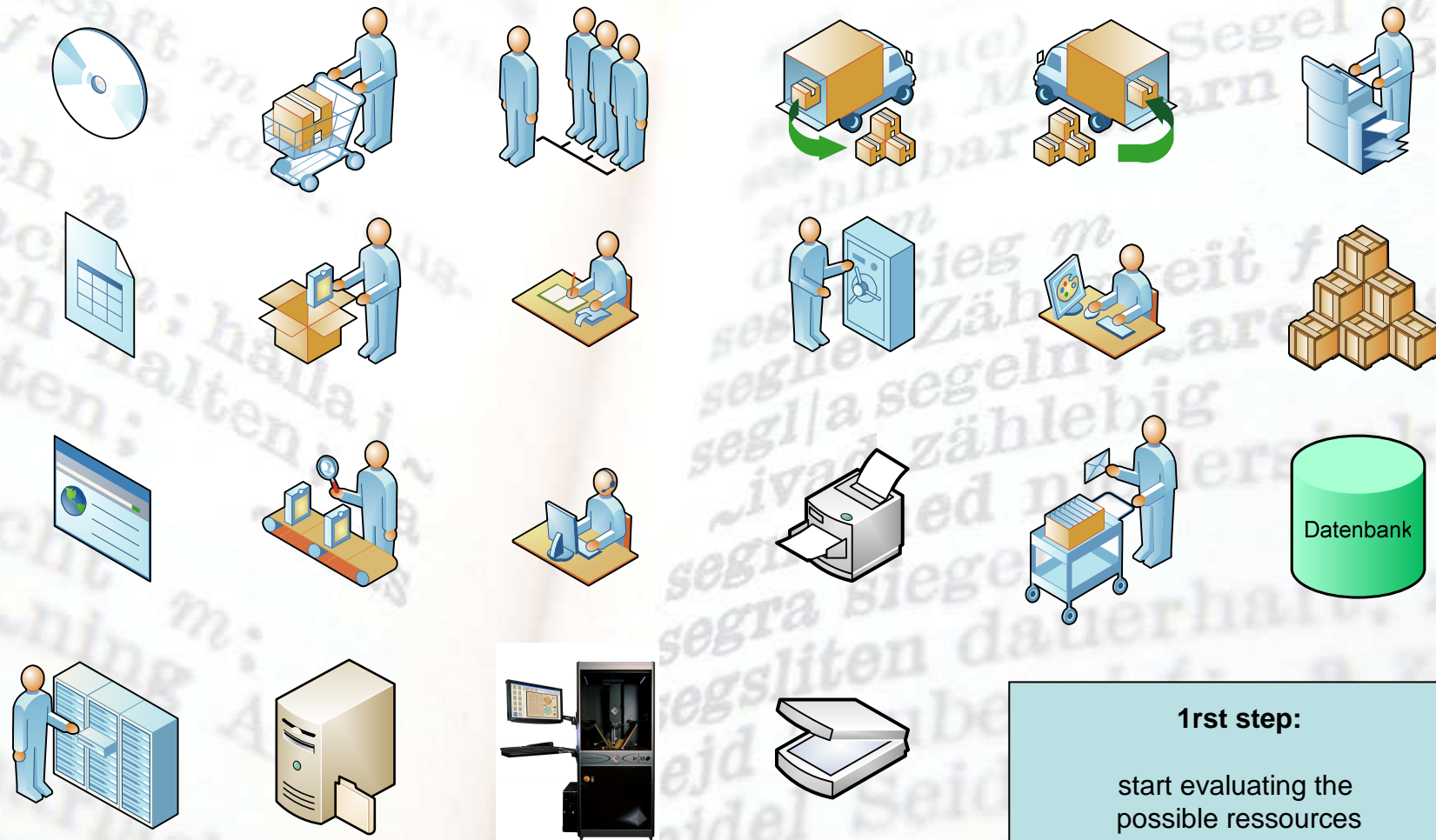
- ▶ not in use
- ▶ not in use
- ▶ in use
- ▶ not in use
- ▶ in use
- ▶ not in use
- ▶ not in use
- ▶ not in use



2.b) workflow example: parallel jobs

Inspection Jobsetup	ScanRobot	document treatment	indexing	archiving & viewing
<ul style="list-style-type: none"> • preparation • classifications • ScanJob setup • book cover • map etc. 	<ul style="list-style-type: none"> • scanning • Metadatea • multiple Image • → DFG-Viewer 	<ul style="list-style-type: none"> • Batchprocessing • book cover • map etc. • Metadatea • PDF and/or reprint 	<ul style="list-style-type: none"> • text recognition OCR • adding bibliografic metadata 	<ul style="list-style-type: none"> • repository • storage/archiv • (digital) long time preservation • Online → WWW presentation
1 x ScanOp 1 x PC 1 x ScanGate light 1 x ScanFeature	1 x ScanOp 1 x ScanRobot 1 x ScanGate	1 x ScanOp 1 x PC 1 x ScanGate light 1 x ScanFeature		→ depends on requirements
				

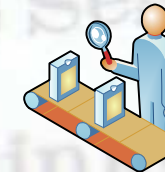
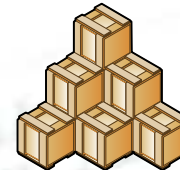
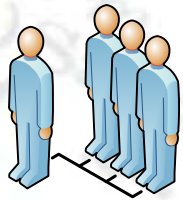
2.b) WF: ressources (1/3) – not ordered



1rst step:

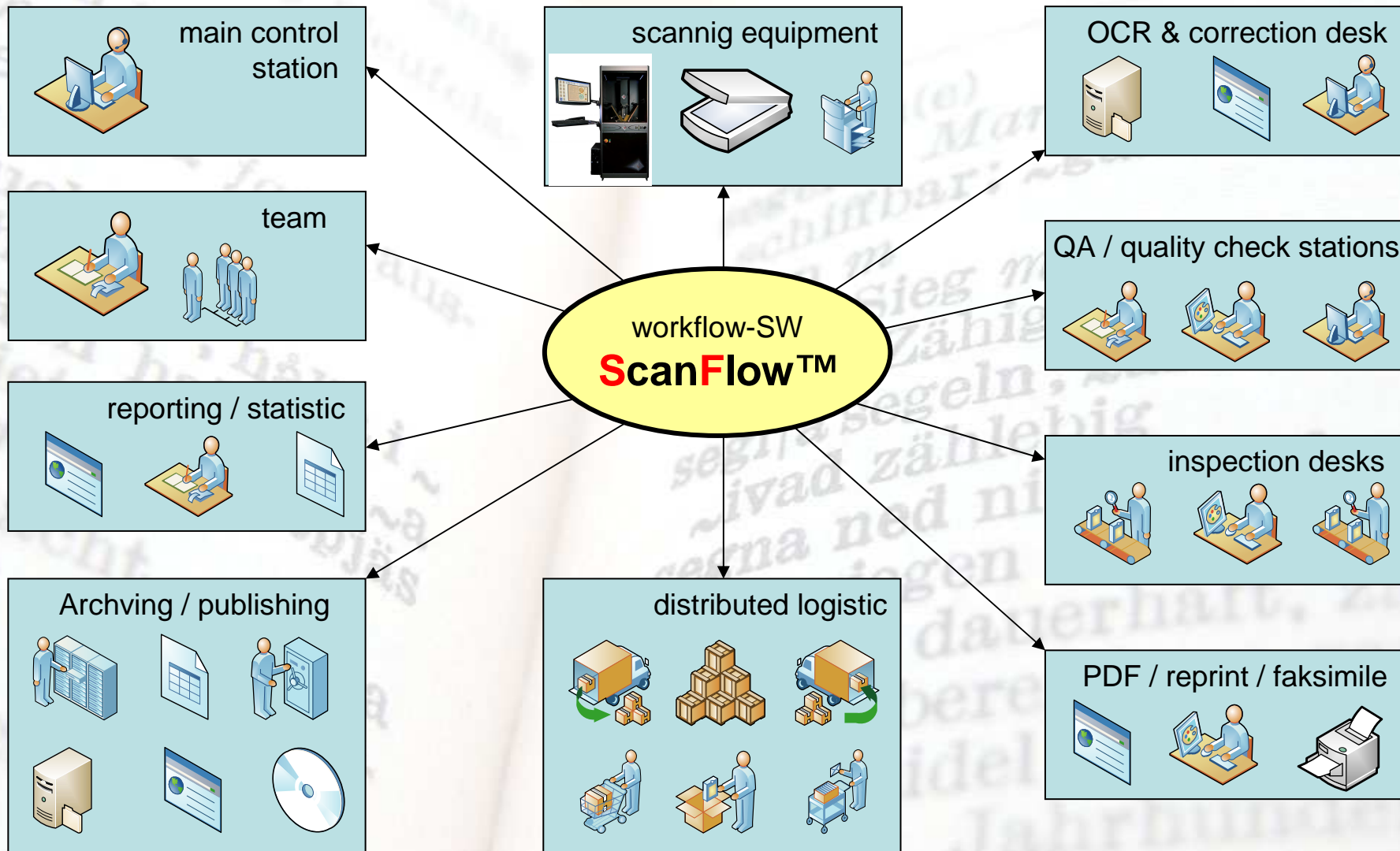
start evaluating the possible resources

2.b) WF: ressources (2/3) – logical groups



2nd step:
start building logical groups
start defining targets (i.e. reprint)

2.b) WF: ressources (3/3) – ordered modules



3.a) Tools for digitizing: scanner

- past: **ONLY** manual planetary scanner/fotografy
 - manual scanning process



- future: new technologies
 - automatic book scanner → ScanRobot®
 - V shape scanner → 60°
 - very gentle to the book
(much of the **cultural heritage books** are old and rare books)
 - constant quality of productivity



4.) TREVENTUS: mass digitization solutions

a) ScanRobot®

b) ScanGate™

c) ScanFlow™

d) Experience – digitization service

4.a) ScanRobot® SR301 → simple solution



- automatic scanner
- opening angle: 60°
- gentle air flows
- minimum illumination
- NO image distortions
- only 1 moving part!!!
- results in realtime
- image treatment (deskew, crop etc.)

4.a) ScanRobot® - the innovative idea

image capturing
without any distortions!!!

Lacke 1749 m

Weiler nördlich des Vigljochs. Bergstation der Vigljochseilbahn. Die Talstation liegt in Ober-Weig. Weg 34 ein Stück hinauf, dann Waldweg 7 weiter und, vorbei am Seehof, hinaus zum Weiler. *knapp 1 Std., leicht und lohnend.* Gasthaus Jocher (R 769/a) auf Weg 9 nordwärts nahezu eben zum Ziel. *HU gering, 20 Min., leicht und lohnend.*

Vigljochkirchlein 1743 m

Steig: R 520. Bergstation der Vigljochseilbahn (R 769/a) auf Waldweg 34 westwärts steigend zum Gasthaus Jocher am Kirchlein. *HU 257 m, knapp 1 Std. und lohnend.* Vom Jocher (R 767) auf Weg 34 A nordwärts durch Wald hinauf zum Ziel. *knapp 2 Std., leicht und lohnend.*

Naturnser Alm 1922 m

Steig: R 503. Vom Jocher (R 769) zuerst auf dem Weg 9 leicht ansteigend südwestwärts rechts ab und auf Weg 30 bis zur Alm. *HU ca. 200 m, 1 Std., leicht und lohnend.*

Naturnser Höhenweg (Höhenweg)

Steig, der das Vigljoch mit dem Kirchbachtal (Seitenast des Uffentalles) verbindet. Vom Jocher (R 769) auf Weg bzw. Steig 9 nordwärts über Almen südwestwärts zum Naturser Hochjoch, links zum Falkomai (2180 m) und auf



in der Gauschlucht

teilweise gesichertem Steig hinunter zum inneren Falkomai (2051 m, R 884) *HU 437 m, 4–5 Std., für Gehgewohnte ausdauernde leicht und lohnend.* – Vom Jocher nach St. Pankraz siehe R 886, in umgekehrter Richtung, Gehzeit 1/3 kürzer.

772 Naturser Hochwart 2180 m

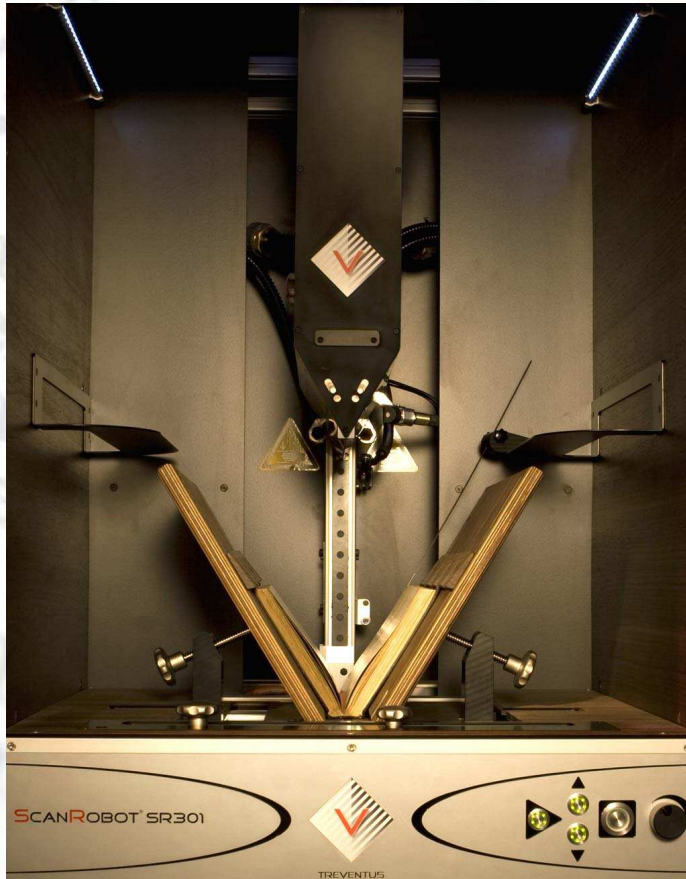
Bedeutende Gipfel in dem vom Vigljoch südwestwärts streichenden Kamm. Vom Vigljoch (R 769) auf Weg bzw. Steig durch Wald und über Almen südwestwärts bis zur Abzweigung des Gipfelsteigs. Von diesem hinauf zum Kamm und über den Gratrücken empord zum Gipfel. *HU 908 m, 3 1/2 Std., für Geübte leicht und lohnend.*

773 Nach Völlan

Charakteristik: R 779.

catturazione d'immagine
SENZA distorsioni!!!

4.a) ScanRobot® - opening angle

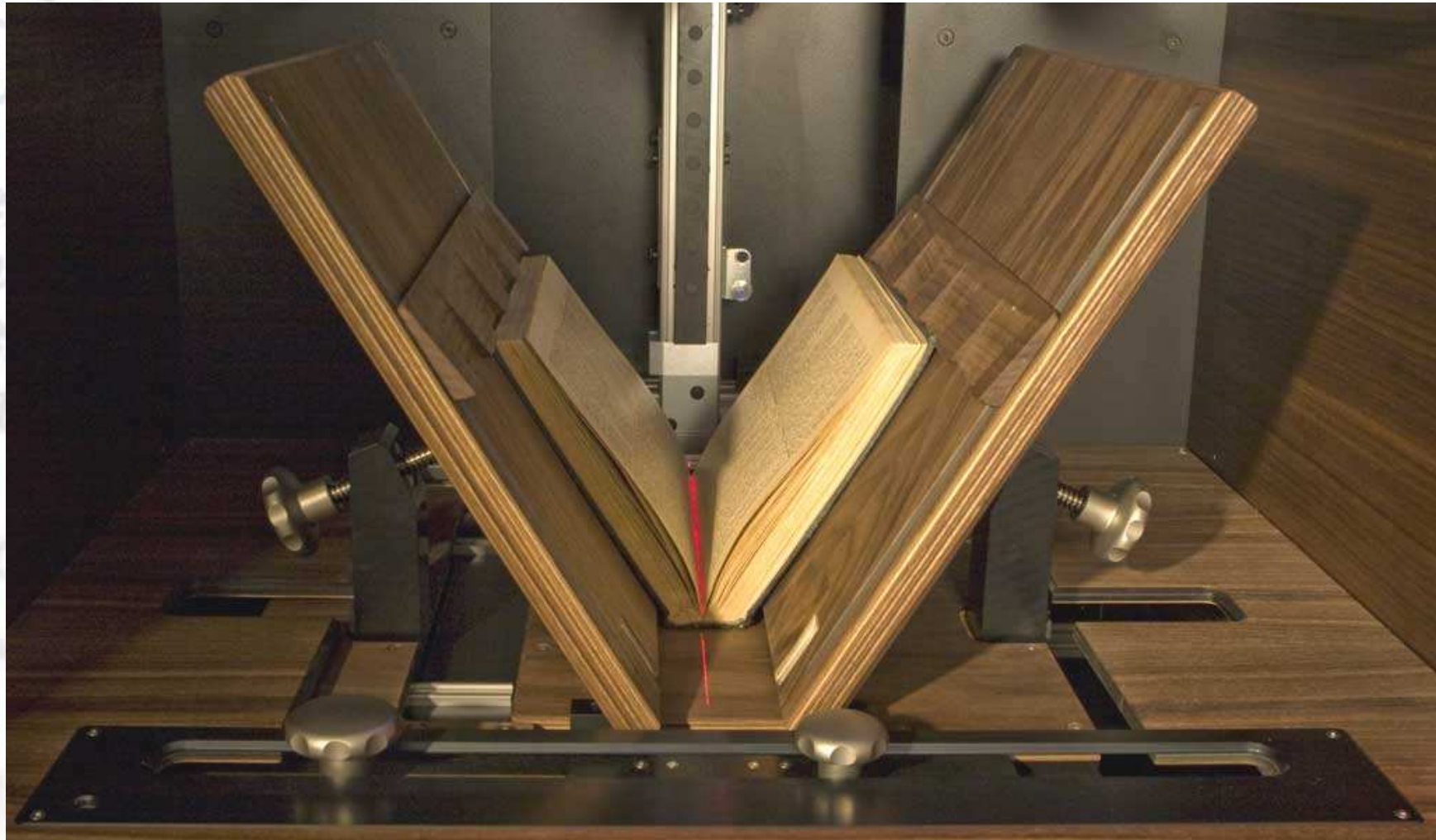


opening angle: 60°



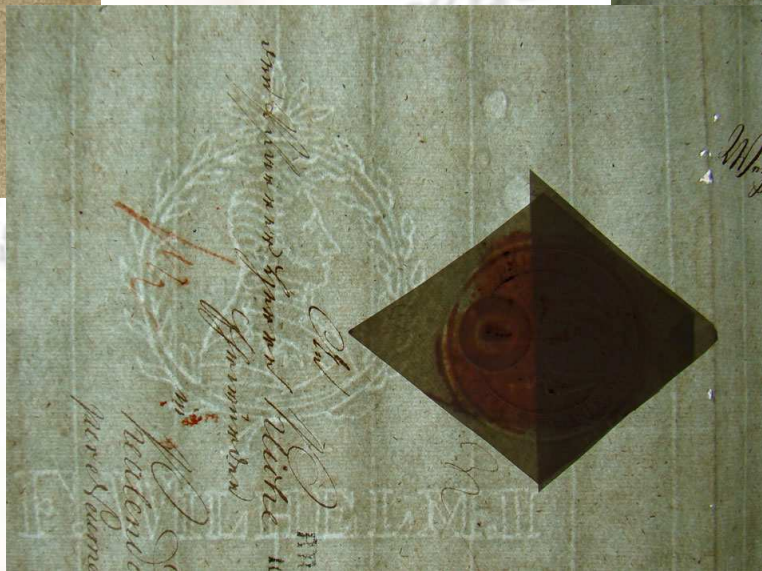
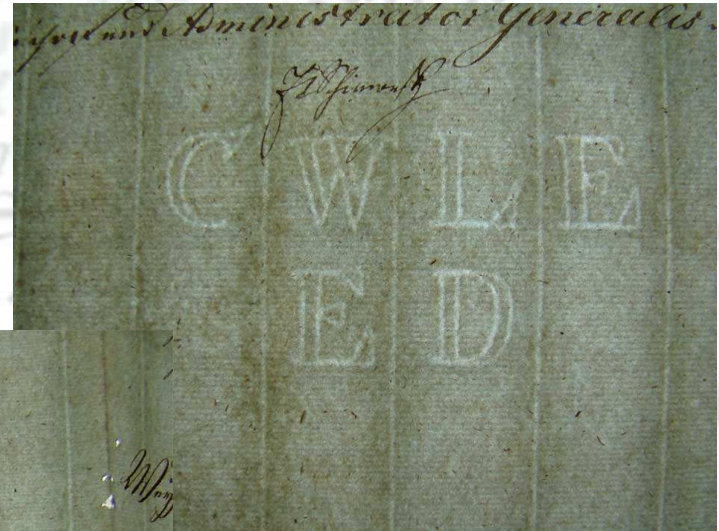
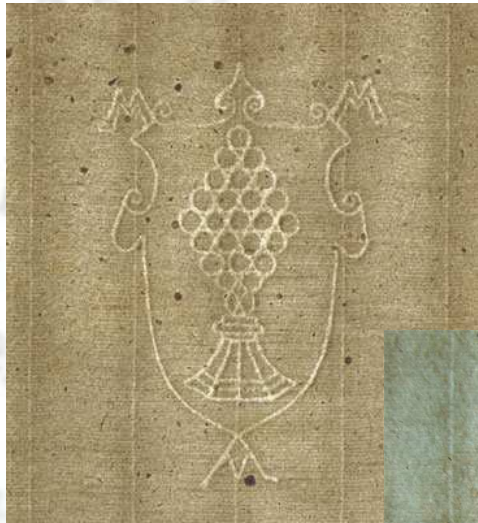
opening angle: 180°

4.a) ScanRobot® - the new book cradle

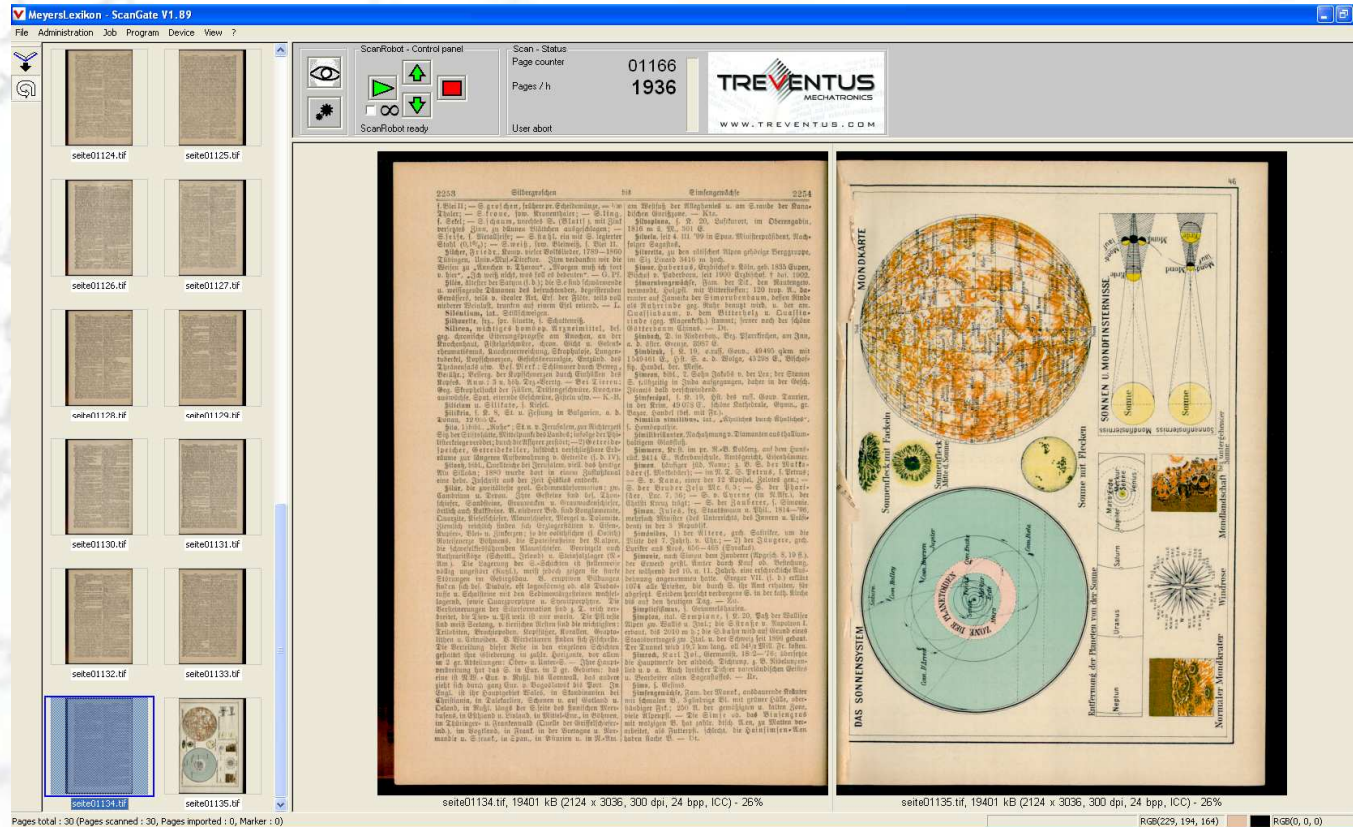


4.a) ScanRobot® - interesting details

- water marks (tested in Göttingen)



4.b) ScanGate™



- easy interface!
- multilanguage

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

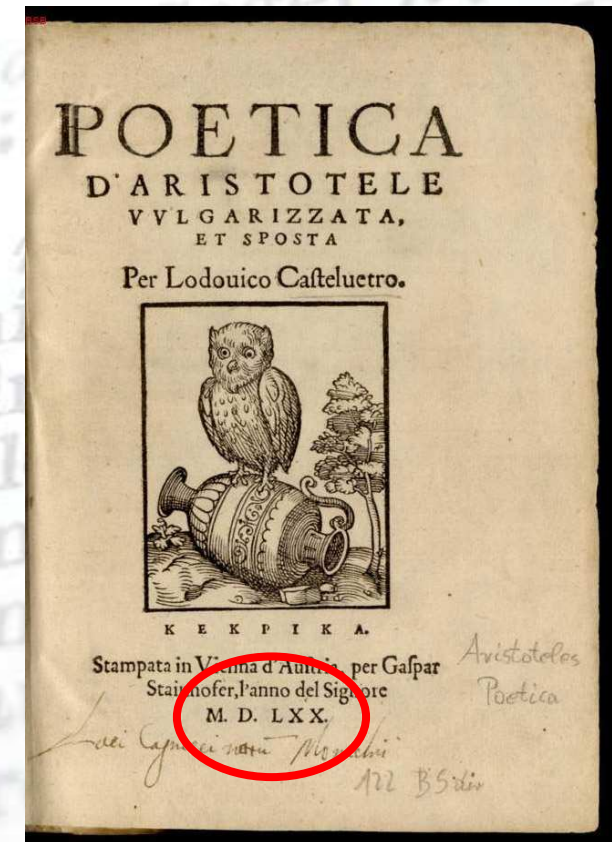
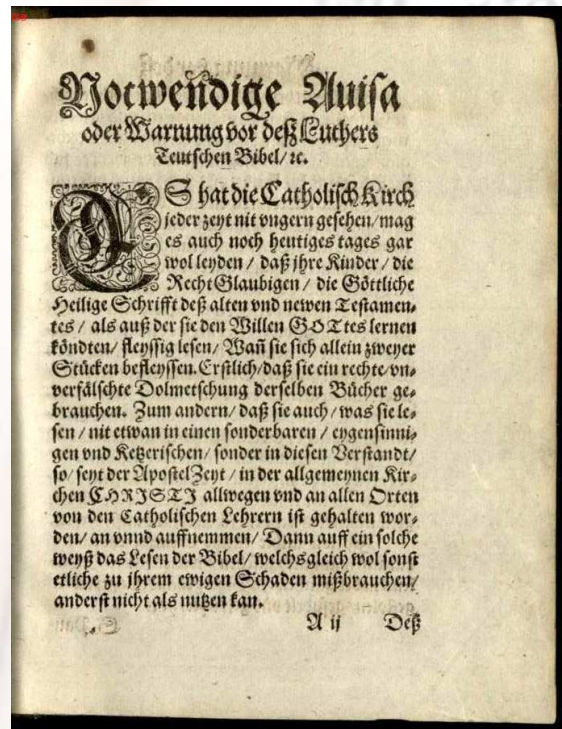
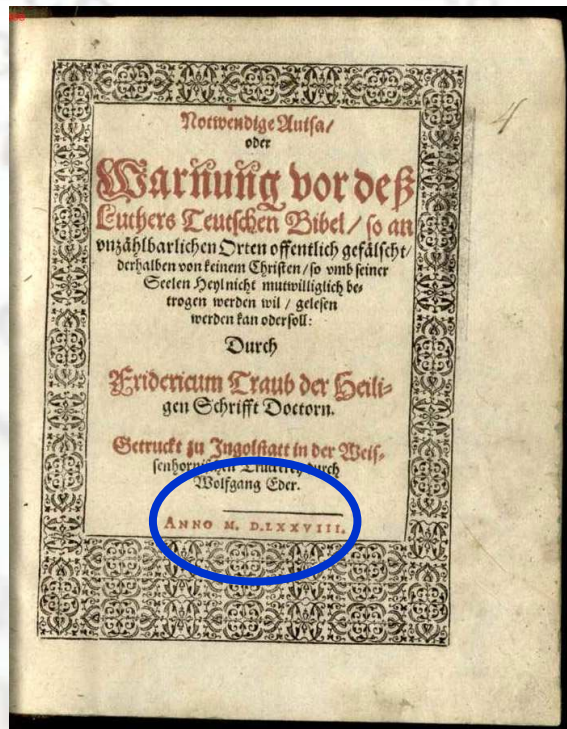
4.c) technical implementation in the field

- Bavarian State Library
 - VD16-2: finished on end of june 2009
 - now: 20th century books using 3 ScanRobot®
- the goal in both projects: racing up productivity
 - “parallelization” of the processes
 - constant quality in every brunch of process



4.c) results online (from the 16th century)

1. Aristoteles / Castelvetro, Lodovico: Poetica, Vienna 1570 [VD16 A 3572]
2. Warning of Luthers german bible, Ingolstadt 1578 [VD16 S 8592]



4.c) experience in mass digitization

scanning different books in mass digitization =
different results in throughput

- BSB → 2 year ~ 1,1 Mio. pages
- de Gryter → 6 months → 1,6 Mio. pages

preservation & accessibility



cardinal
Franziskus Eherle
(1845 - 1934)

preservation

accessability





www.treventus.com