SCANFLOW™
stay on Track!

Workflow Best Practices
3 examples

by
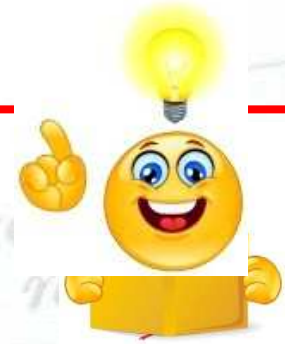Maximilian Kleiss
Wolfgang Novak
Stephan Tratter

TREVENTUS - Overview

SCAN → PROCESS → MANAGE → PRESENT

ScanRobot® 2.0 MDS
Automatic Book Scanner

ScanGate™
Image Processing

ScanFlow™
Workflow Management

Nainuwa™
Digital Library

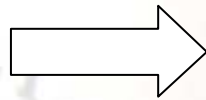# Introduction I – The idea

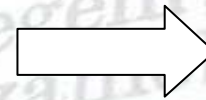Physical Material → Digital Library

Introduction II – Scanning

Physical Material → scanning → **Digital Library**
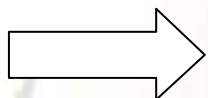
Introduction III – Digitizing
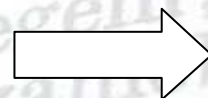
inhomogenous

exceptions

staff

preparation

quality control

scanning

Physical Material → digitizing → **Digital Library**

exceptions

output formats

exceptions

storage

OCR

archiving

metadata

languages

exceptions

TREVENTUS

Introduction IV – Workflow

Workflow

inhomogenous   exceptions   staff
scanning   quality control   preparation

Physical Material   ⇒   digitizing   ⇒   Digital Library

output formats   OCR   exceptions
exceptions   exceptions   storage
metadata   languages   archiving

TREVENTUS
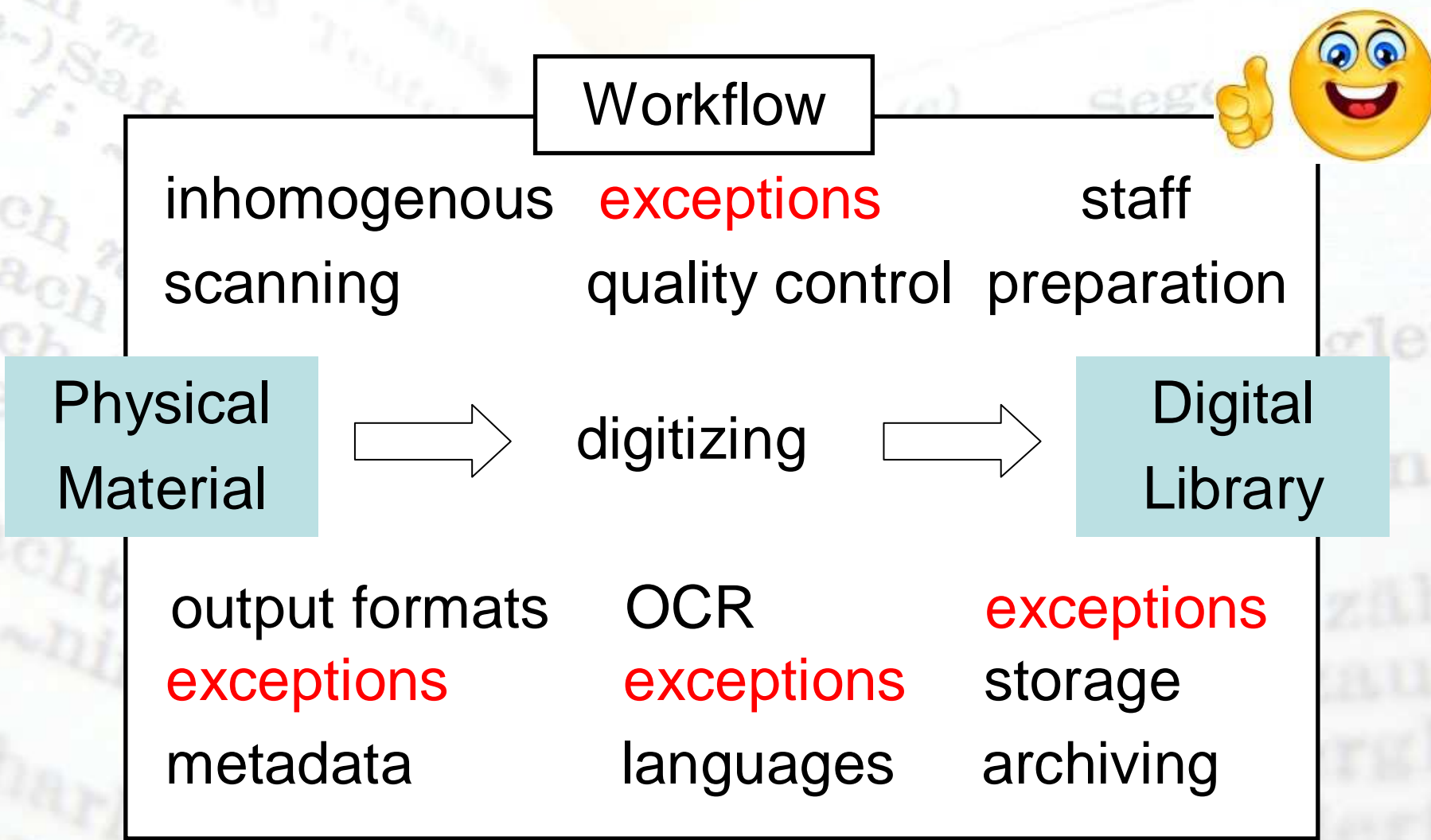
Workflow: examples

Complexity depends on:
- Desired end results
- Required quality
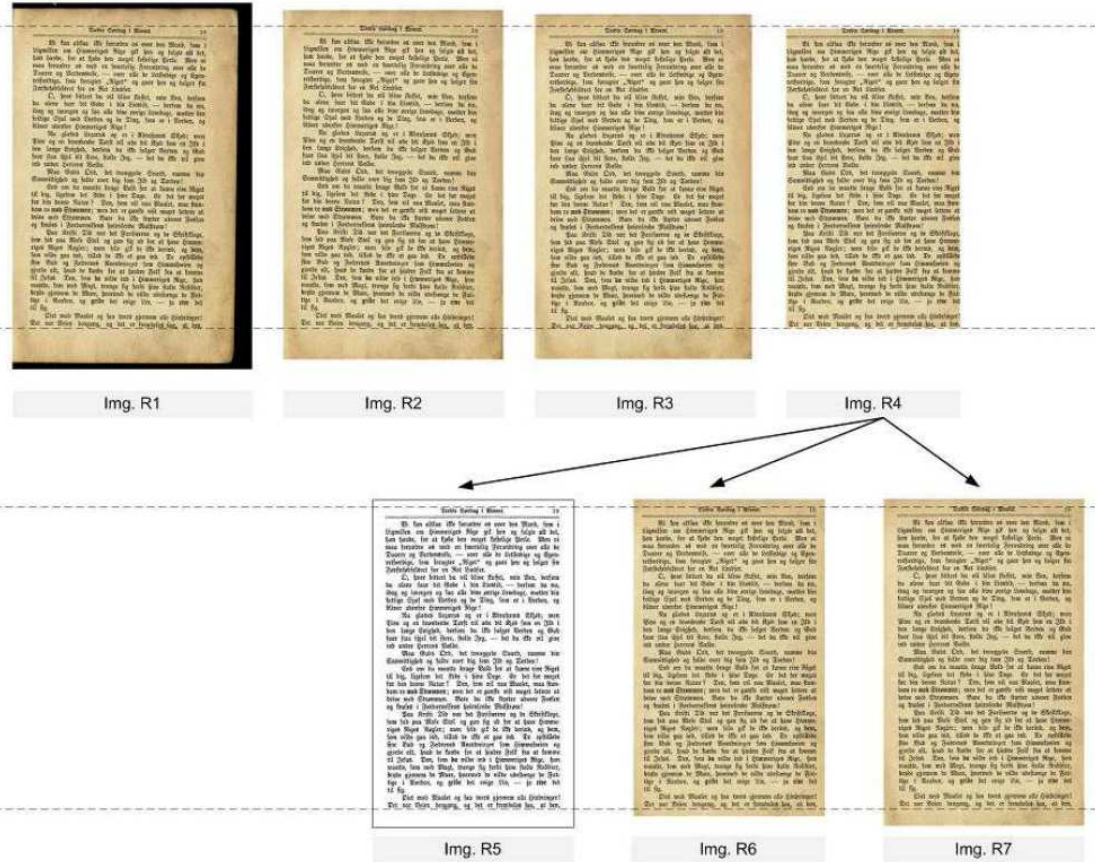

Image 1, Master and various derivates

Workflow: examples

Complexity depends on:
- Desired end results
- Required quality

→ Simple workflows:

#	Task	Mode	Format
00	Scan with BookEye / planetary scanner & check split line		
01	Page splitting	Master	TIFF

(01)   (01)

#	Task	Mode	Format
00	Release for backup		
01	BACKUP - Clean job	Master	TIFF
02	BACKUP - Move to backup folder		

(01)   (02)

Tasks

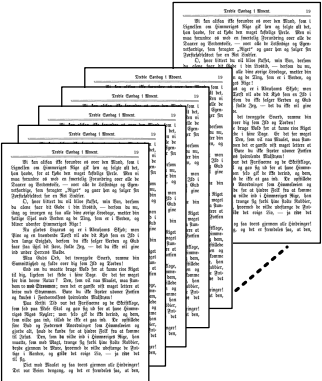| #	Task	Mode	Format
00	Scan with ScanRobot		
01	Create jpegs	Master	TIFF
02	Calculate borders and image attributes	Master	JPEG
03	Upload to the storage		
04	Calculate borders and image attributes, if not available	Master	JPEG
05	Check empty pages	Master	JPEG
06	Delete empty pages	Master	TIFF
07	Set image shape functions, title page. mark color, grey scale images and abstract	Master	JPEG
08	PRO - Create work copy of the master tif-files	Master	TIFF
09	PRO - Cropping and Extrapolation	Work	TIFF
10	PRO - Create jpegs for quality control (work)	Work	TIFF
11	OCR - Send marked images to the OCR	Work	JPEG
12	OCR - Get OCR File Back		
13	REN - Rename pages, all pages	Master	TIFF
14	REN - Rename pages, title page	Master	TIFF
15	BIN - Binarize, left and right pages	Work	JPEG
16	BIN - Despeckle	Bitonal	TIFF
17	PRO - Foldouts (normalize to A4)	Work	TIFF
18	PRO - Create Color exception (original size)	Work	TIFF
19	PRO - Create Grey exception (original size)	Work	TIFF
20	OUT - Clean output folder		
21	OUT - Copy output all tifs (folder - tif)	Bitonal	TIFF
22	OUT - Copy foldouts (folder - sup)	Bitonal	TIFF
23	OUT - Copy Ocr Abstract		
24	Release for backup		
25	BACKUP - Clean job		
26	BACKUP - Move to backup folder		

(04)   (23)

manual	automatic

TREVENTUS

ScanFlow™

Best Practice:

Example A

„Law Files"

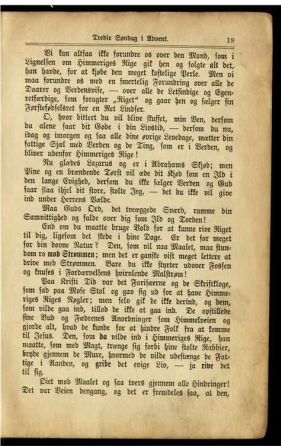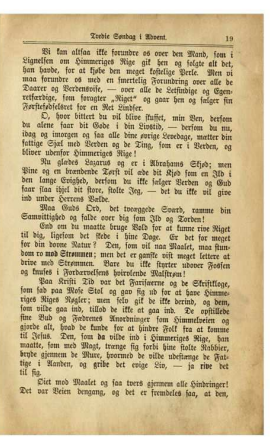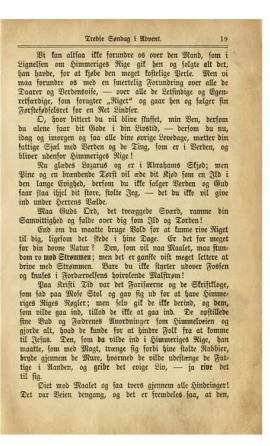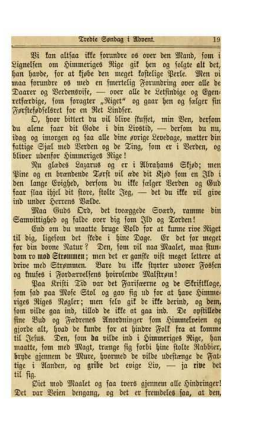# Example 1: Material / Documents

BATCH →	LAW FILE →	CATEGORY →	Single Sheets
000001	BAX120304	BAX120304-01	BAX120304-01_001.ext

- Total pages: ca. 1,5 Mio. pages (pilot project)
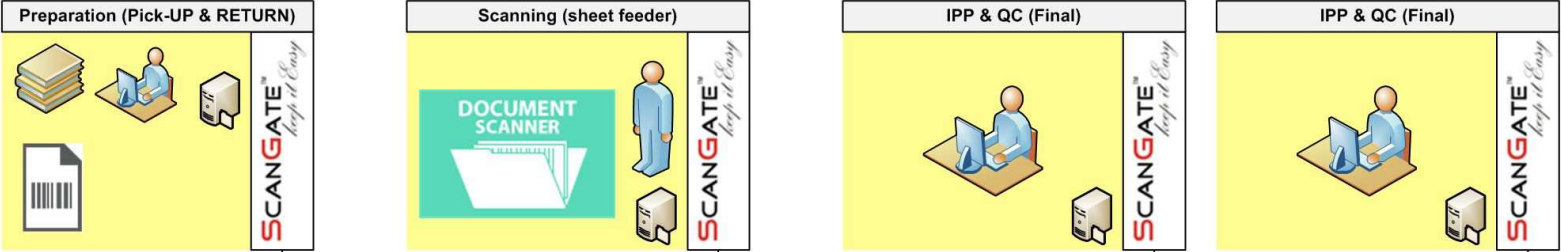
TREVENTUS

Example 1: Requested end result

- 1 physical page
  - → 1 MASTER file – TIFF
  - → 1 DERIVATE file – JPG
  - → 1 file searchable – PDF

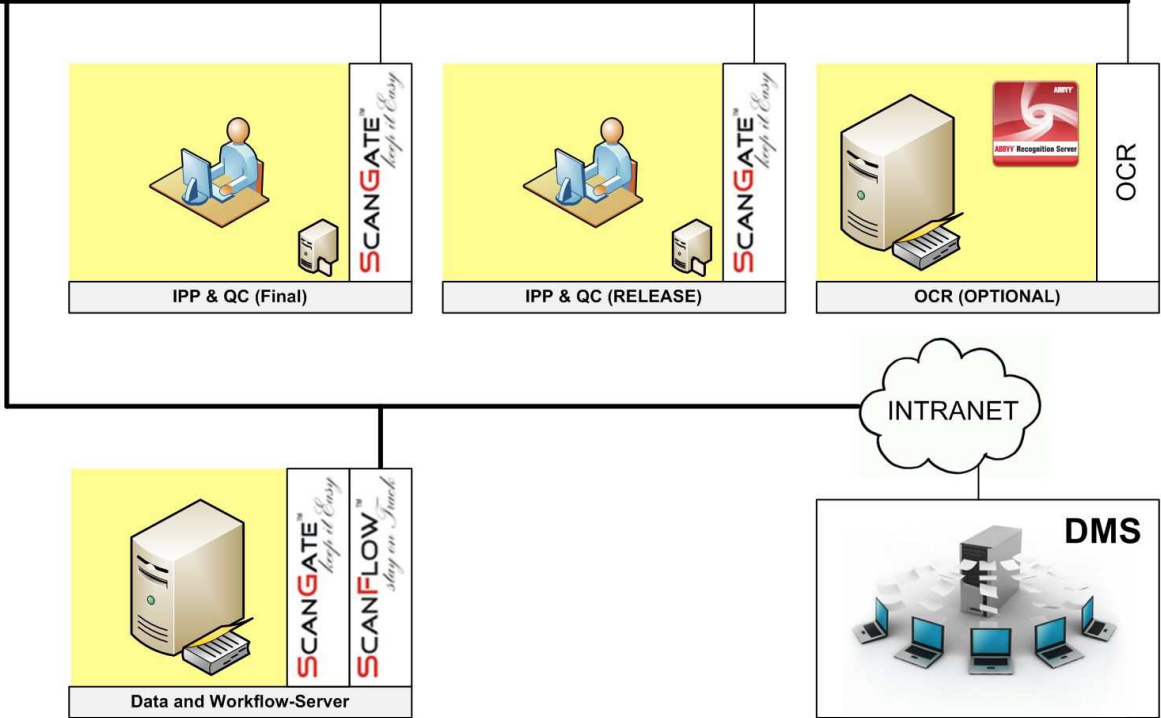| RAW - SCAN	CROP-1	DESKEW	CROP-2	Extrapolate & Background Homogenization
MASTER		→		DERIVATE

Example 1: Infrastructure

General overview: Digitization Project „Law Files"

Example 1: Challenges

- **Sheet Feeder Scanner output**
  - Convert to target structure
  - Quality control: Completness check
  - Renaming based on metadata from clients DB
- **Synchronization with clients DMS (handshake)**
  - Finished ScanJobs
  - Finished batches
- **Etc.**

Example 1: File & Folder - Renaming

Example 1: Synchronization w.extern SW

Sheet Feeder	ScanFlow	DMS Software (client)

CNVERSION
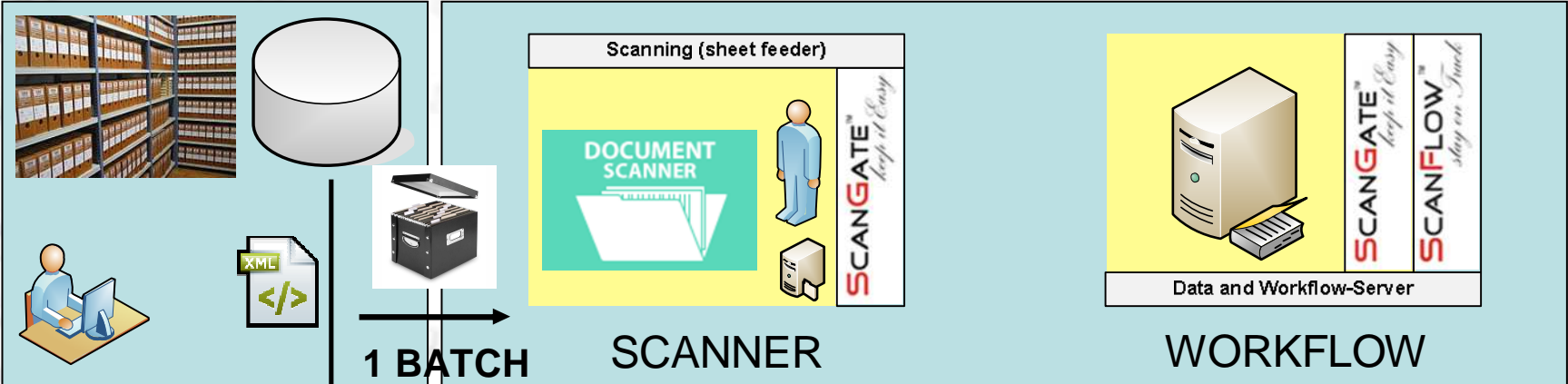
```
db_dump01.xml
1  <Batch-ID:15>
2      ID_01=BAX030406
3      ID_02=BAX030407
4      ID_03=BAX030412
5      ID_04=BAX030414
6      ...
7  </Batch-ID:15>
```

000001 ·······► BAX030406

Rename:
+) Folders
+) Files

Batch is finished!

LOG

Batch-end.log

TREVENTUS
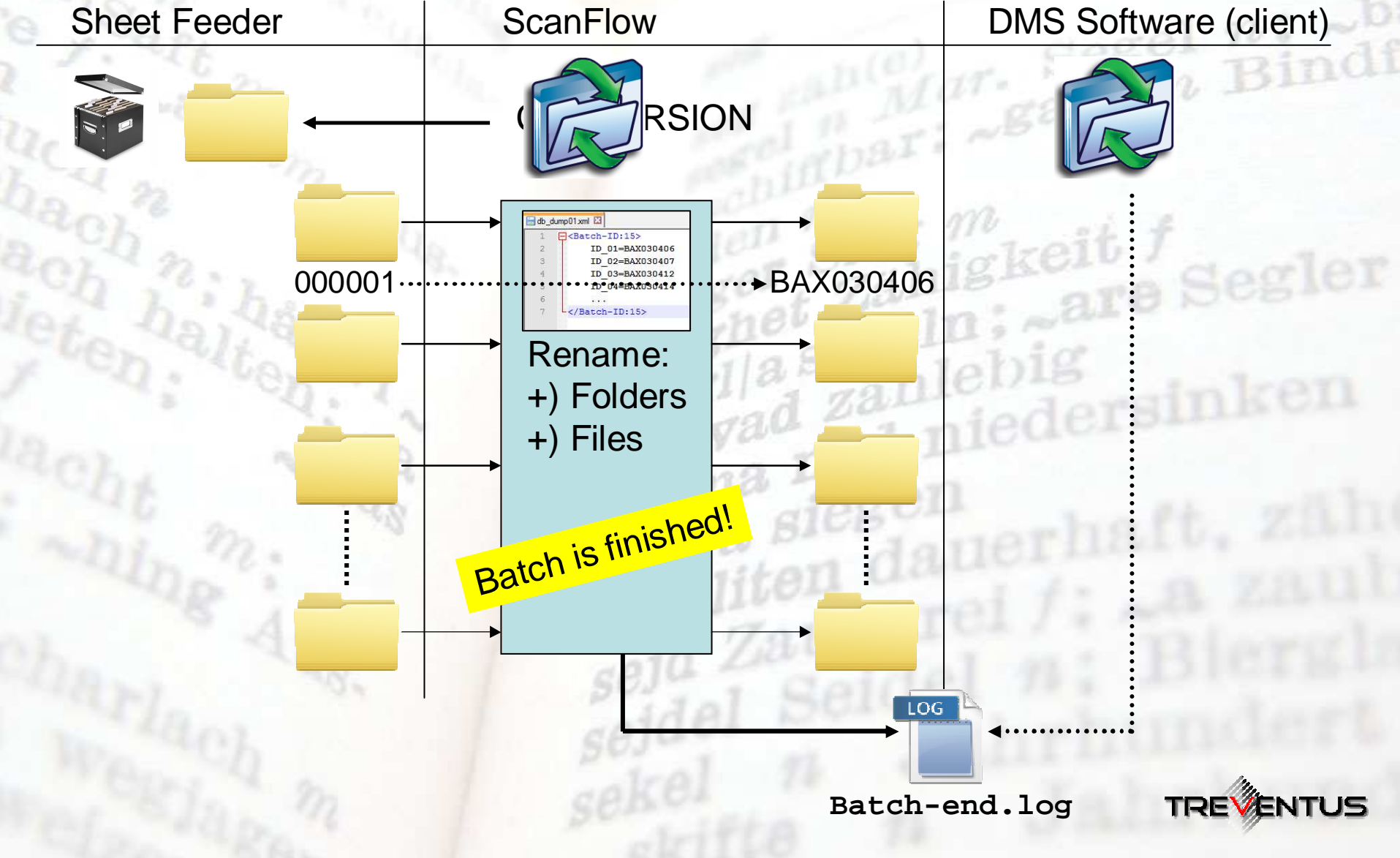
Example 1: Numbers and Figures

- Project time: 12 months
- Full throughput: ca. 120.000 pages/month
  - Including: scanning
  - Image treatment
  - Various Quality control steps

TREVENTUS

ScanF**low**™

Best Practice:

Example B

"Register archive – QR codes"

Example 2: Material / Documents

Project size:

Register books > 6.000

Pages   > 2 Mio

TREVENTUS

Example 2: Requirements

One register consists of:	
Colored & grey pages	→ Normal processing
Empty pages	→Scanned & saved → not processed
Maps	→ Scanned separately, → Imported to the correct place automatically, → Processed differently.
Page count & File naming	→ For each register separate page count. → One register goes over more books → Content split into its several chapters

One register

Book 1   Book 2

Ch 1   Chapter 2   Ch 3   Ch 4
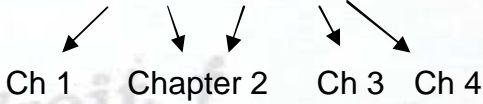
Customer wants the whole work-flow steered by QR codes.

TREVENTUS

Example 2: Requested end result

Requested result:

- Table of Content: PDF searchable
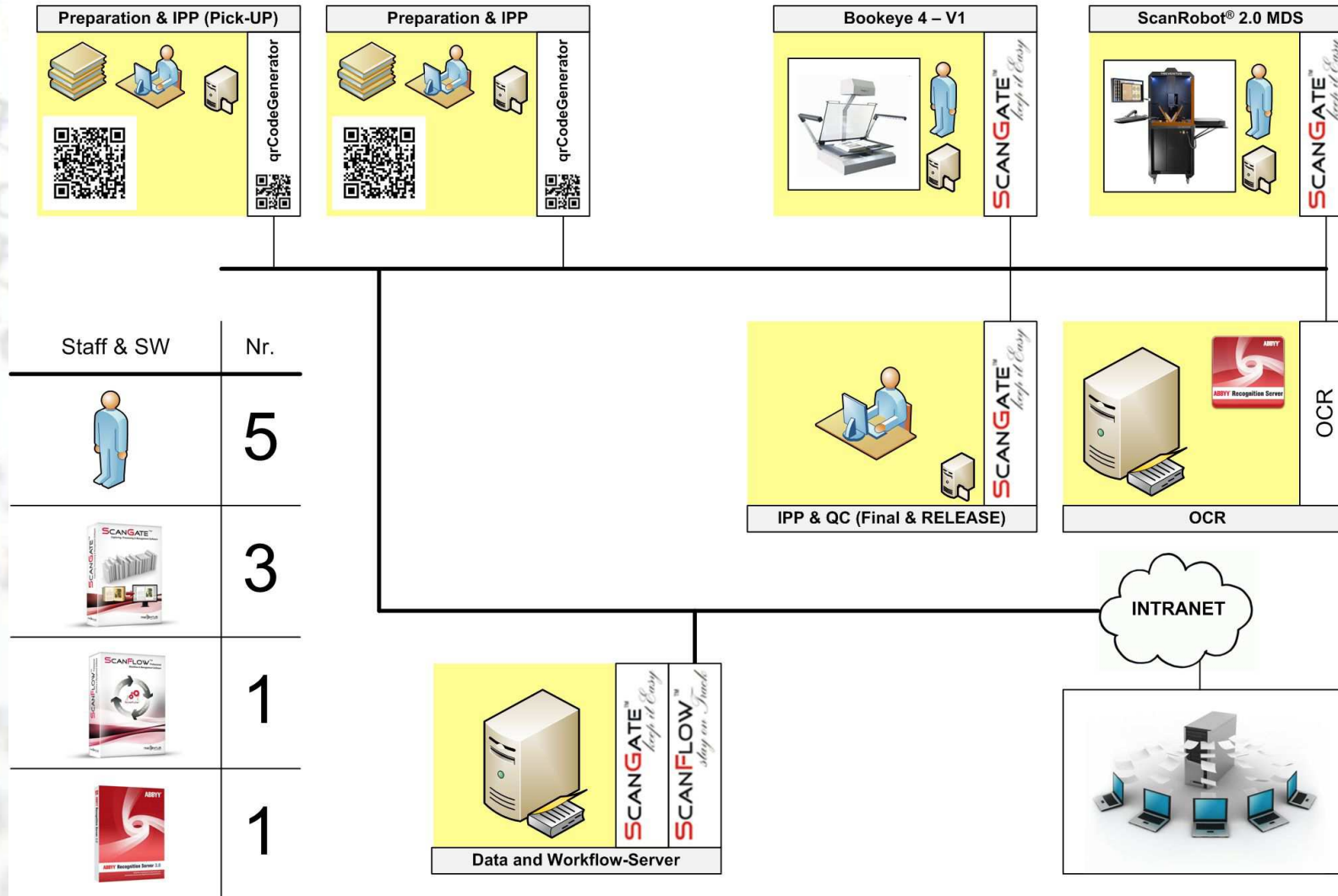
- Chapter wise: 1 PDF per chapter

- Single pages: JPEG

Register

PDF OCR	Table of content

+

PDF PDF PDF	Chapter 1
	Chapter 2
	Chapter 3
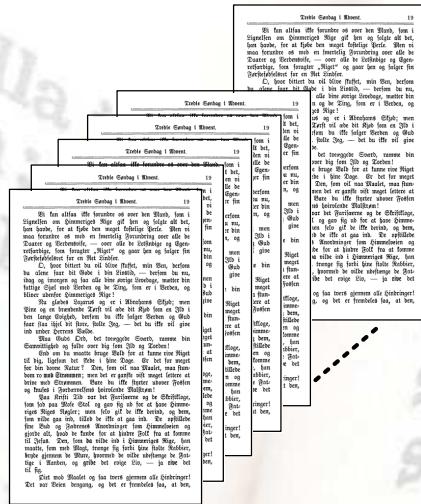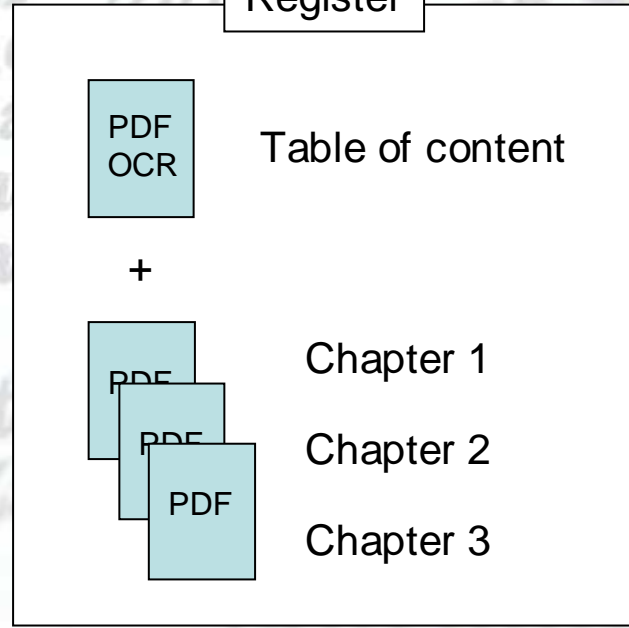
Example 2: Infrastructure

General overview: Digitization Project „BARCODES / QR-Codes"

Preparation & IPP (Pick-UP) — qrCodeGenerator

Preparation & IPP — qrCodeGenerator

Bookeye 4 – V1

ScanRobot® 2.0 MDS

IPP & QC (Final & RELEASE)

OCR

Data and Workflow-Server

INTRANET

Staff & SW	Nr.
	5
	3
	1
	1

Example 2: Solution

Solution:

- Developing a QR code system with small codes (<1 x 1cm) → no information lost
- Spooler function in QR code generator → avoids half empty pages → saves money

Using QR-Codes for steering
→ CQ-Codes are stuck on the pages where sequence changes:

- start and stop scanner,
- completeness check by page count information in QR code,
- meta data (page content) added from QR code
- place holder added,…

Example 2: qrCodeGenerator

Example 2: Challenges

Challenges:
1. Work-flow and scanner steered by QR codes.
2. If multiple QR Codes on page → detect correct QR code.
3. Different scanners → SW sorts the scanned pages automatically.

4. QR codes → as small as possible
5. QR codes → information about
   a. image treatment
   b. job handling

QR code

Scanner steering
Meta data
Page information
Processing information

SCANFLOW™

SCANROBOT® 2.0 MDS
MASS DIGITIZATION SYSTEM

Planetary scanner

TREVENTUS
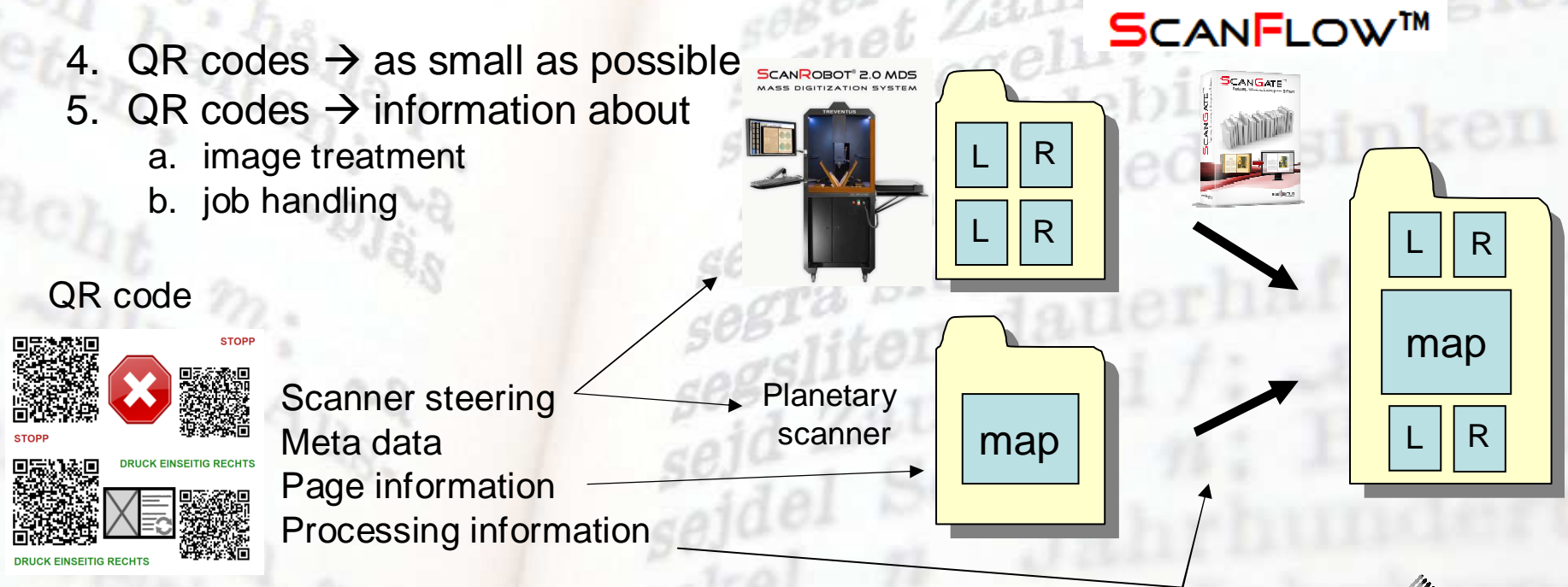
ScanFlow™

Best Practice:

Example 3

„Dissertations / Diploma Thesis"

Example 3: Material & basic conditions

- 5000 Dissertations & diploma thesis 3-4 times per year
- ~150 pages/thesis
- Format: A4 with exceptions
- Thesis with simplex and duplex print
- Content: mainly text but also images, diagrams, tables
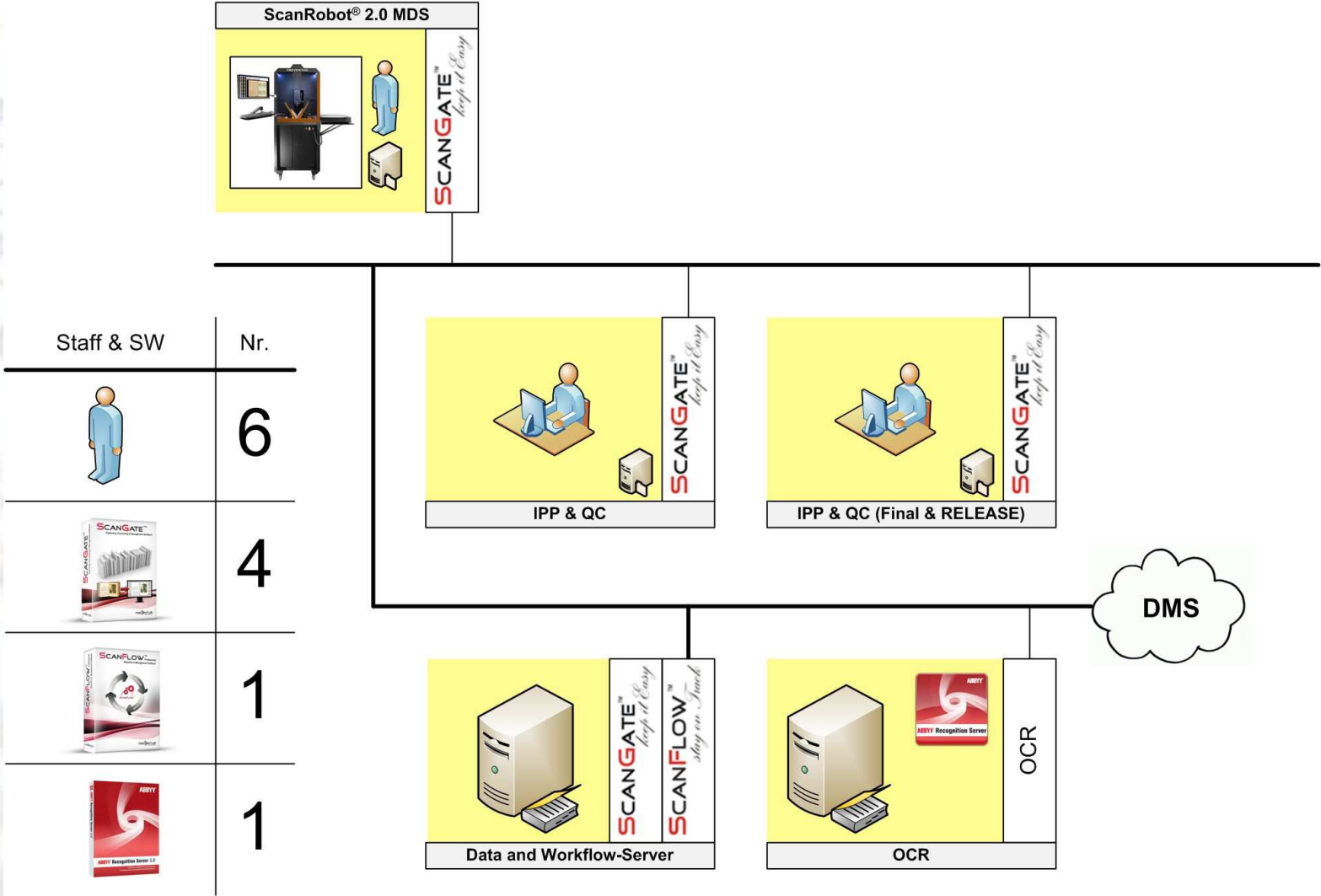- ~ 2.5 Million pages/year; ~ 750k per subset
- Digitization of each subset within 6-8 weeks

TREVENTUS

Example 3: Infrastructure

General overview: Digitization Project „Dissertations / Diploma Thesis"

ScanRobot® 2.0 MDS

Staff & SW	Nr.
(person)	6
SCANGATE	4
SCANFLOW	1
ABBYY	1

IPP & QC

IPP & QC (Final & RELEASE)
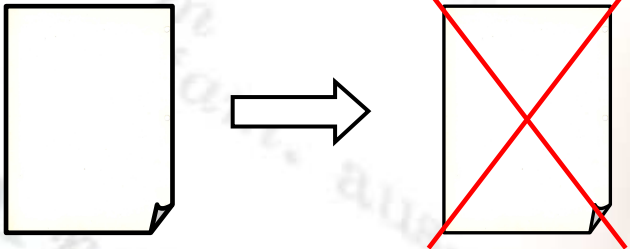
DMS

Data and Workflow-Server

OCR

Example 3: Requested end result (selection)

- ✓ Resolution: 400 dpi
- ✓ Cropping
- ✓ Scaling
- ✓ Naming convention (00001t.tif, 00002.tif)
- ✓ **Removing blank pages**
- ✓ **File fromat: TIFF** (CCITT group 4 & LZW compressed)
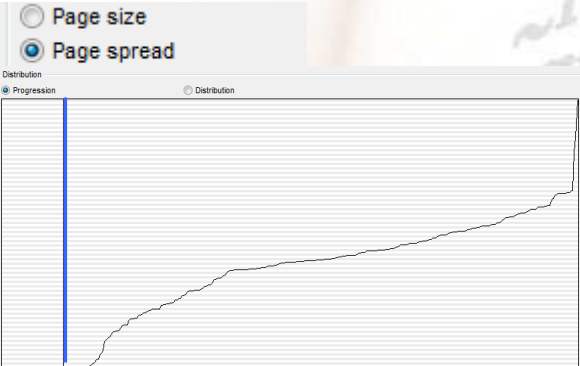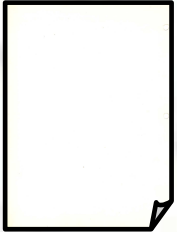- ✓ **OCR for abstract only**

TRE**V**ENTUS

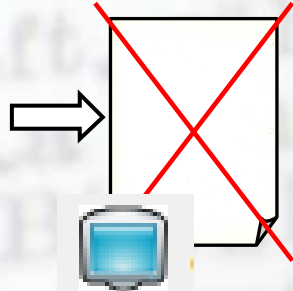# Example 3: Challenges & Exceptions I

☑ Deleting blank pages

Workflow

Example 3: Challenges & Exceptions II

☑ File format: TIFFs depending on content
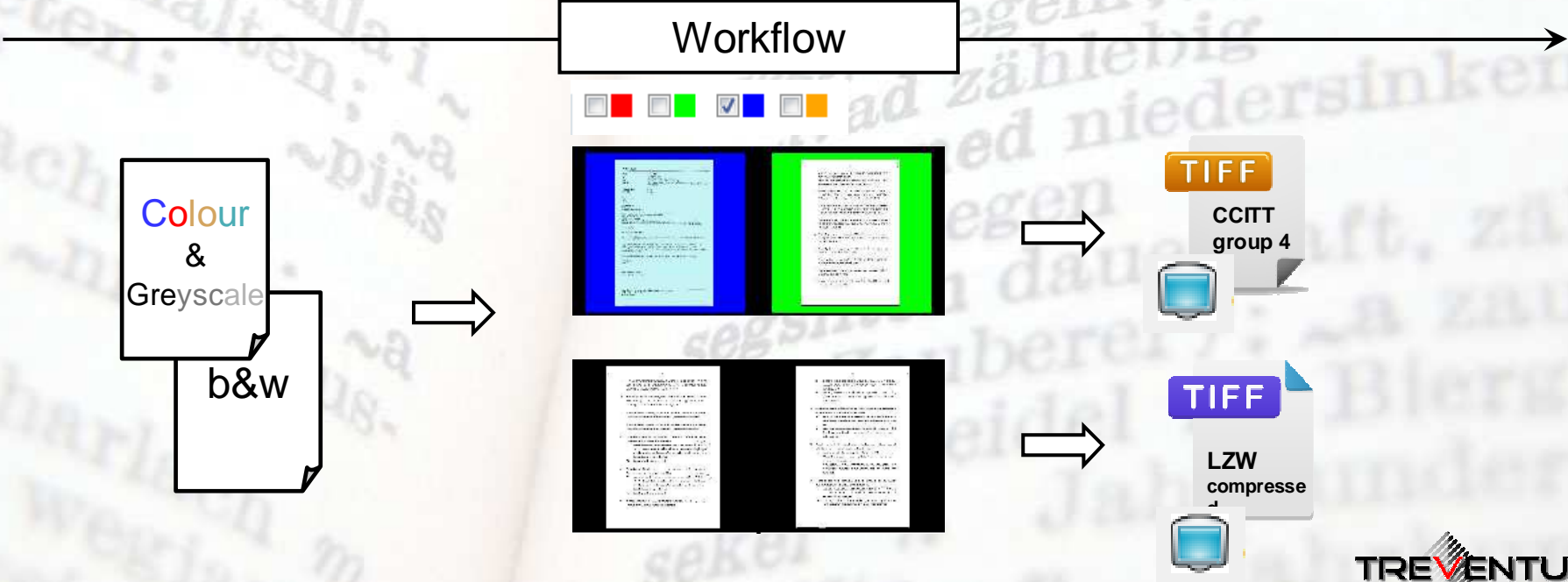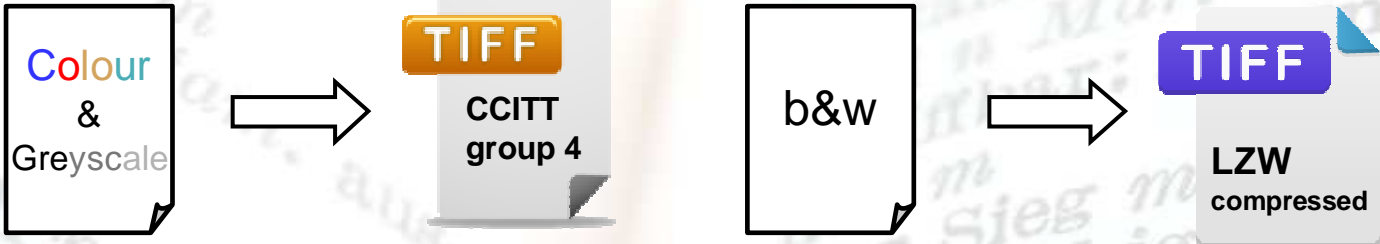
Example 3: Challenges & Exceptions III

☑ OCR for abstract pages only

workflow

CONCLUSIO

Conclusio: Digitization Workflow characteristics

workflow-software
ScanFlow™

- Embraces **all areas** of digitization (Metadata, OCR, QC, final results, DB handshakes,...)
- **structures the flow** of work from the book to the final result
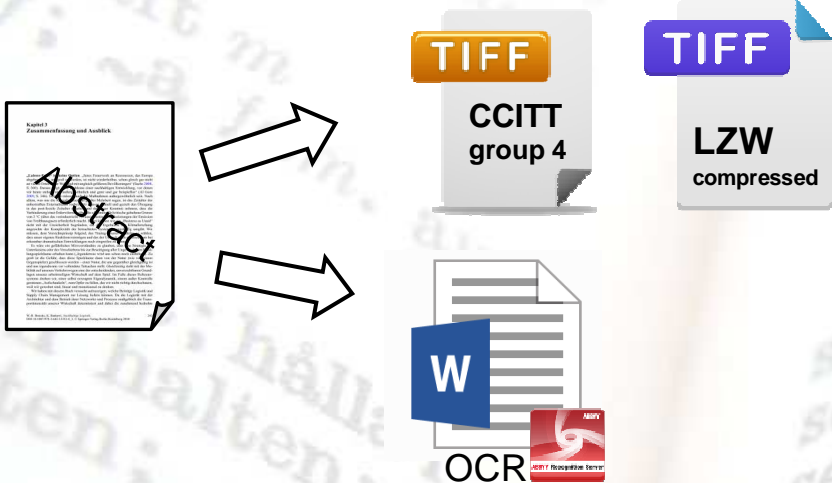- **reduces communication** overhead ("*He, does anyone know if the missing pages of this big brown book has already been rescanned?*")
- enables you **define quality control** steps according to your needs
- enables you to **split the work** to subtasks and distribute it according to your resources
- enables you to **make changes** to all areas of your workflow at any time (exeptions, unexpected challenges)
- gives you an **overview and status** of all of your digitizing projects

TRE**V**ENTUS

Conclusio: Digitization Workflow

workflow-software
ScanFlow™

To put it in a nutshell:

"A good solution for your digitization projects makes your life easier, reduces the required resources and helps to increase the output in terms of quality and quantity."

TREVENTUS

microform

Partner in SK
Alena Kulikova

Digital Document Professionals

Partner in PL
DDP Sp. z o.o

www.treventus.com

tratter@treventus.com

TREVENTUS