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)

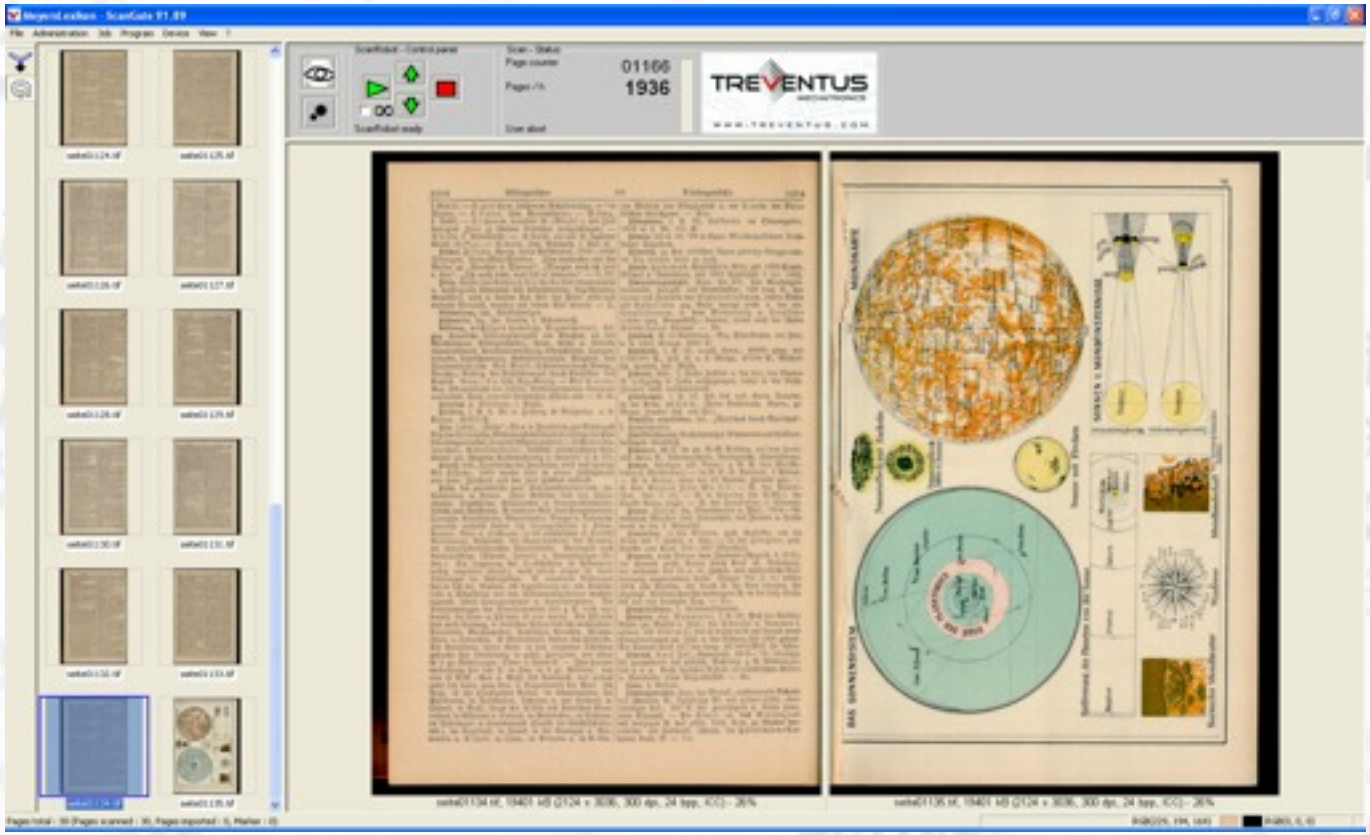
TREVENTUS

ScanRobot® SR301 → simple solution

- opening angle: 60°
- gentle air flows for page turning
- minimum illumination!!!
- only 1 moving part!!!
- results in realtime onscreen
  → **WYSIWYG** (**w**hat **y**ou **s**ee **i**s **w**hat **y**ou **g**et)
- NO image distorsions
- NO clambs in the result images!!!
- NO permanent calibration needed

TREVENTUS

ScanGate™



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

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

TREVENTUS

ScanRobot® - opening angle

Öffnungswinkel: **60°**

Öffnungswinkel: **180°**

TREVENTUS

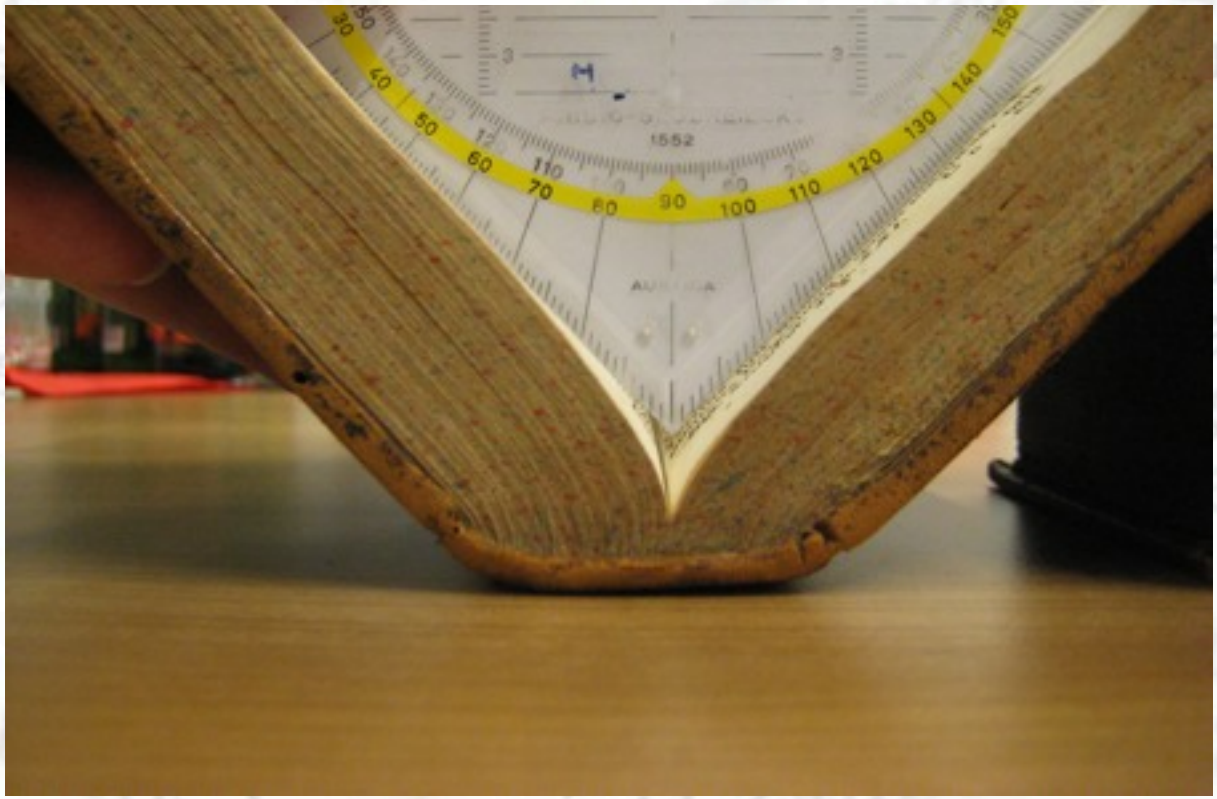
ScanRobot® - opening angle

Öffnungswinkel: **60°**

Öffnungswinkel: **180°**

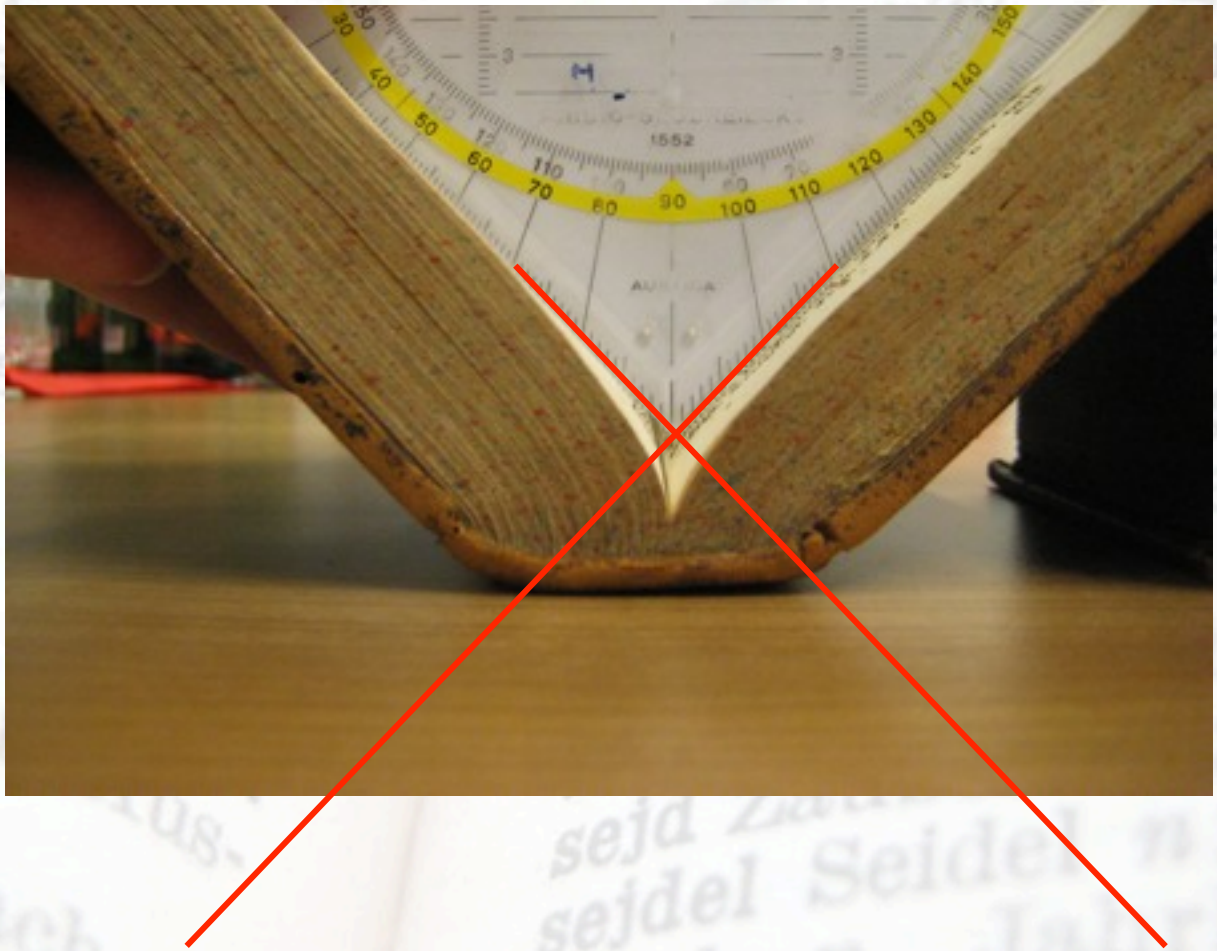
TREVENTUS

opening angle: 90° vs. 60°



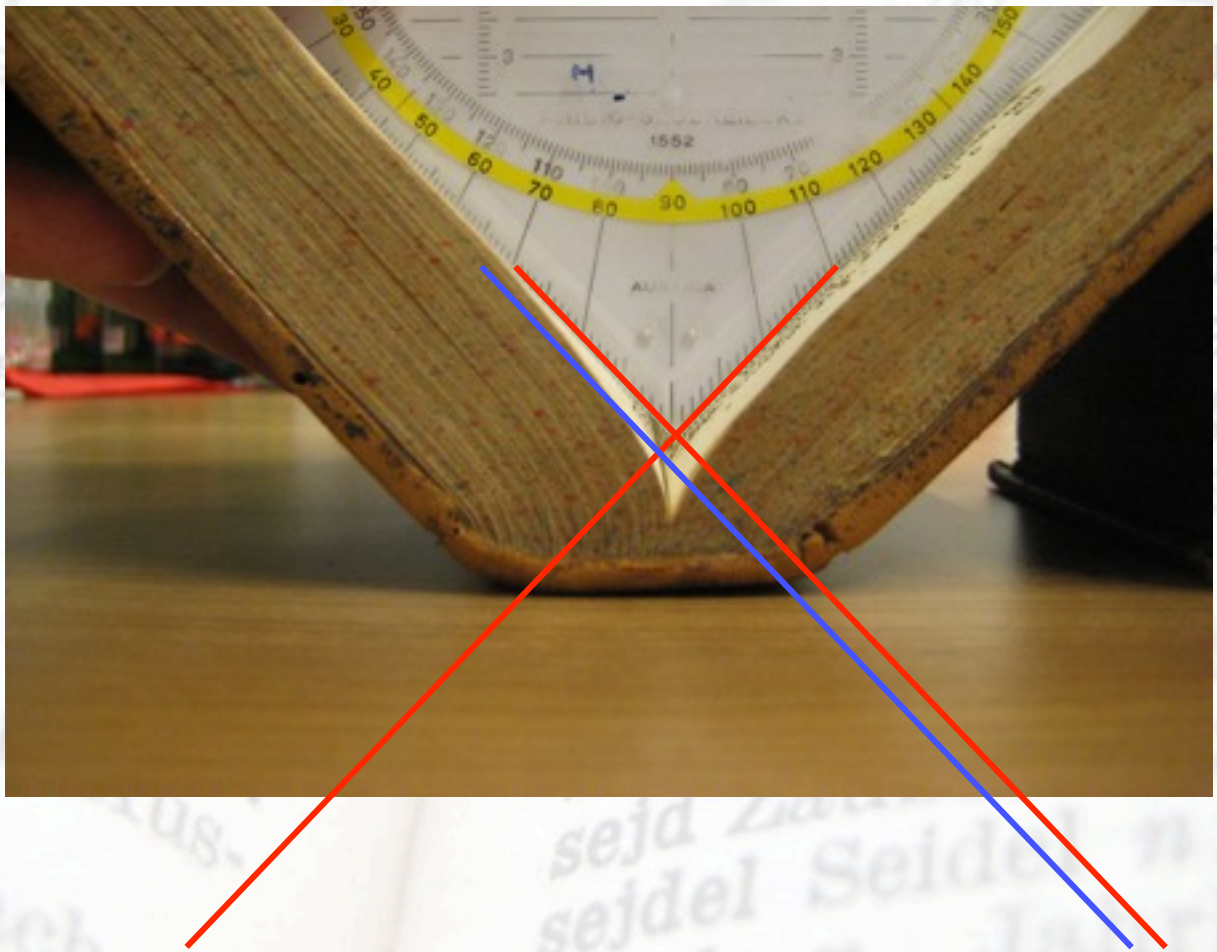
TREVENTUS

opening angle: 90° vs. 60°



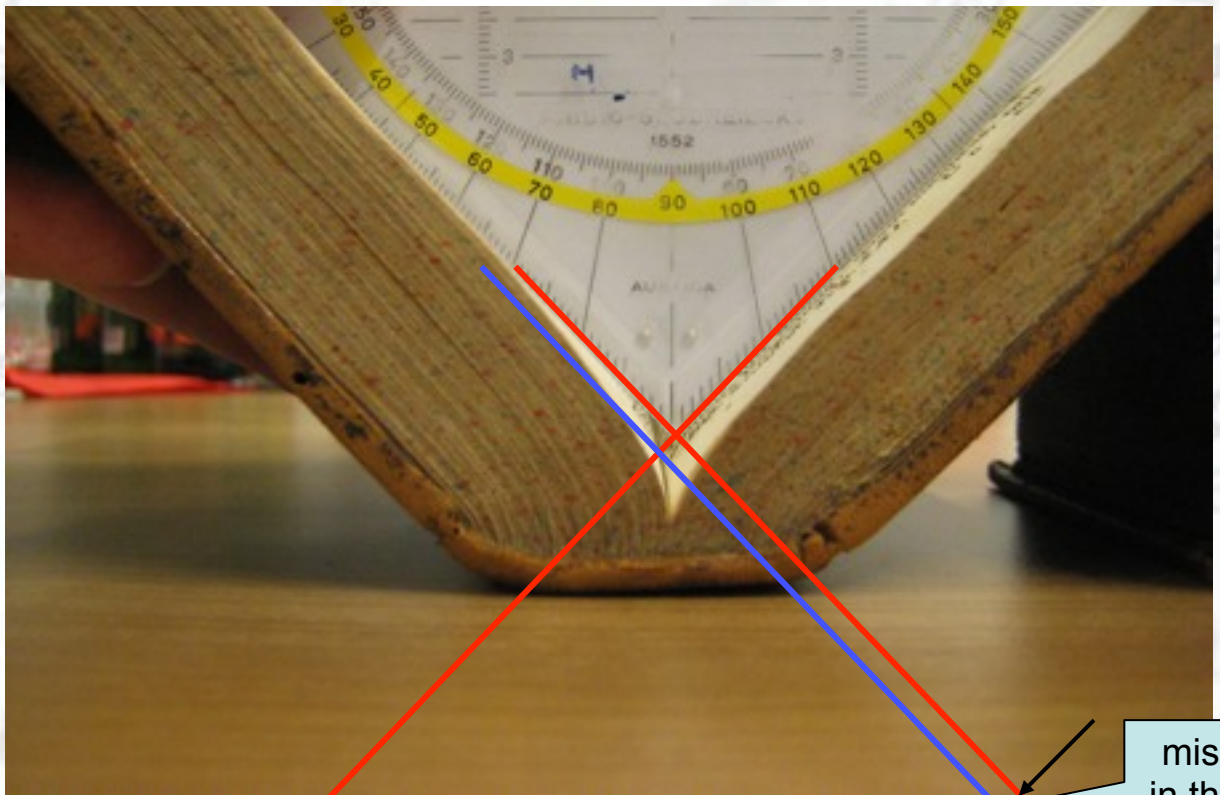
TREVENTUS

opening angle: 90° vs. 60°



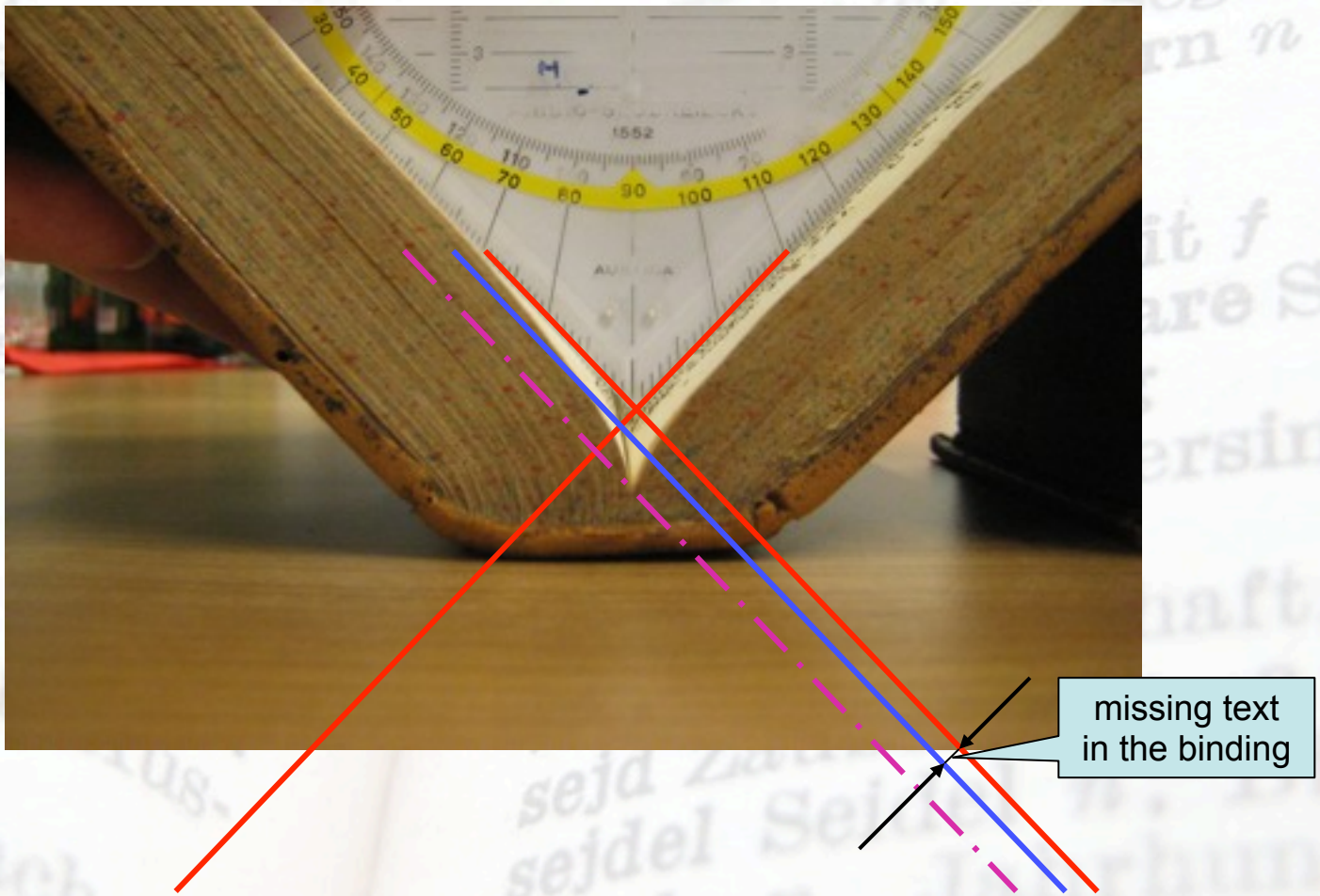
TREVENTUS

opening angle: 90° vs. 60°



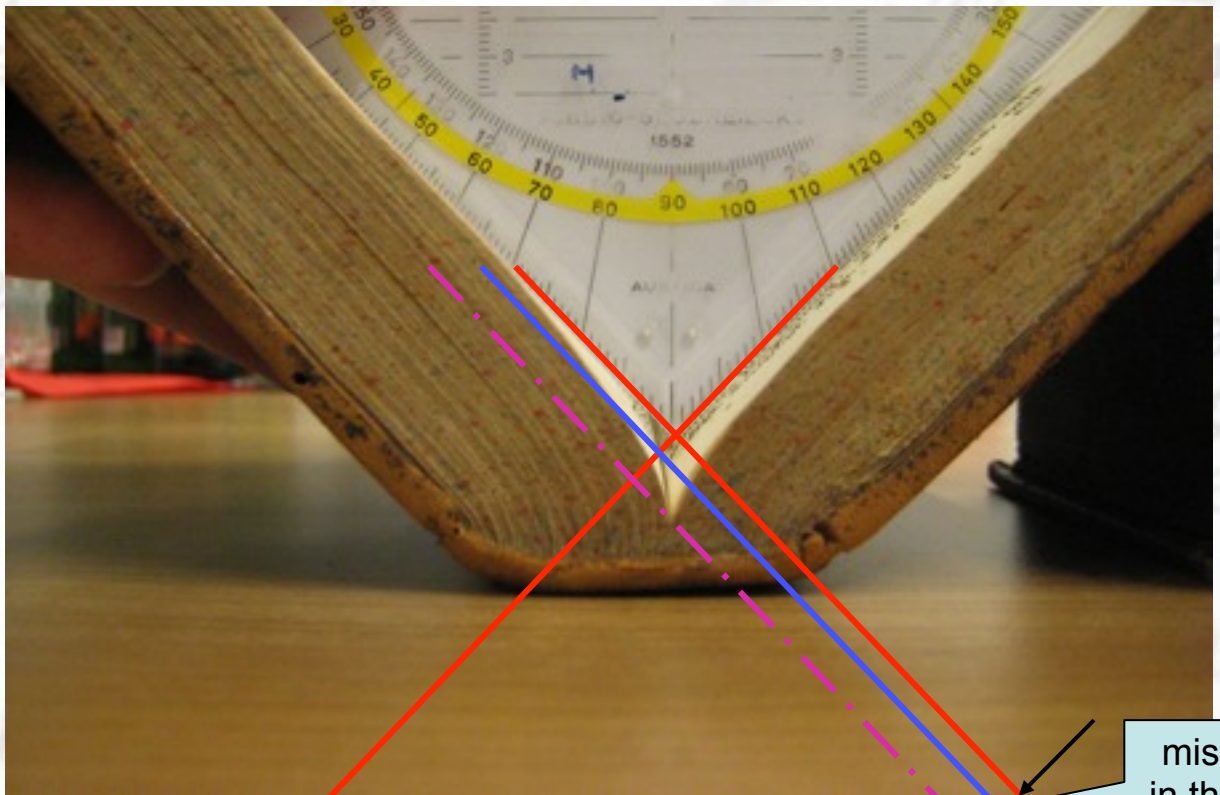
TREVENTUS

opening angle: 90° vs. 60°



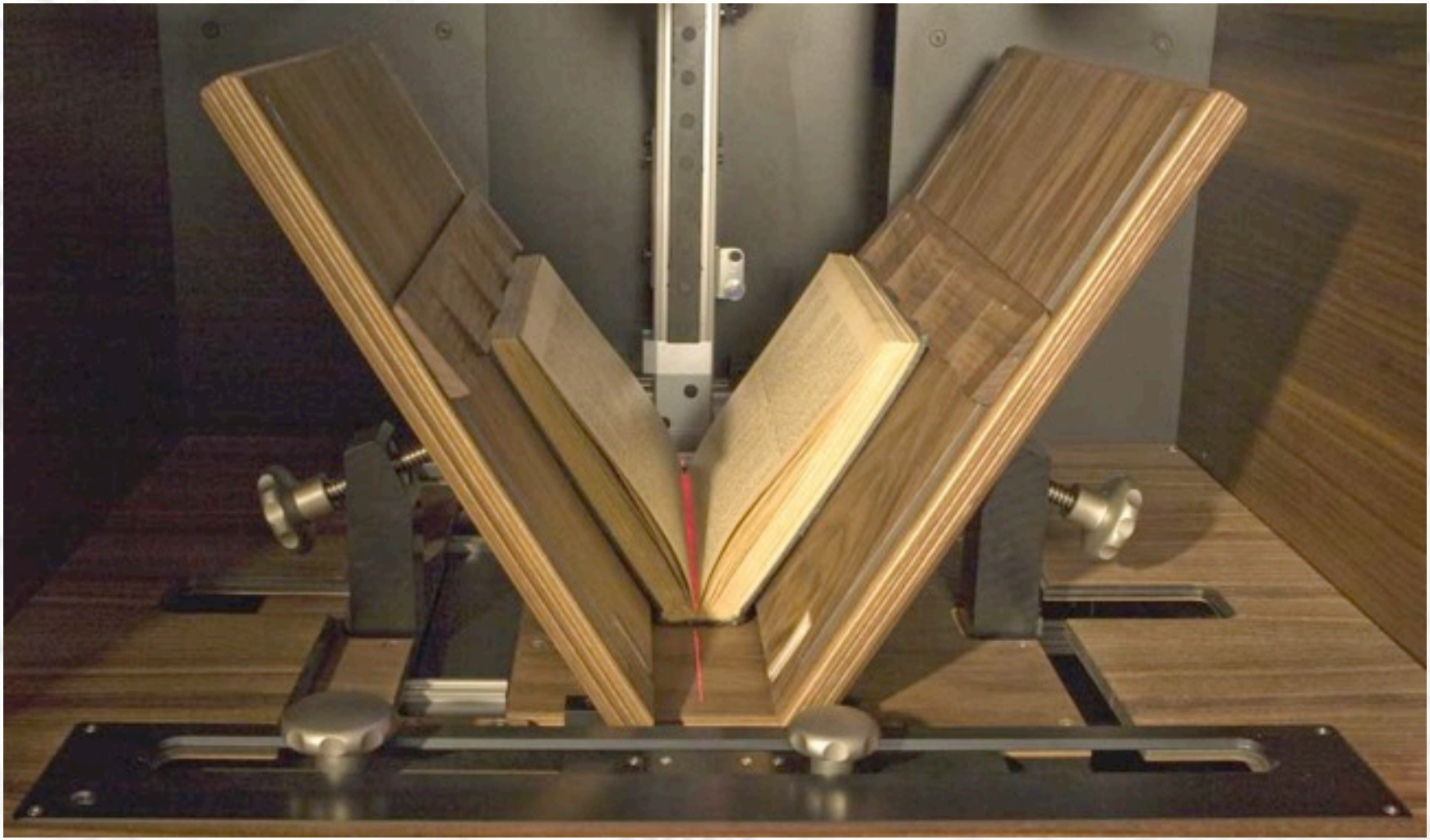
TREVENTUS

opening angle: 90° vs. 60°



TREVENTUS

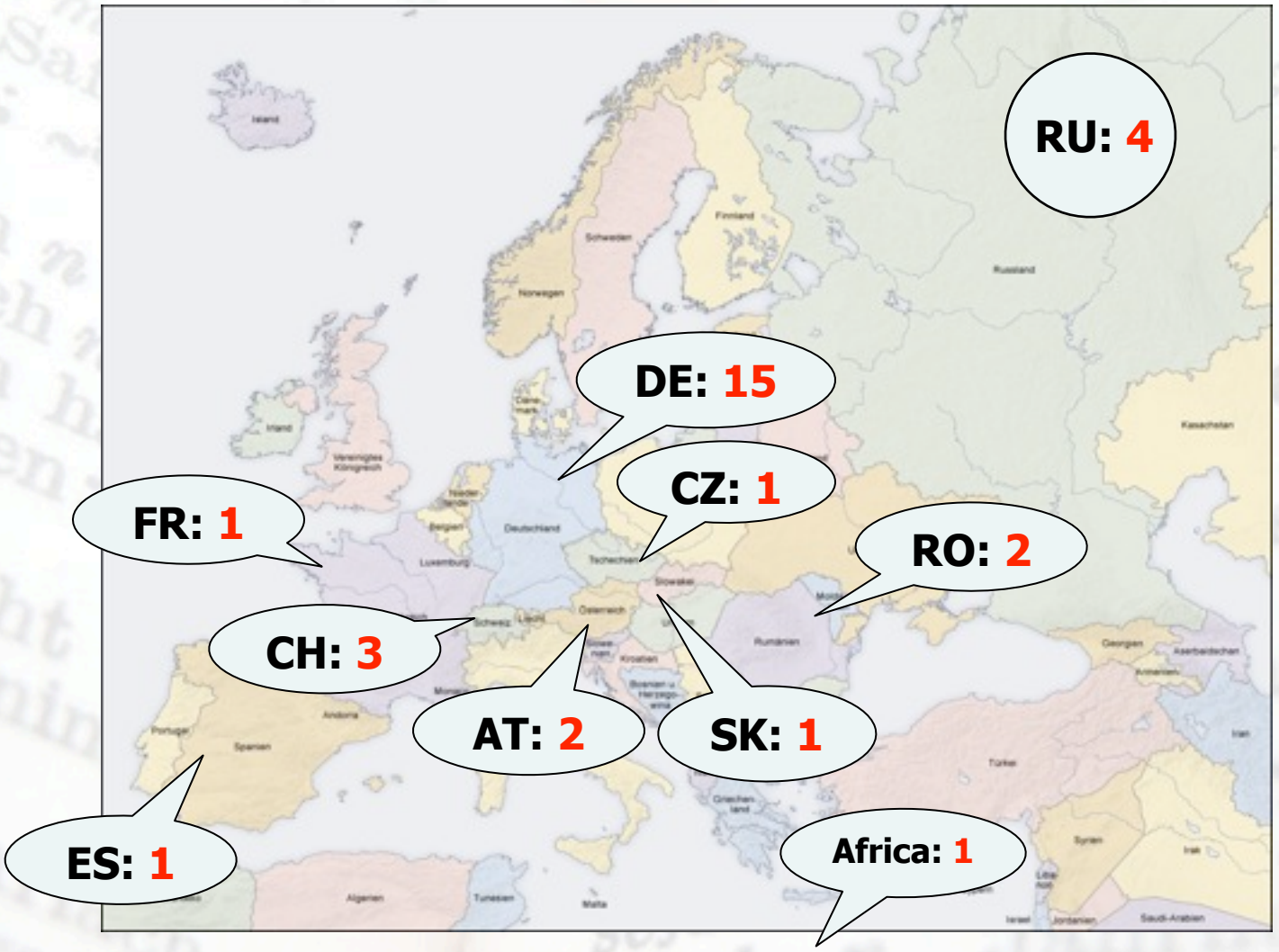
ScanRobot® - the new book cradle



opening angle: 60 degree (→ Treventus book cradle)

TREVENTUS

ScanRobot® - Europa



31 ScanRobot® sold world wide

(→ reliable technology | last update: 01.12.2009)

TREVENTUS

Treventus digitization service

best practice in house

DS: Digitization Service

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

TREVENTUS

ScanFlow™

a brief overview

TREVENTUS

Workflow - What is the principal aim?

Workflow - What is the principal aim?

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

TREVENTUS

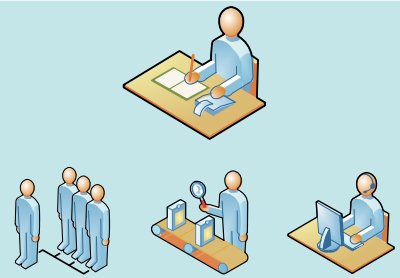
Workflow - What is the principal aim?

- Digital object of a page/book
  - images with description files
  - meta data
    - bibliografic 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)

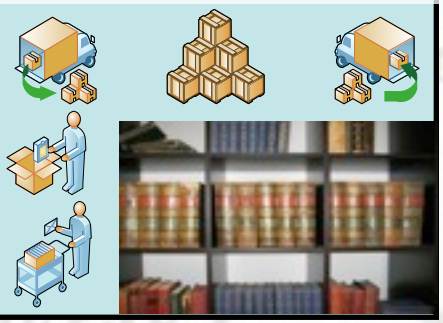
TREVENTUS

ScanFlow™ – handels the ressources

coordination - staff



books - docs - logistic



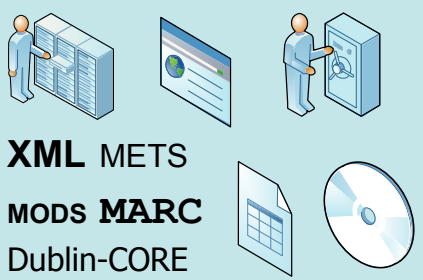
workflow-software

ScanFlow™

scan- & server-hardware

Quality-Control

files: images, meta-data



XML METS

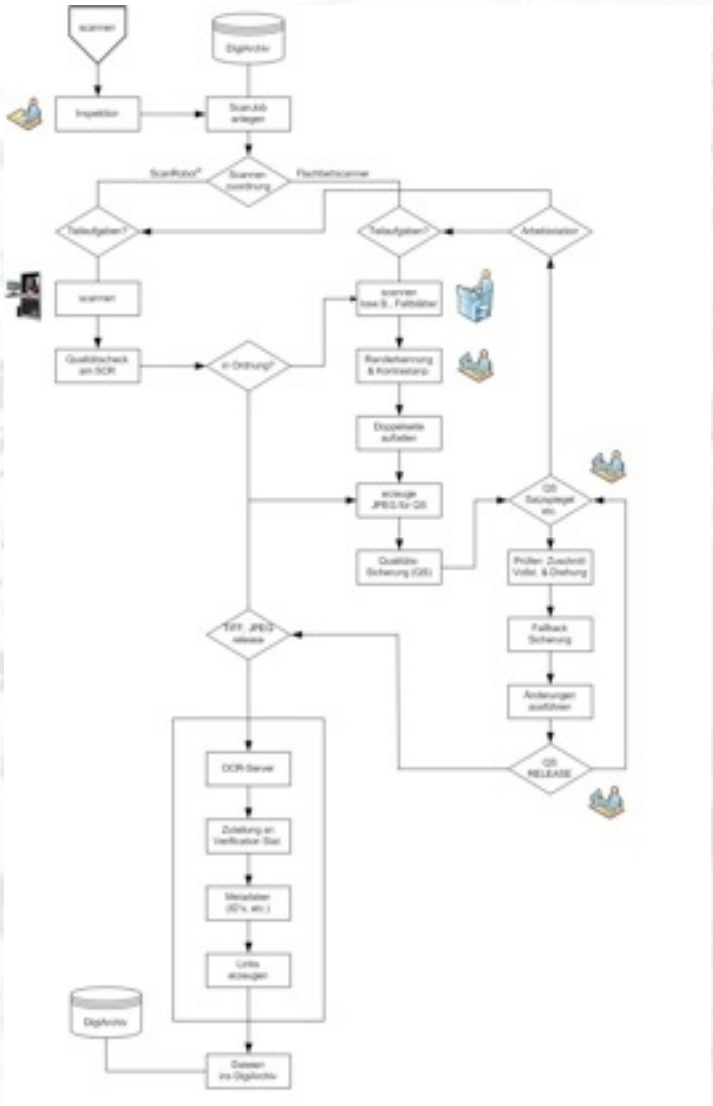
MODS **MARC**

Dublin-CORE

TREVENTUS

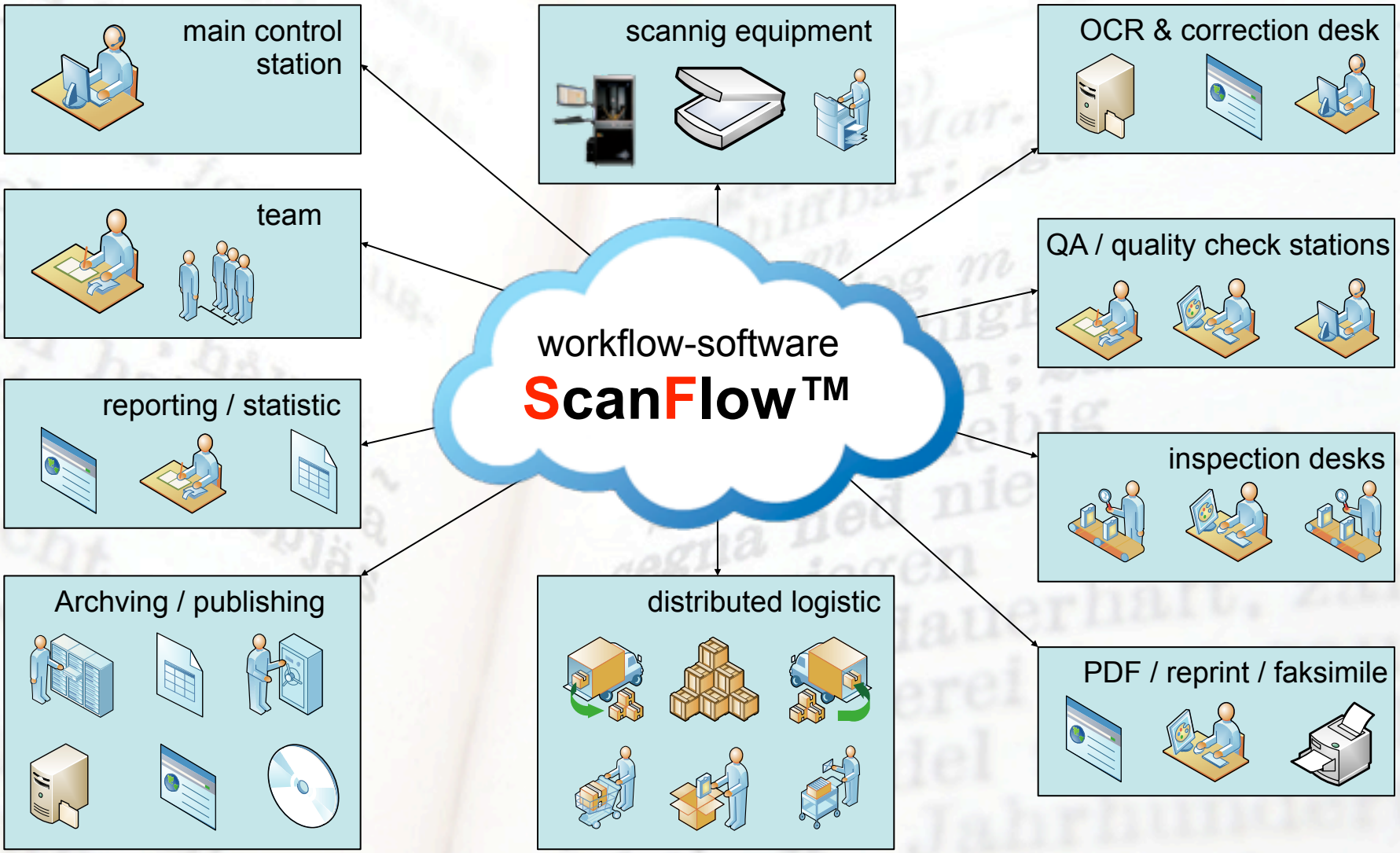
The heart of ScanFlow™?

- flow diagram
  - organized work
  - structure of the enviroment
  - tasks and decisions
- statistic
  - understanding and optimization
- monitoring
  - Is all ready, up and running?



TREVENTUS

ScanFlow™: Ressource Management



TREVENTUS

ScanFlow™: organizing resources - strategy

?

archivation server
rep.ulib.sk

doc. deliv. server
mvs.ulib.sk

workflow server
TREVENTUS

OCR server
TREVENTUS

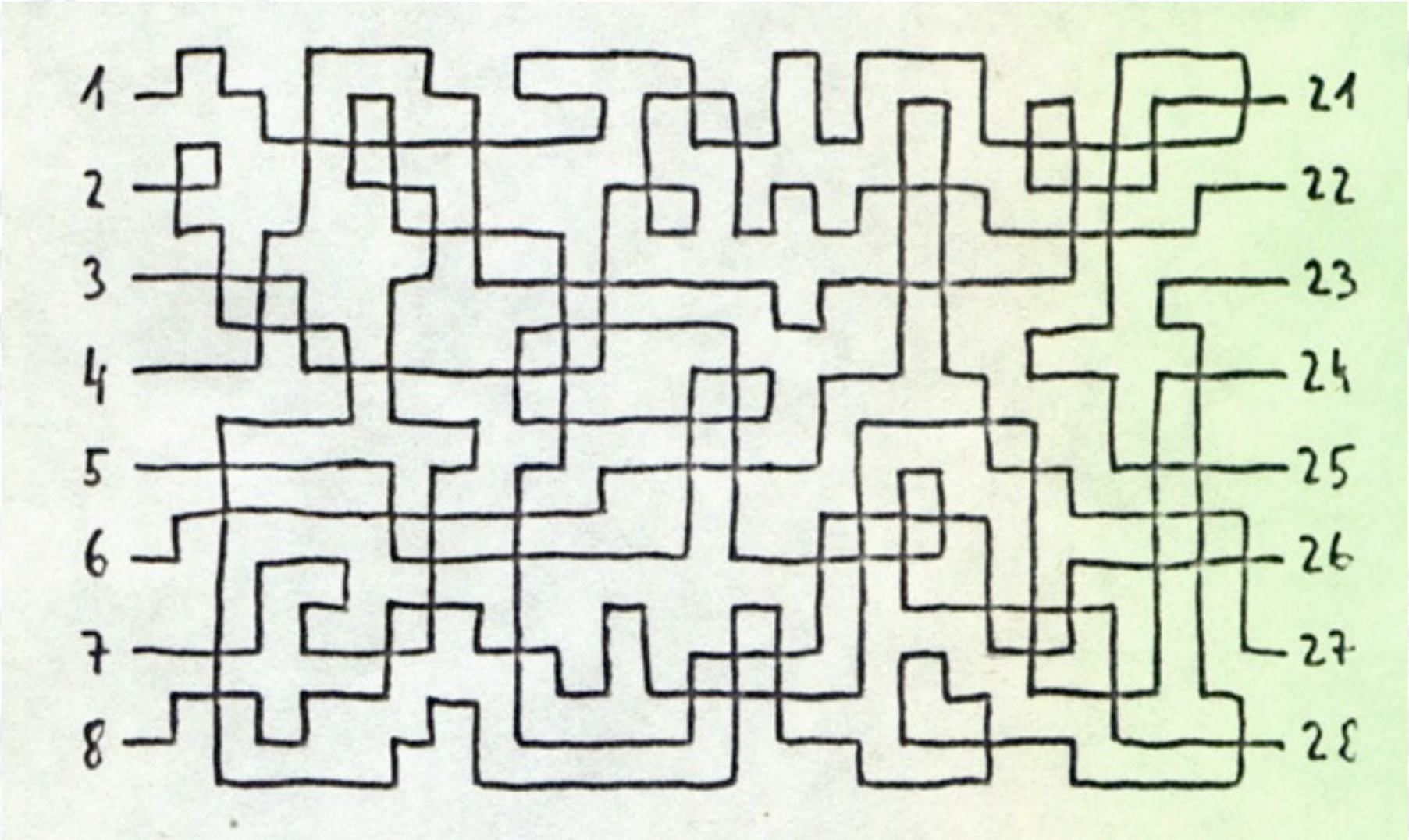
archive DVD service
TREVENTUS

administrator
UKB

work storage

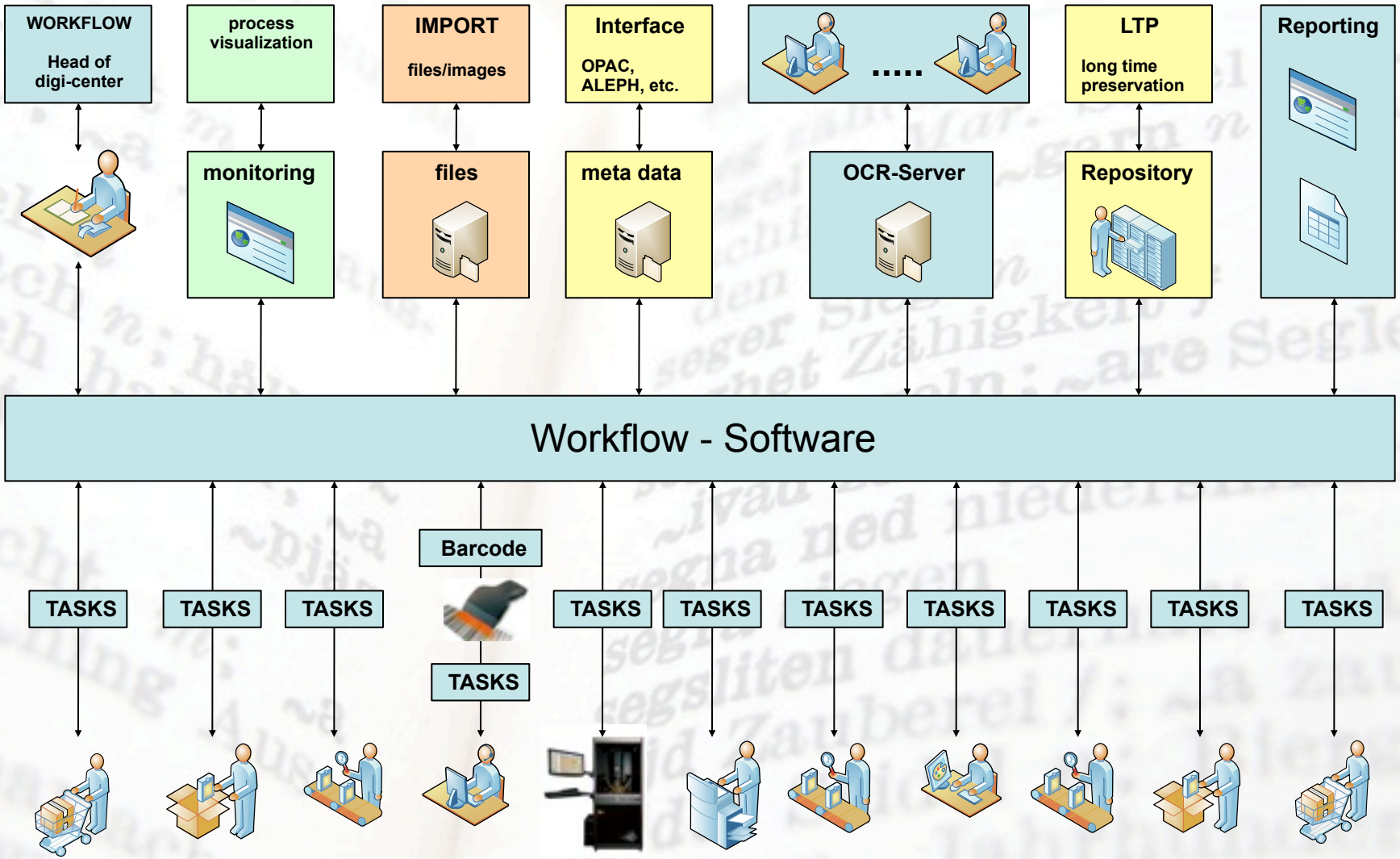
TREVENTUS

ScanFlow™: organizing resources - strategy



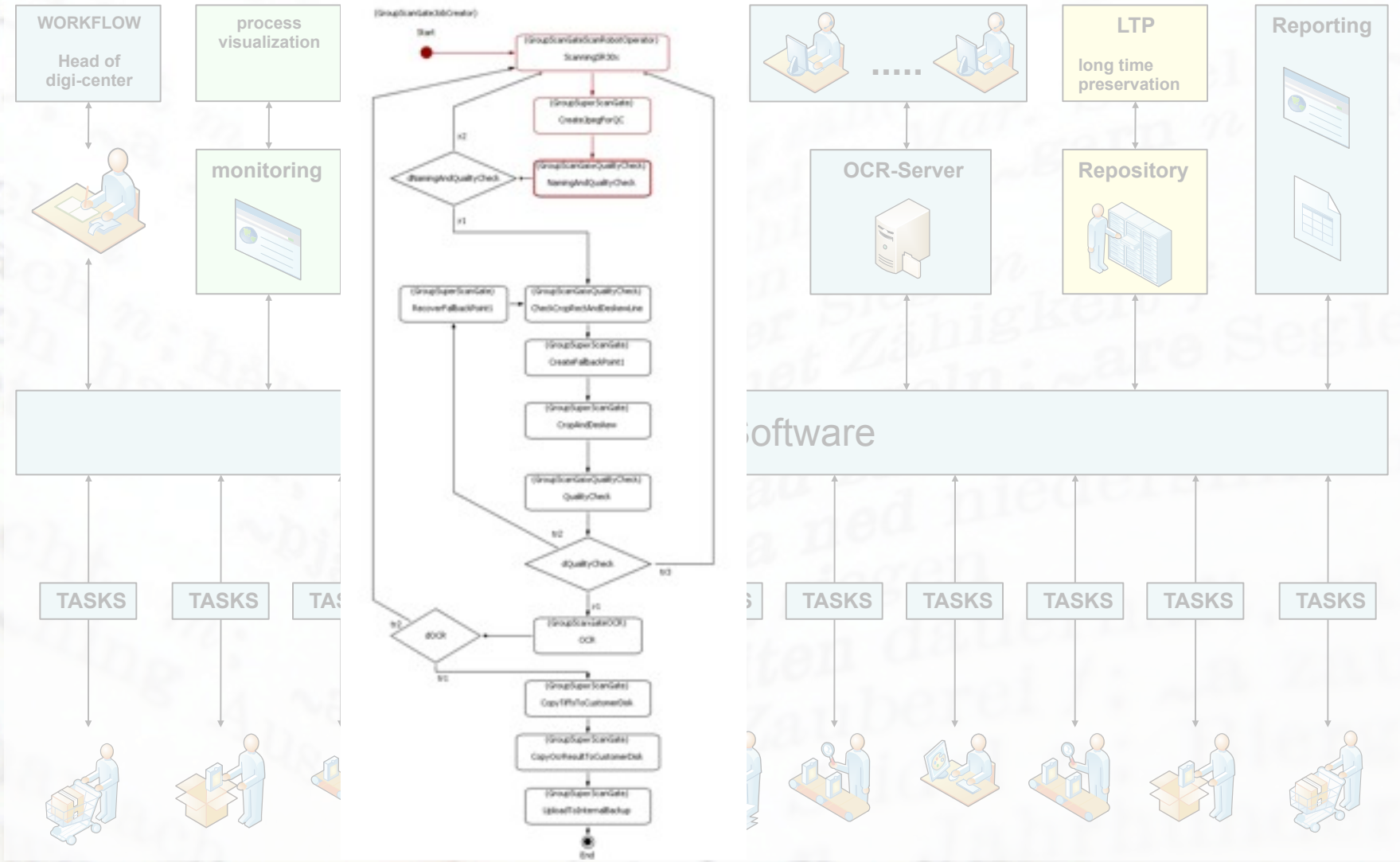
TREVENTUS

ScanFlow™: Management of resources



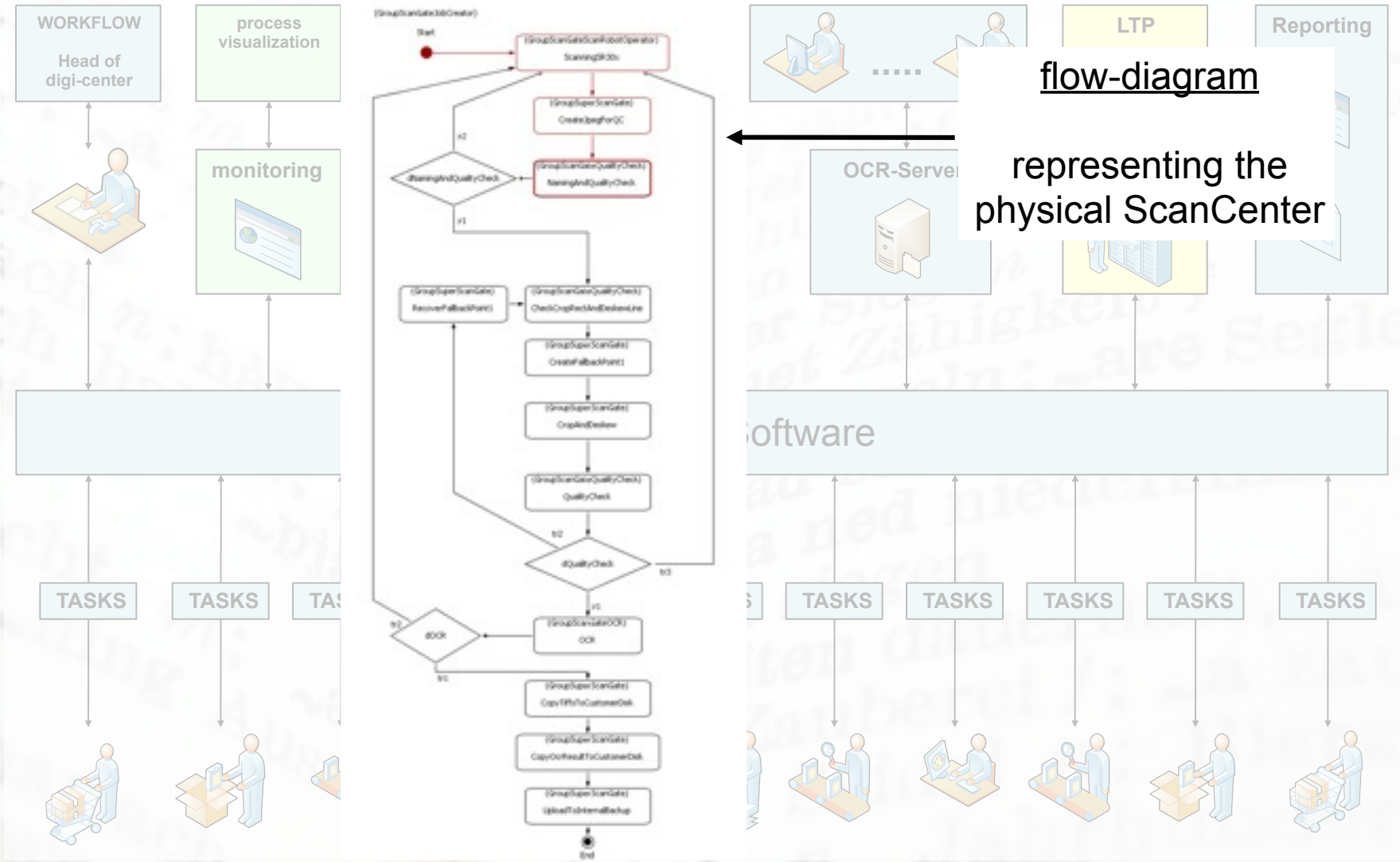
TREVENTUS

ScanFlow™: Management of resources



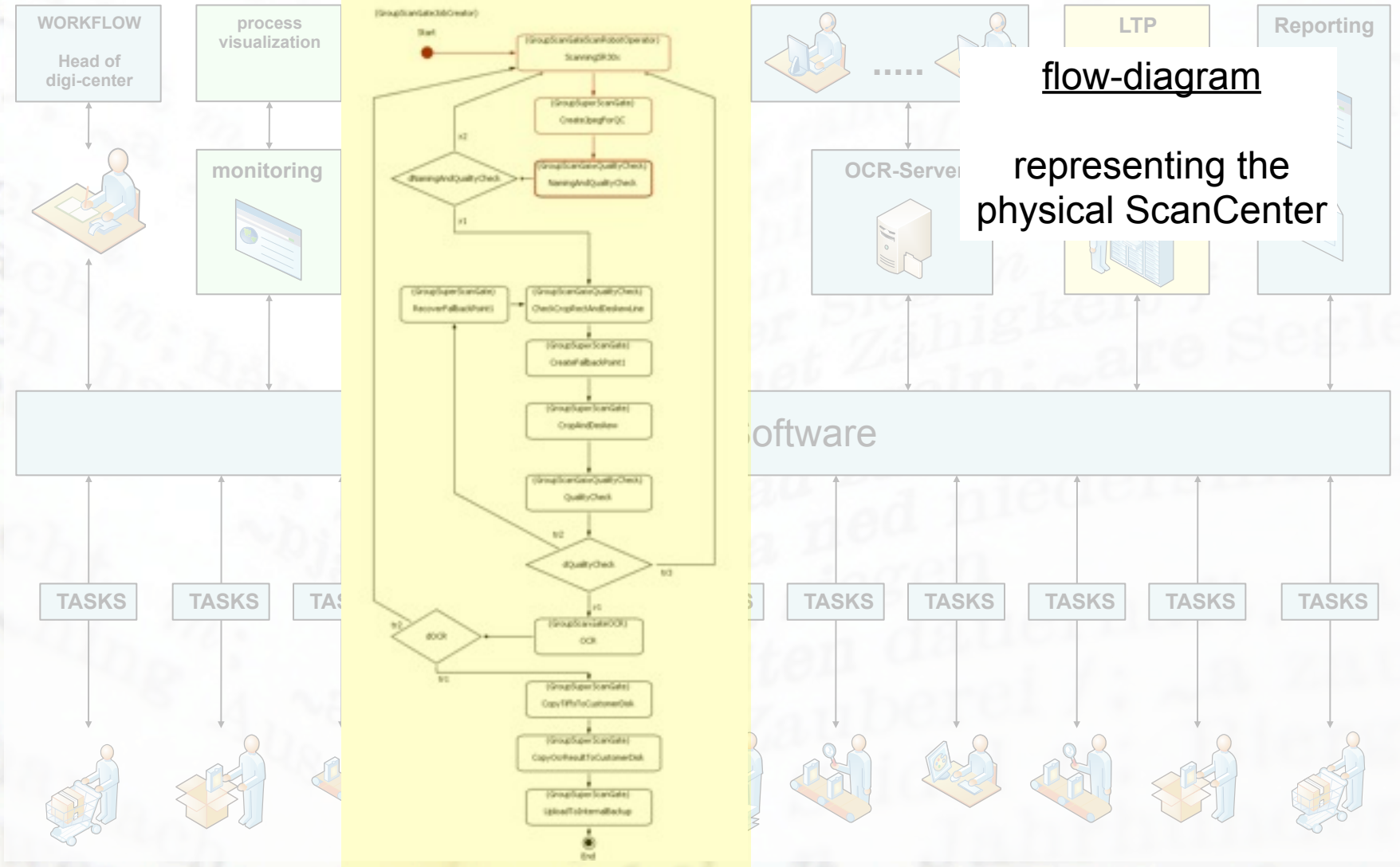
TREVENTUS

ScanFlow™: Management of resources

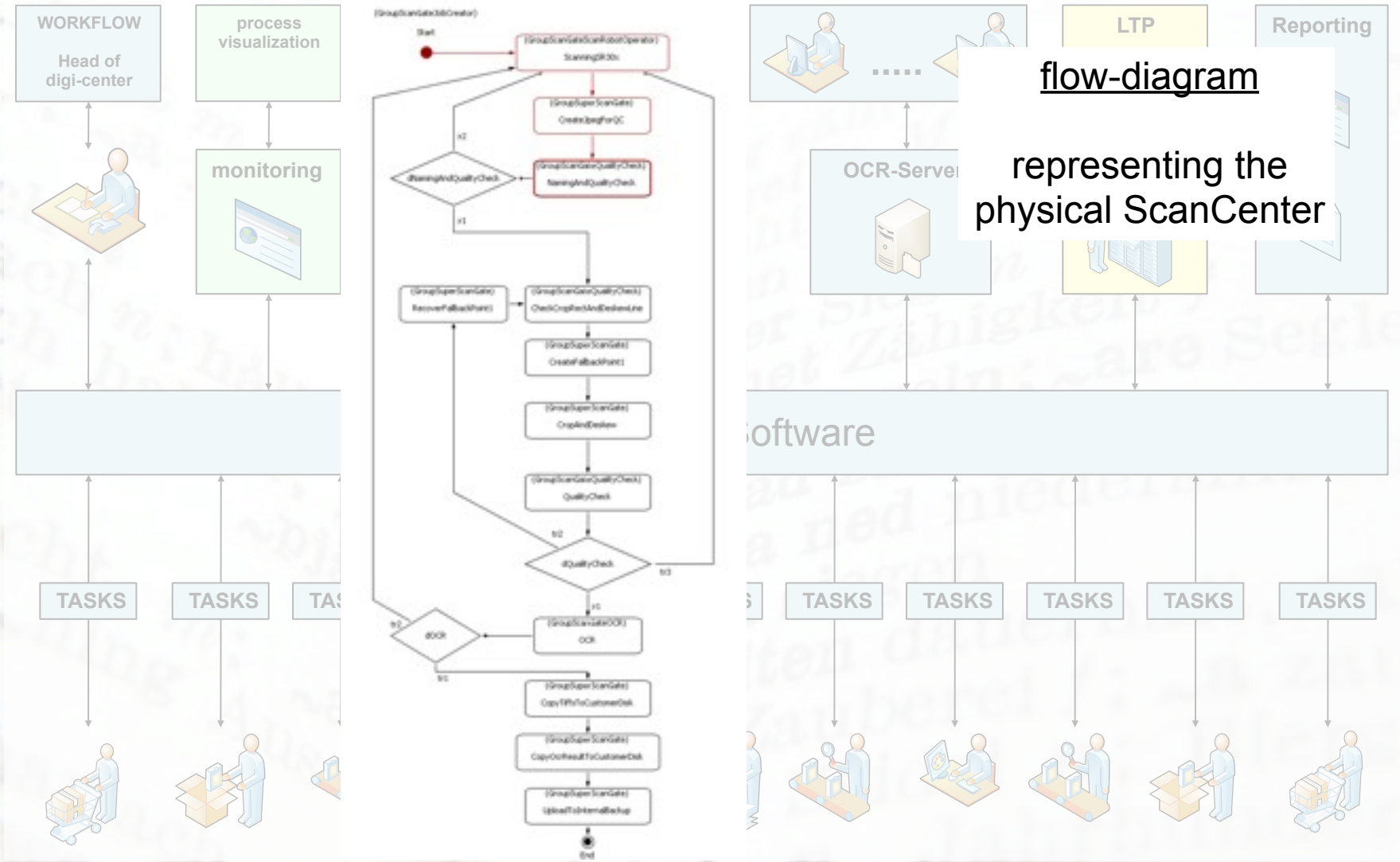


TREVENTUS

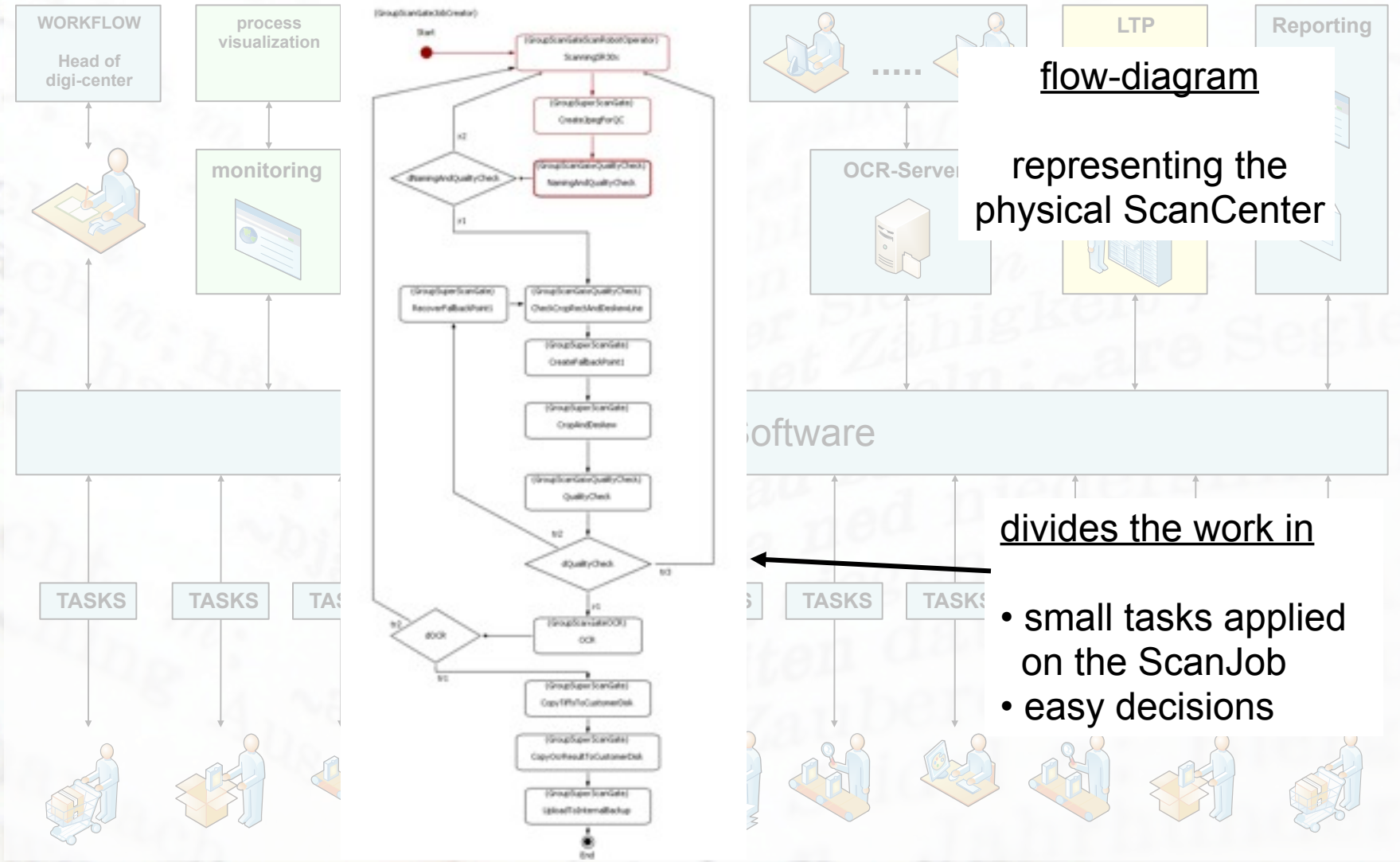
ScanFlow™: Management of resources



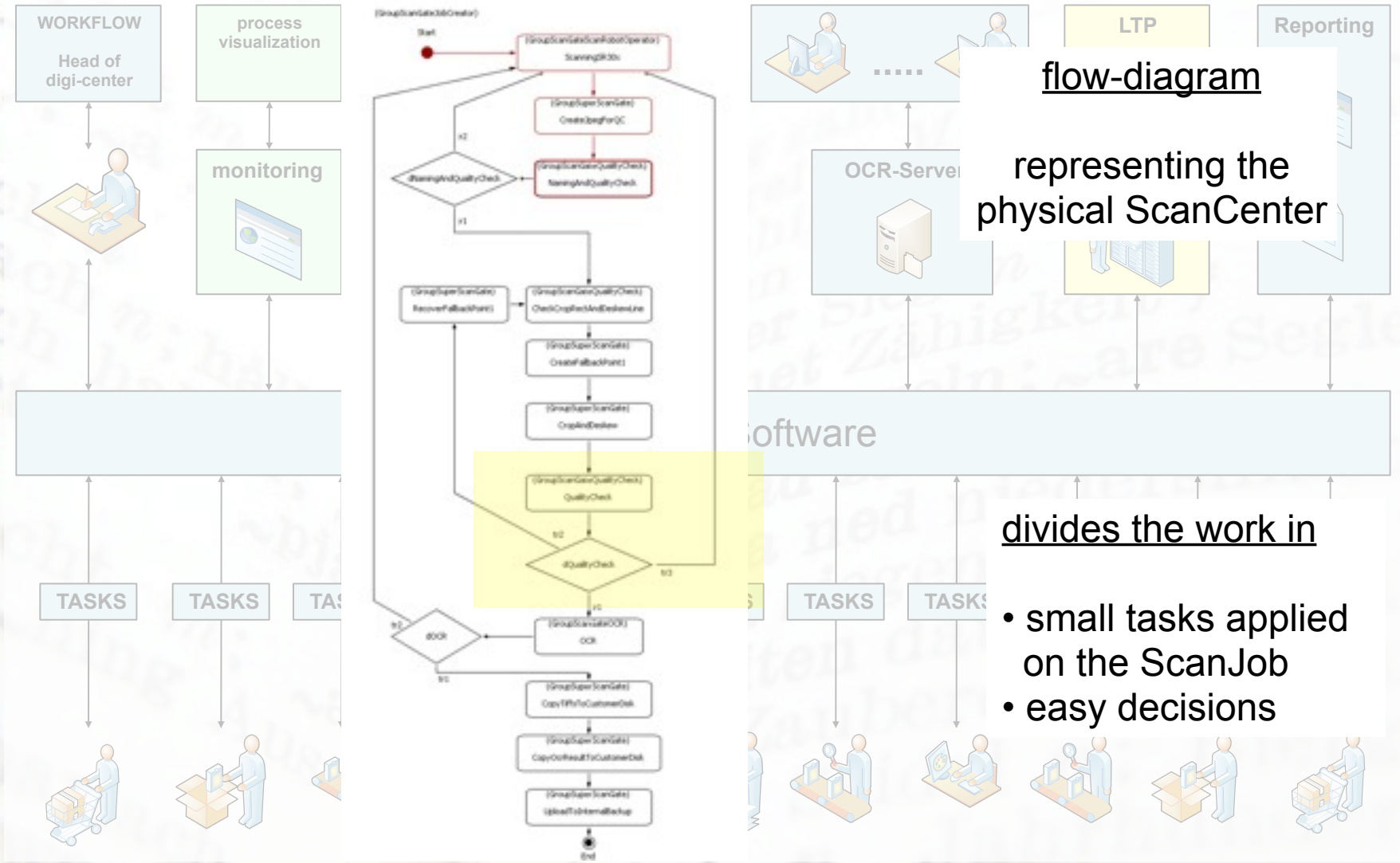
ScanFlow™: Management of resources



ScanFlow™: Management of resources



ScanFlow™: Management of resources



flow-diagram

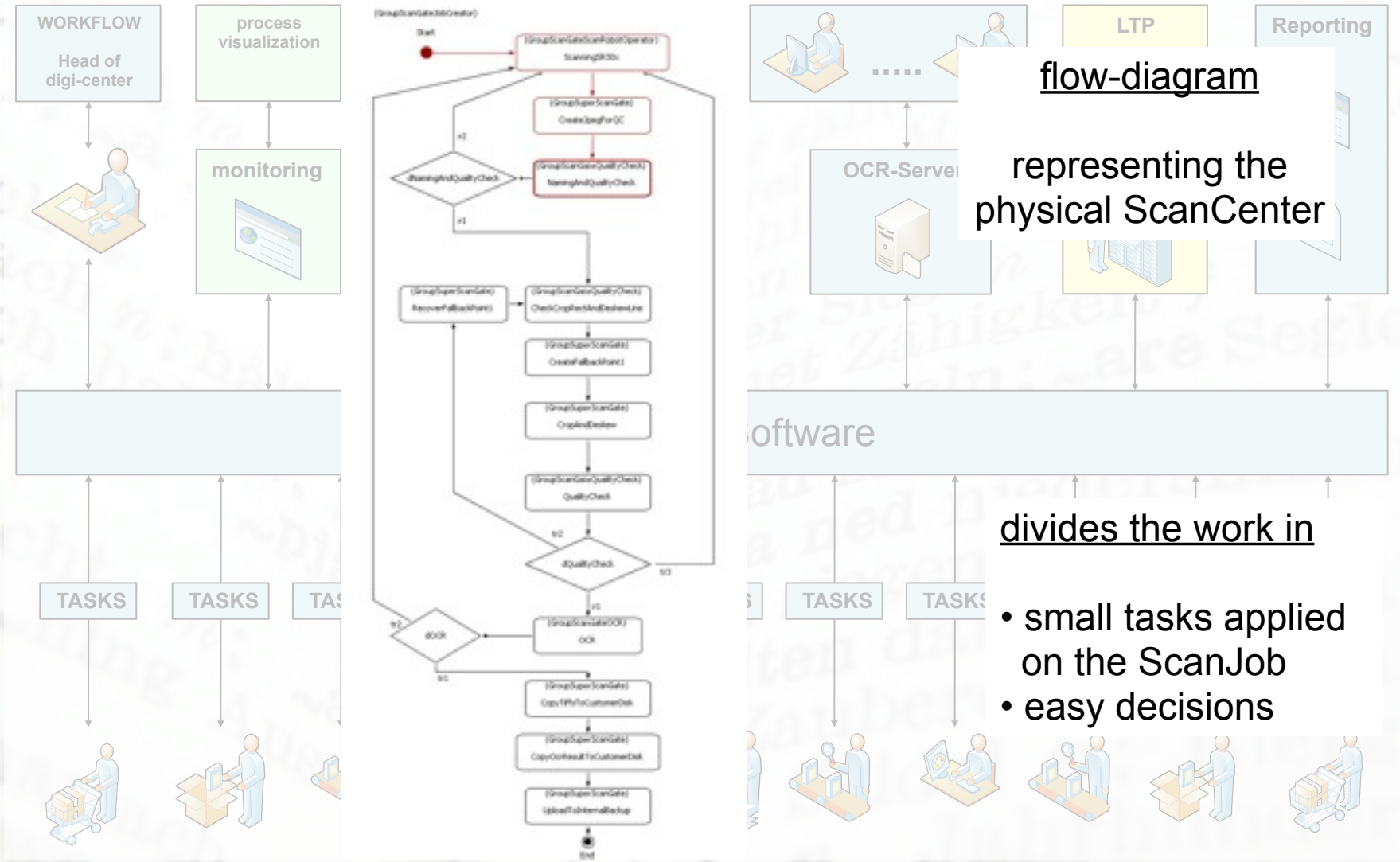
representing the physical ScanCenter

divides the work in

- small tasks applied on the ScanJob
- easy decisions

TREVENTUS

ScanFlow™: Management of resources



flow-diagram

representing the physical ScanCenter

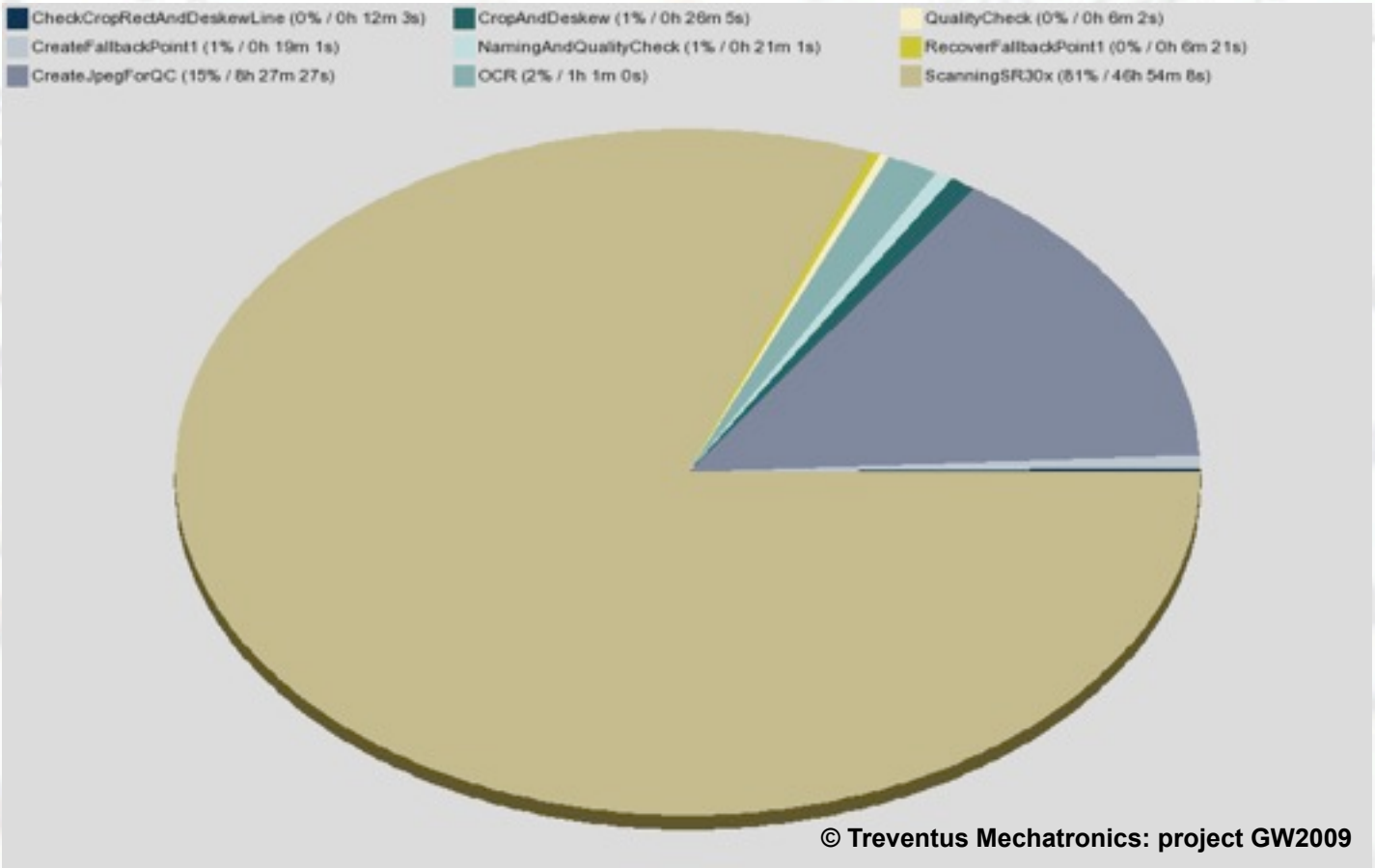
divides the work in

- small tasks applied on the ScanJob
- easy decisions

TREVENTUS

ScanFlow™: Statistic Analysis

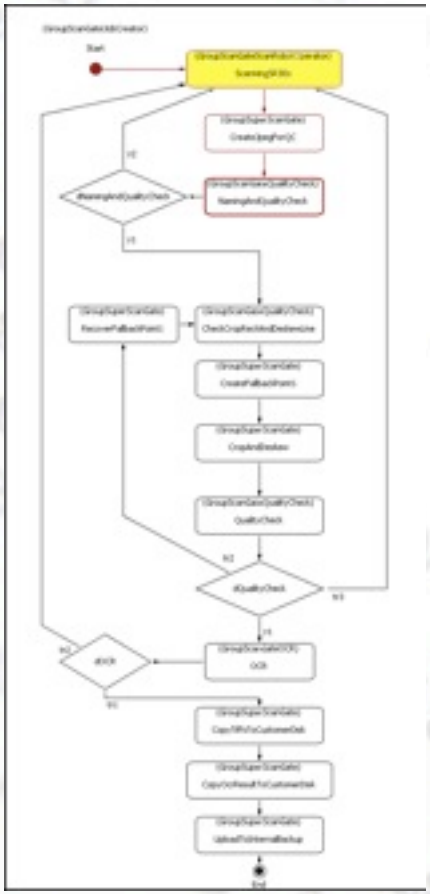
- Tasks listed in a **pie chart**
- to understand the **bottlenecks** & to see possible **optimizations**

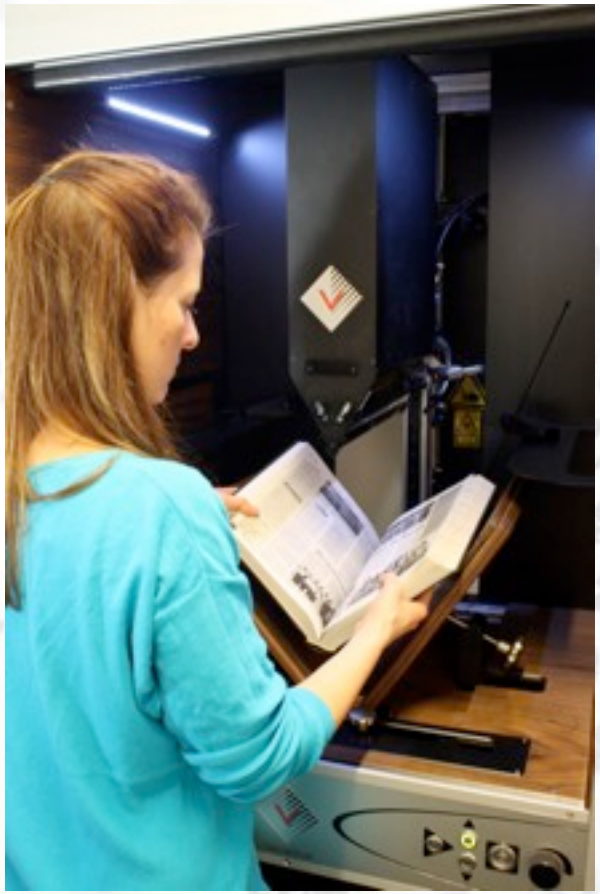


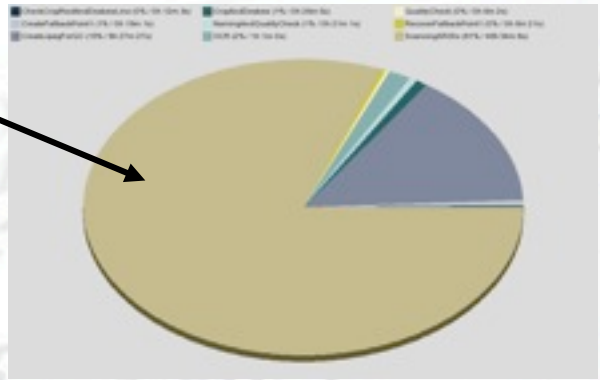
© Treventus Mechatronics: project GW2009

TREVENTUS

ScanFlow™: Statistic Analysis I

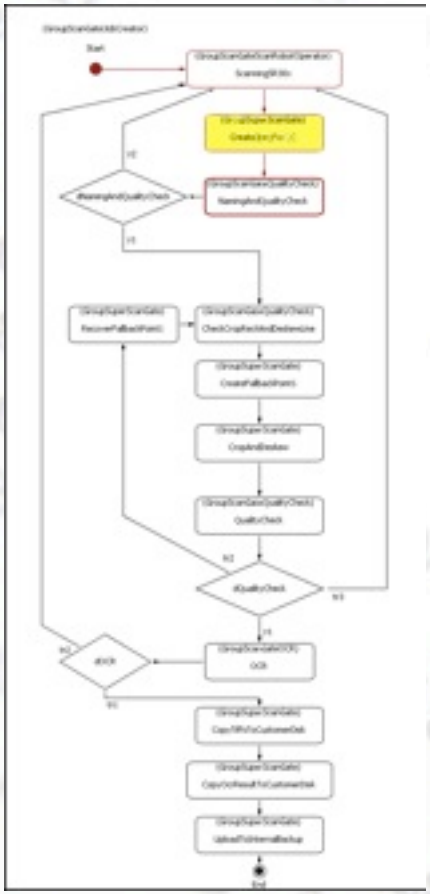


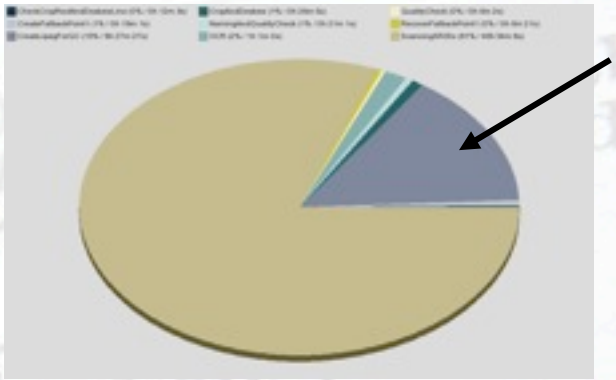




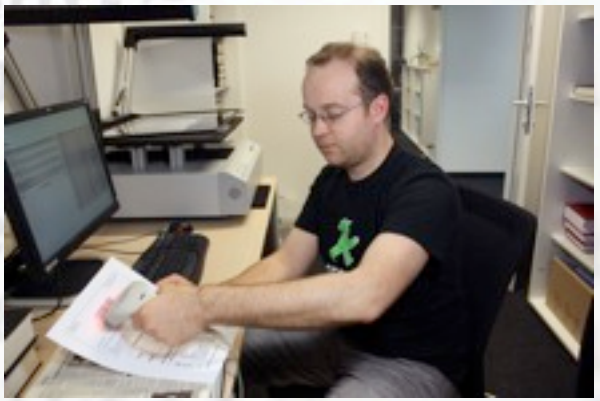
TASK: Scanning

ScanFlow™: Statistic Analysis I



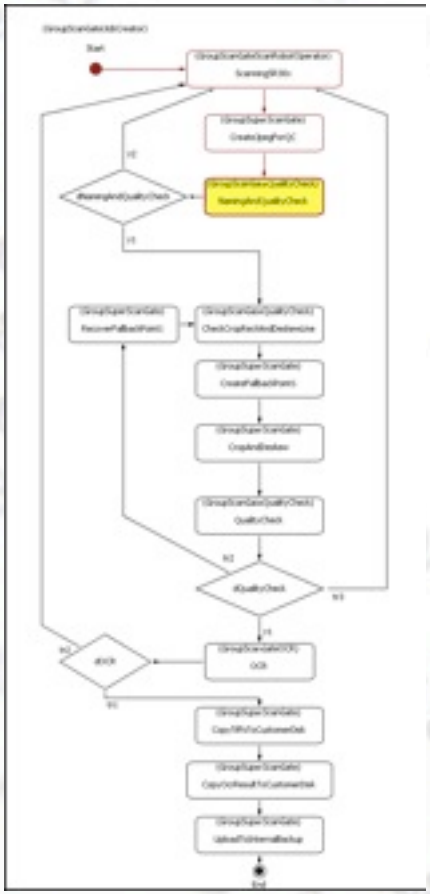


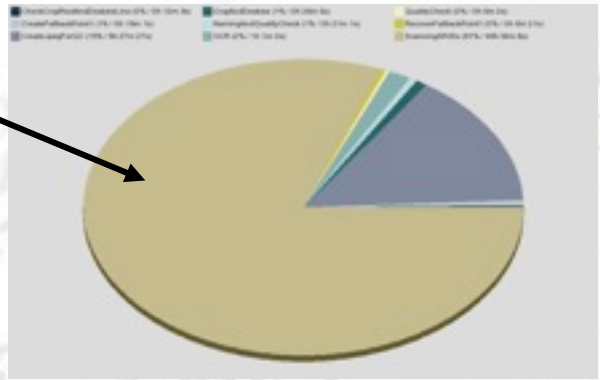
TASK: Quality Check 1



TREVENTUS

ScanFlow™: Statistic Analysis I

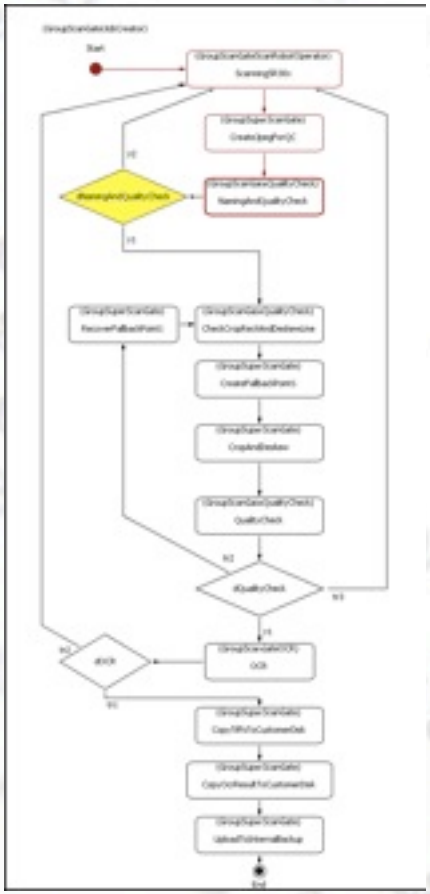


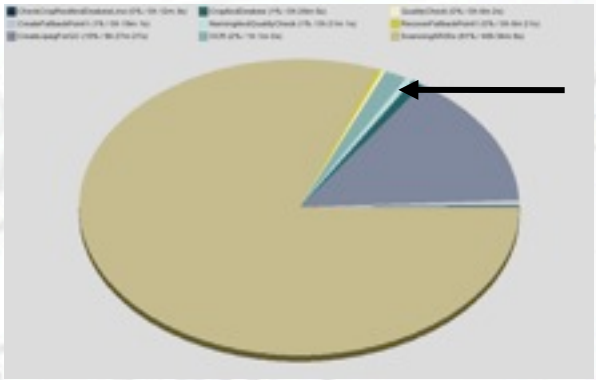


TASK: renaming the pages

TREVENTUS

ScanFlow™: Statistic Analysis I

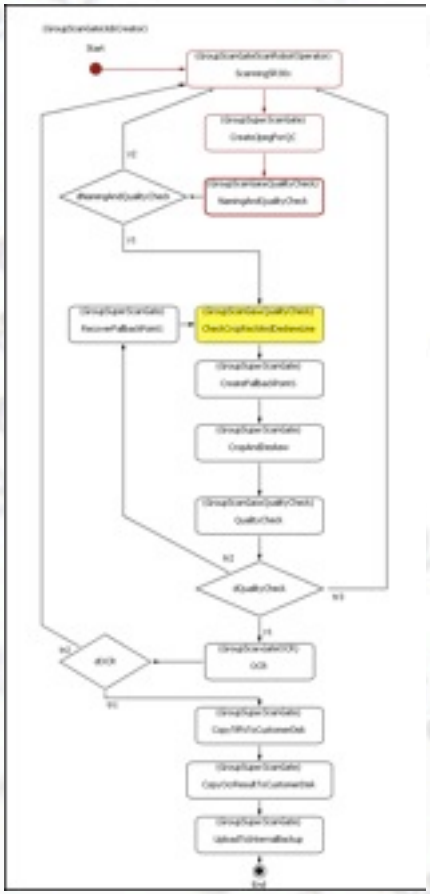


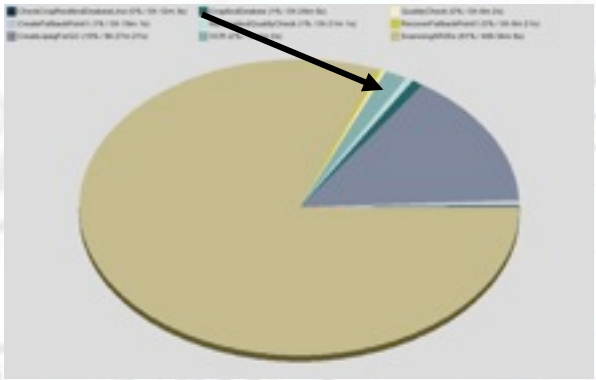


TASK: Quality Check 2

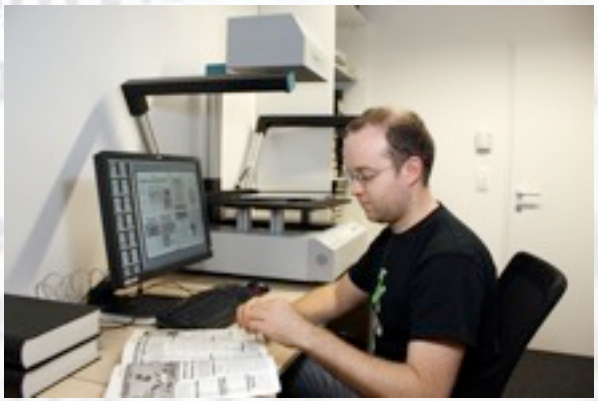
TREVENTUS

ScanFlow™: Statistic Analysis I



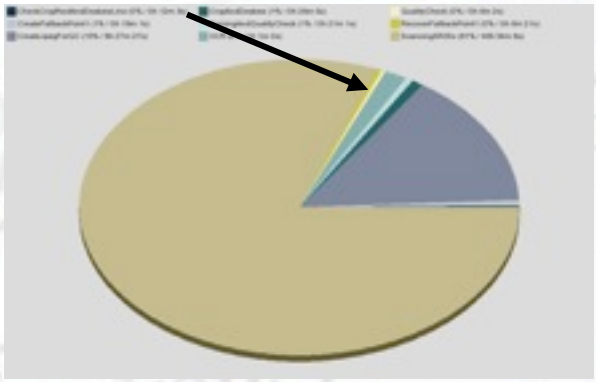


TASK: check crop & deskew



TREVENTUS

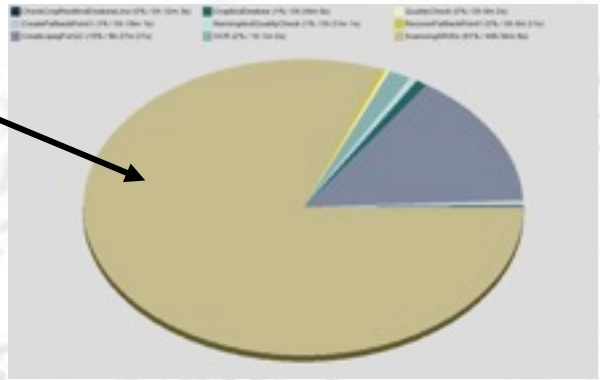
ScanFlow™: Statistic Analysis I



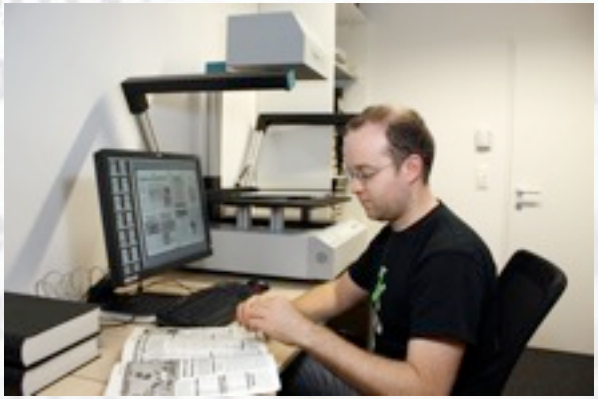
TASK: create fallback point

TREVENTUS

ScanFlow™: Statistic Analysis I

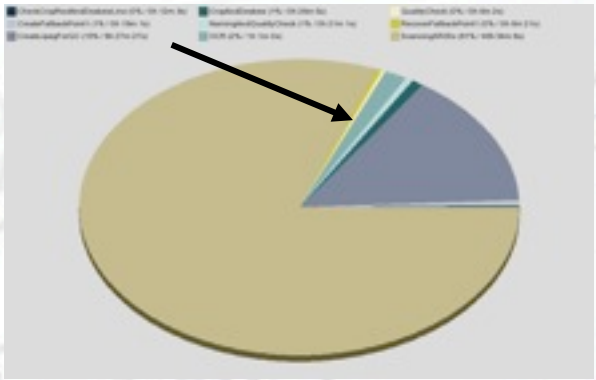


TASK: adjust crop & deskew



TREVENTUS

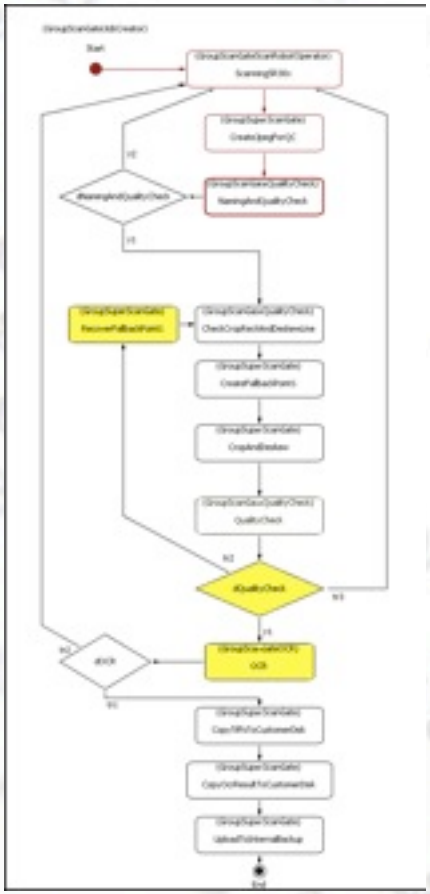
ScanFlow™: Statistic Analysis I

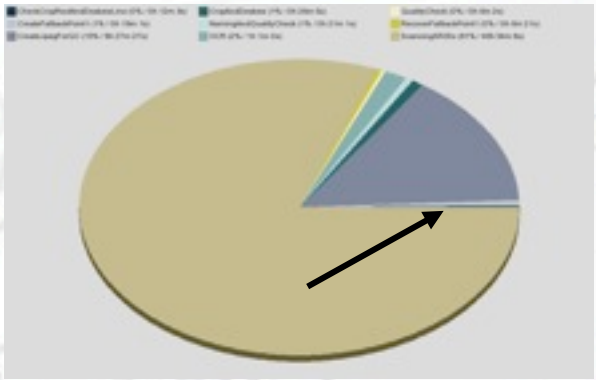


TASK: Quality Check 3

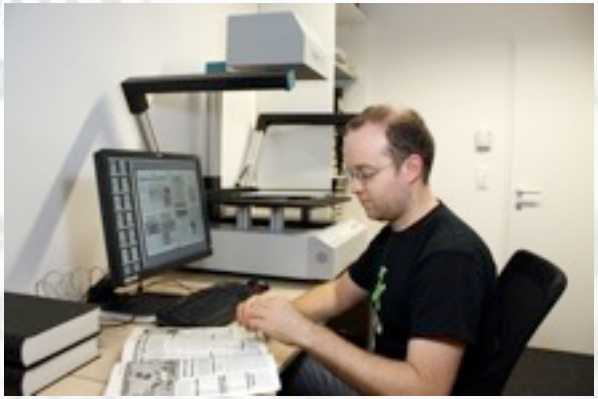
TREVENTUS

ScanFlow™: Statistic Analysis I



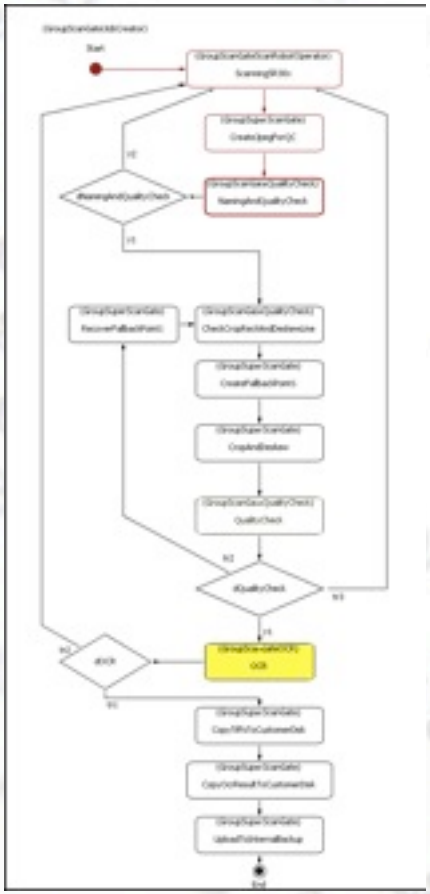


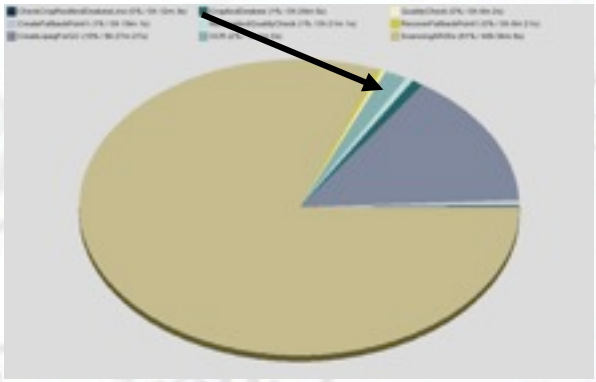
TASK: OCR or recover FBP1?



TREVENTUS

ScanFlow™: Statistic Analysis I

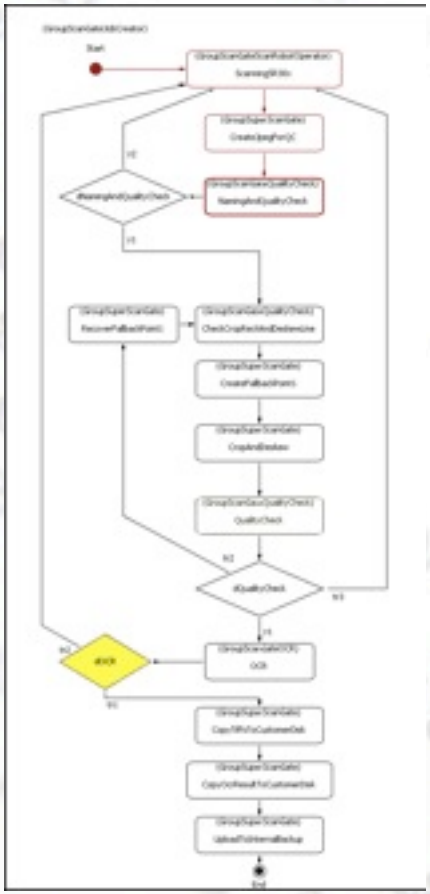


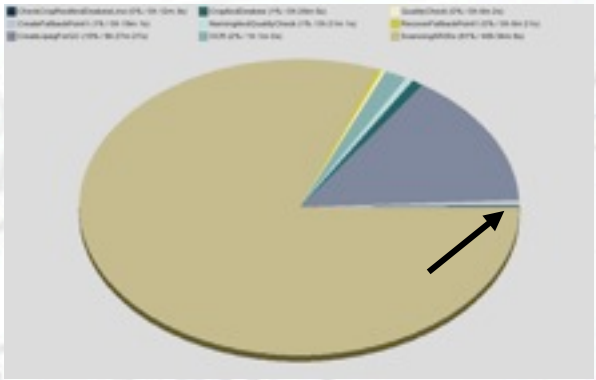


TASK: OCR is running... (auto)

TREVENTUS

ScanFlow™: Statistic Analysis I

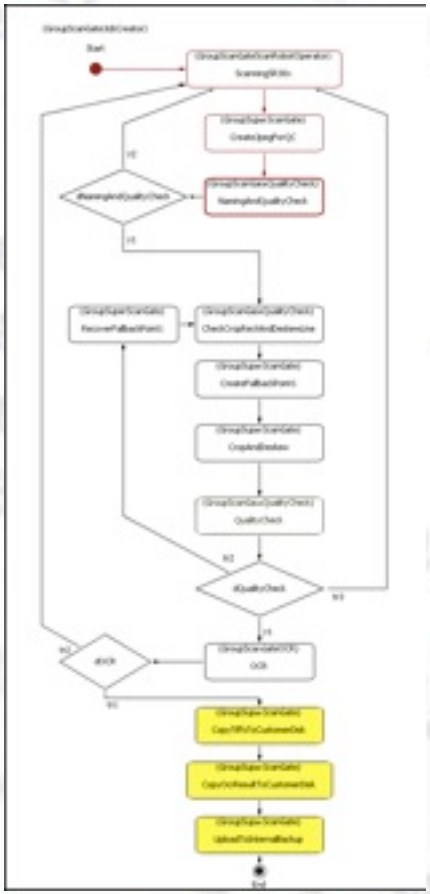


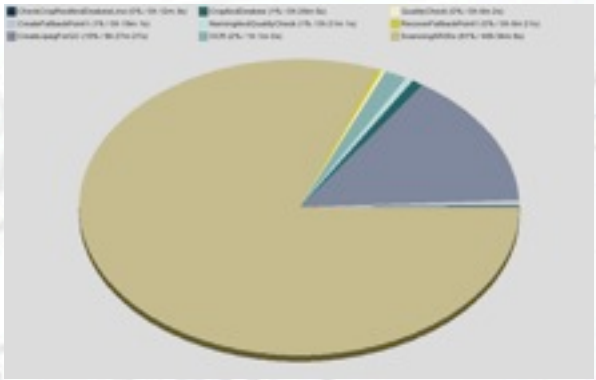


TASK: OCR Quality Check

TREVENTUS

ScanFlow™: Statistic Analysis I

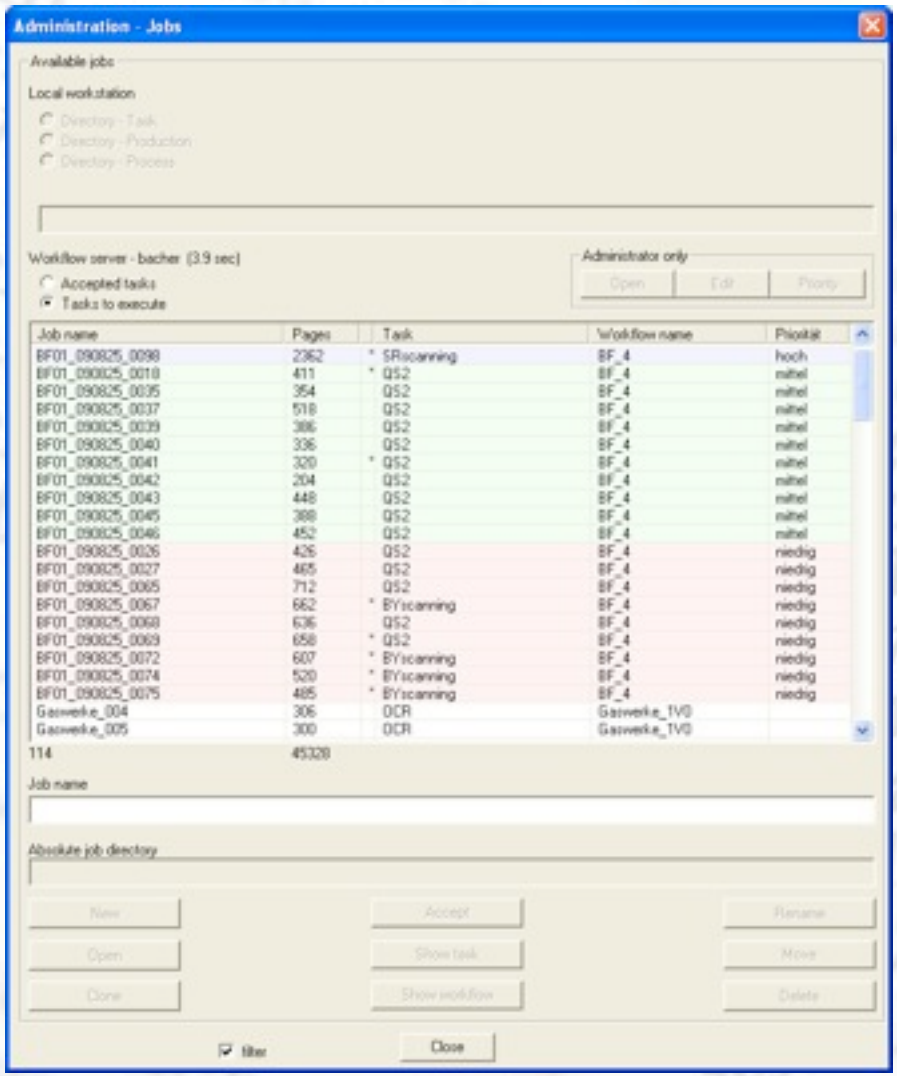




TASK: server tasks (i.e. copy)

TREVENTUS

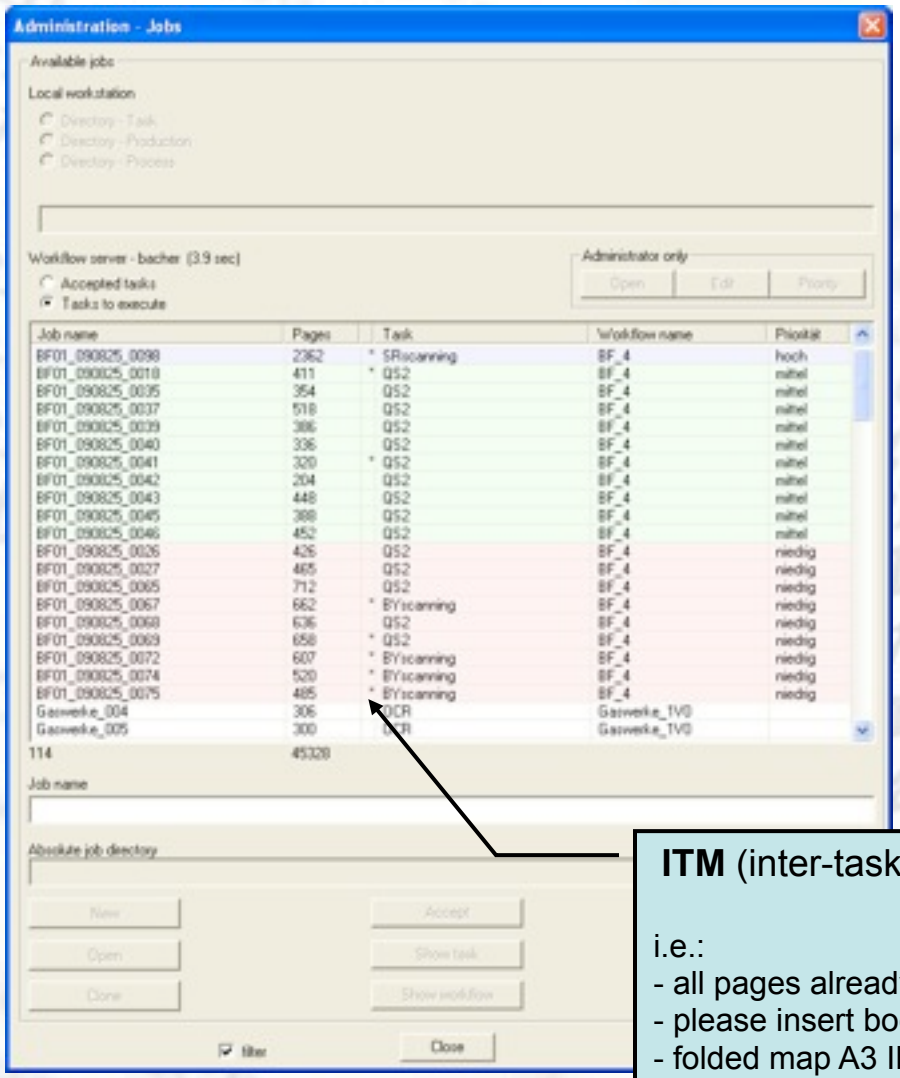
ScanFlow™: Job administration



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

TREVENTUS

ScanFlow™: Job administration



ITM (inter-task message)

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

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

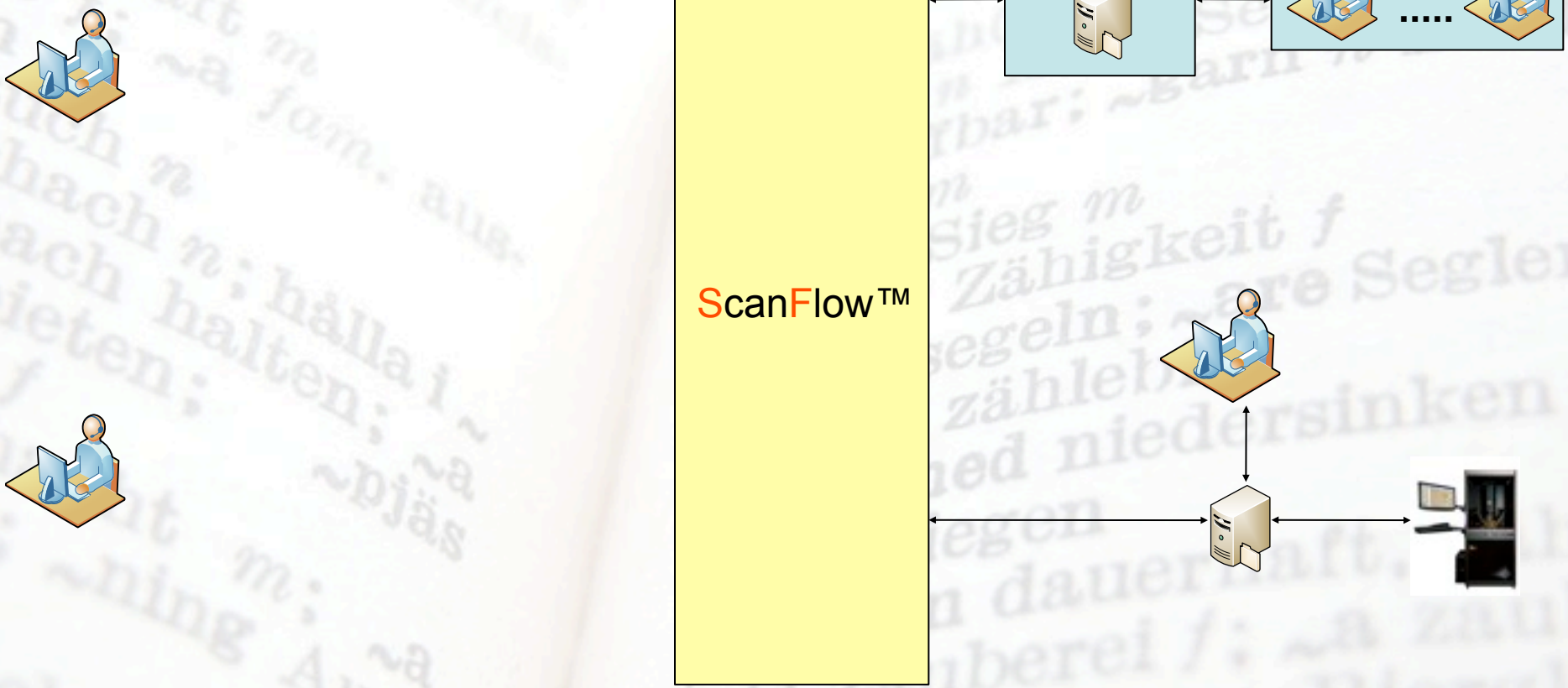
ITM

deskew = ok
cropping = ok
1 folded map (middle)

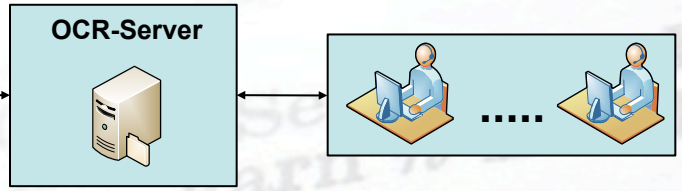
ask Pablof!

TREVENTUS

ScanFlow™: Example



ScanFlow™: Example



TREVENTUS

ScanFlow™: Example

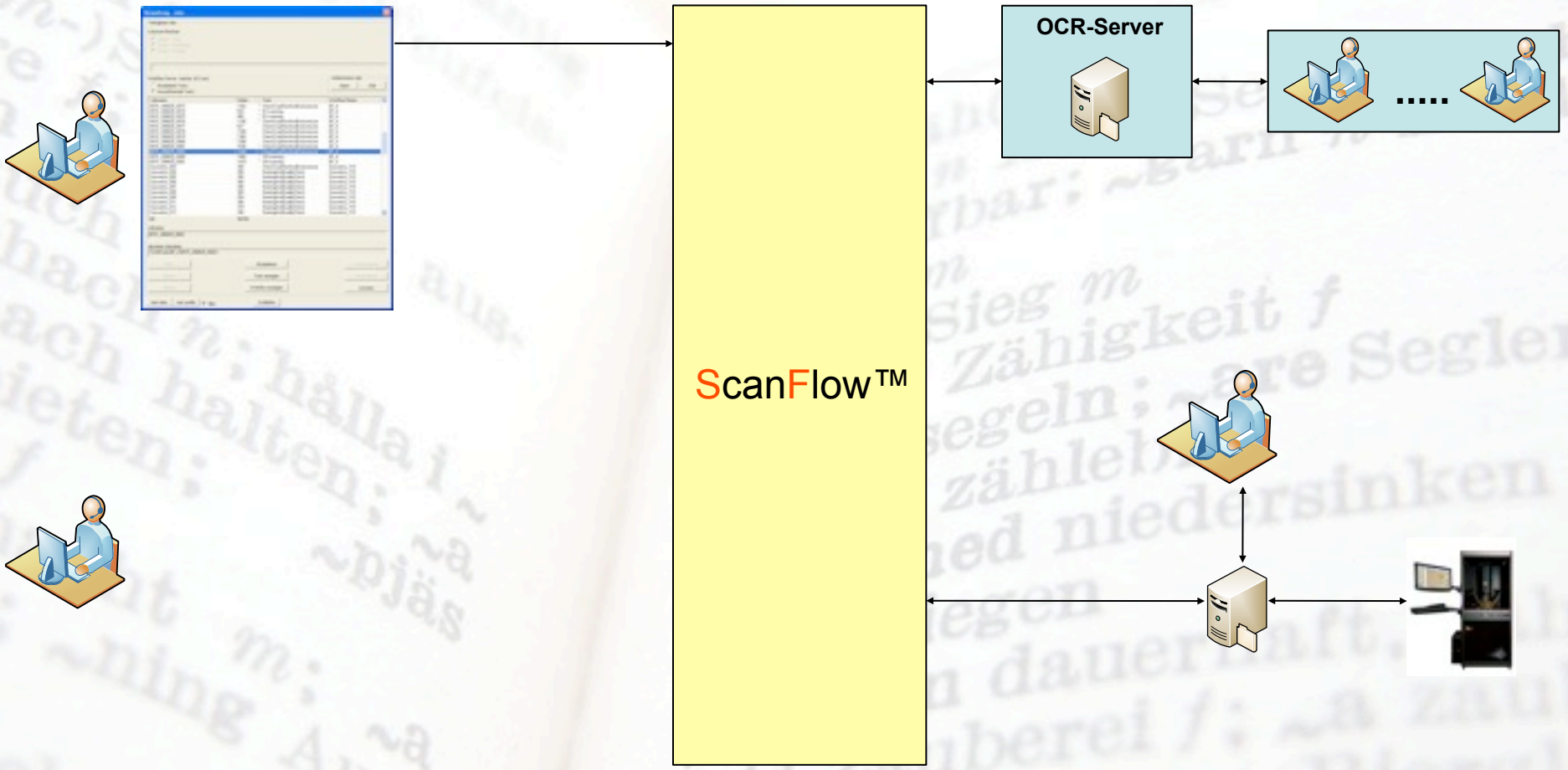
Operator 1 login

ScanFlow™

OCR-Server

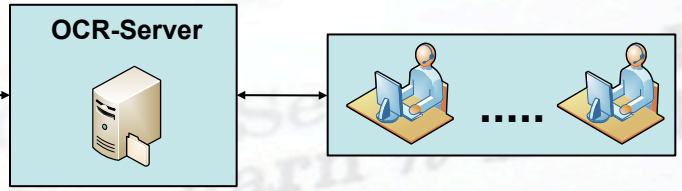
TREVENTUS

ScanFlow™: Example



TREVENTUS

ScanFlow™: Example

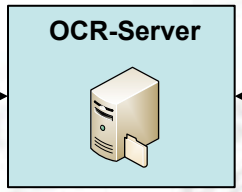


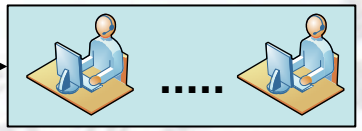
TREVENTUS

ScanFlow™: Example

List of all jobs

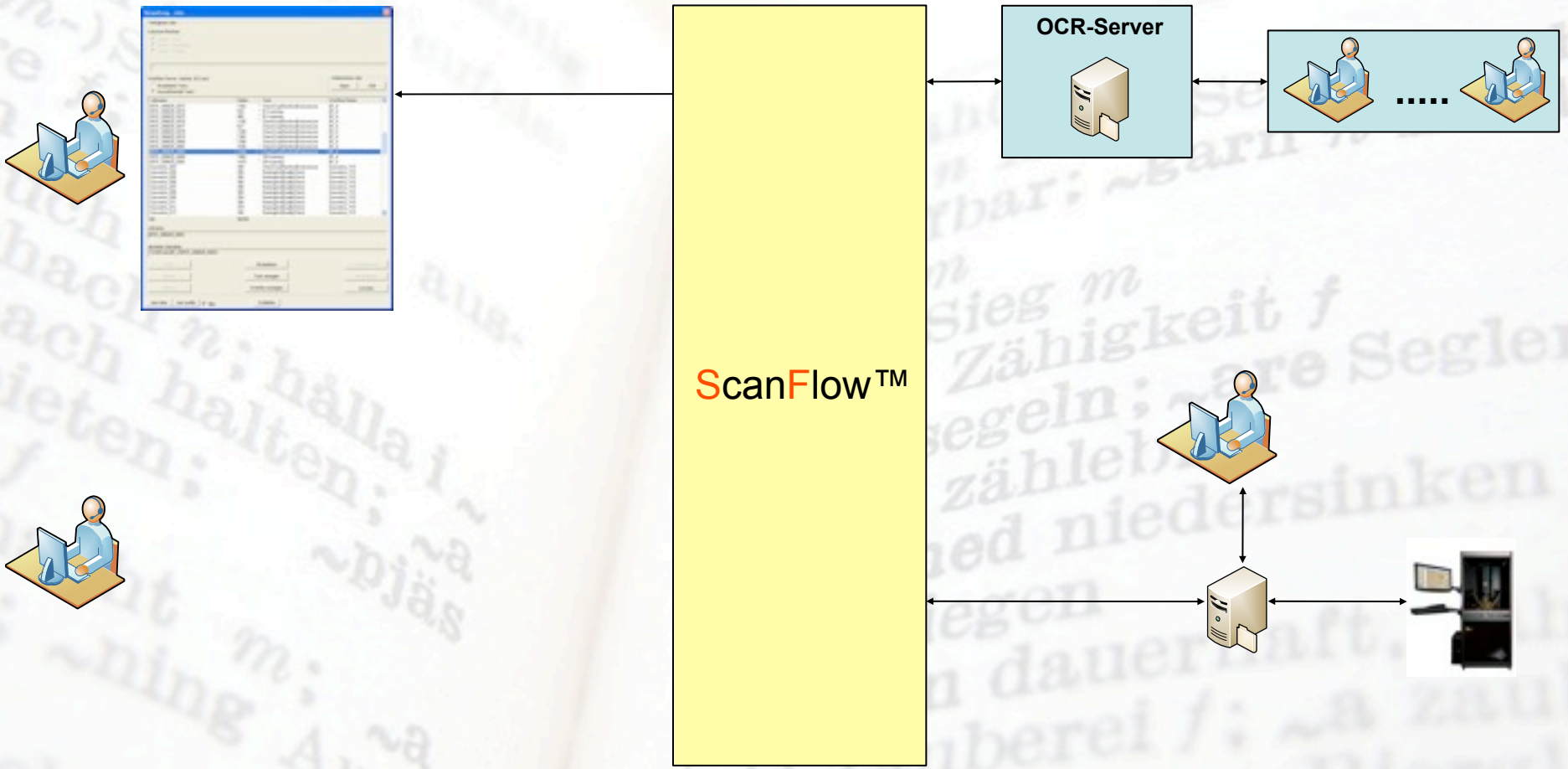
ScanFlow™



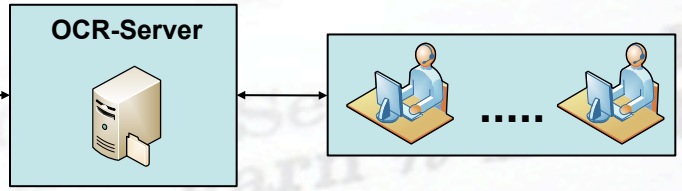


TREVENTUS

ScanFlow™: Example

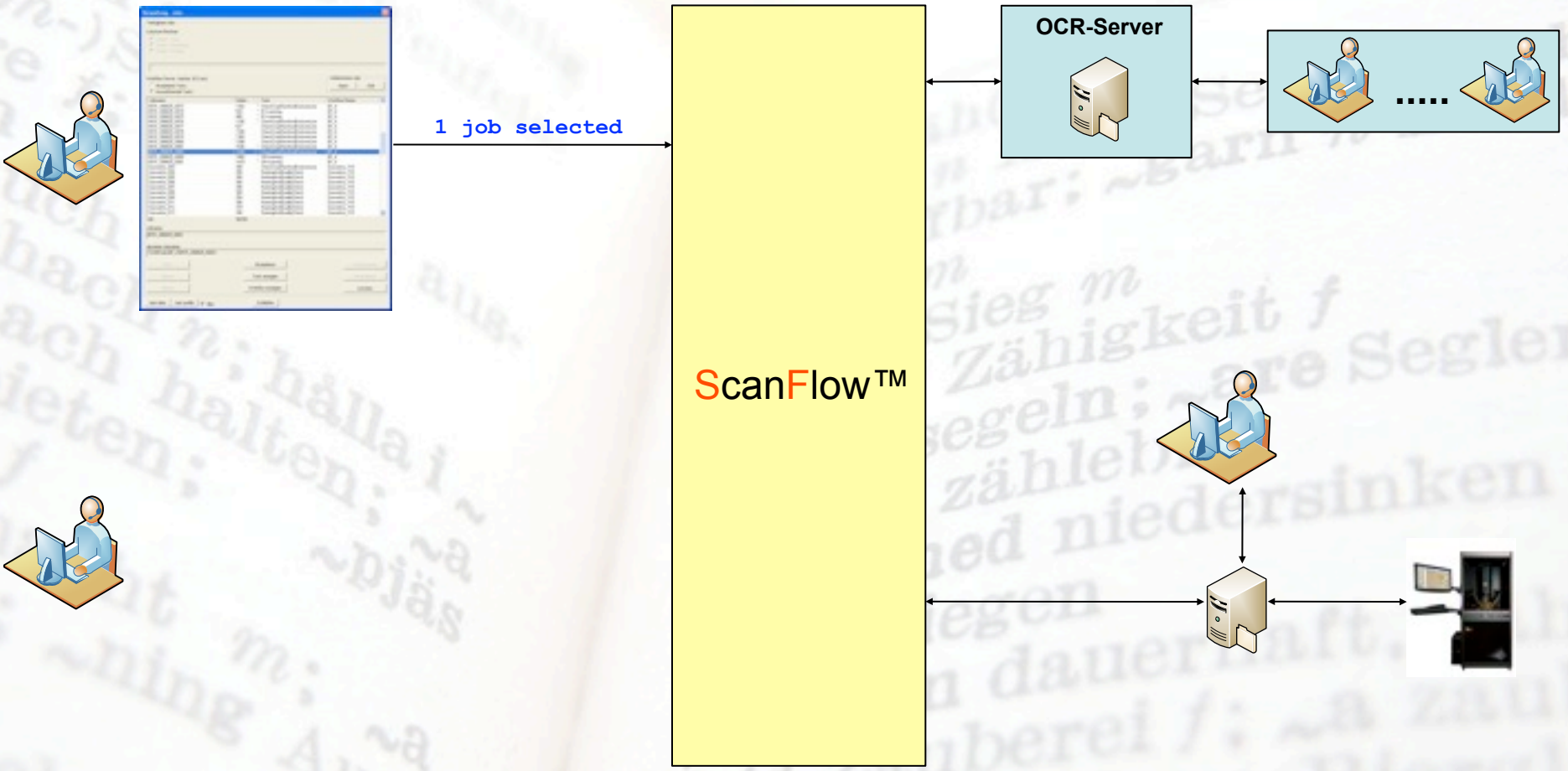


ScanFlow™: Example

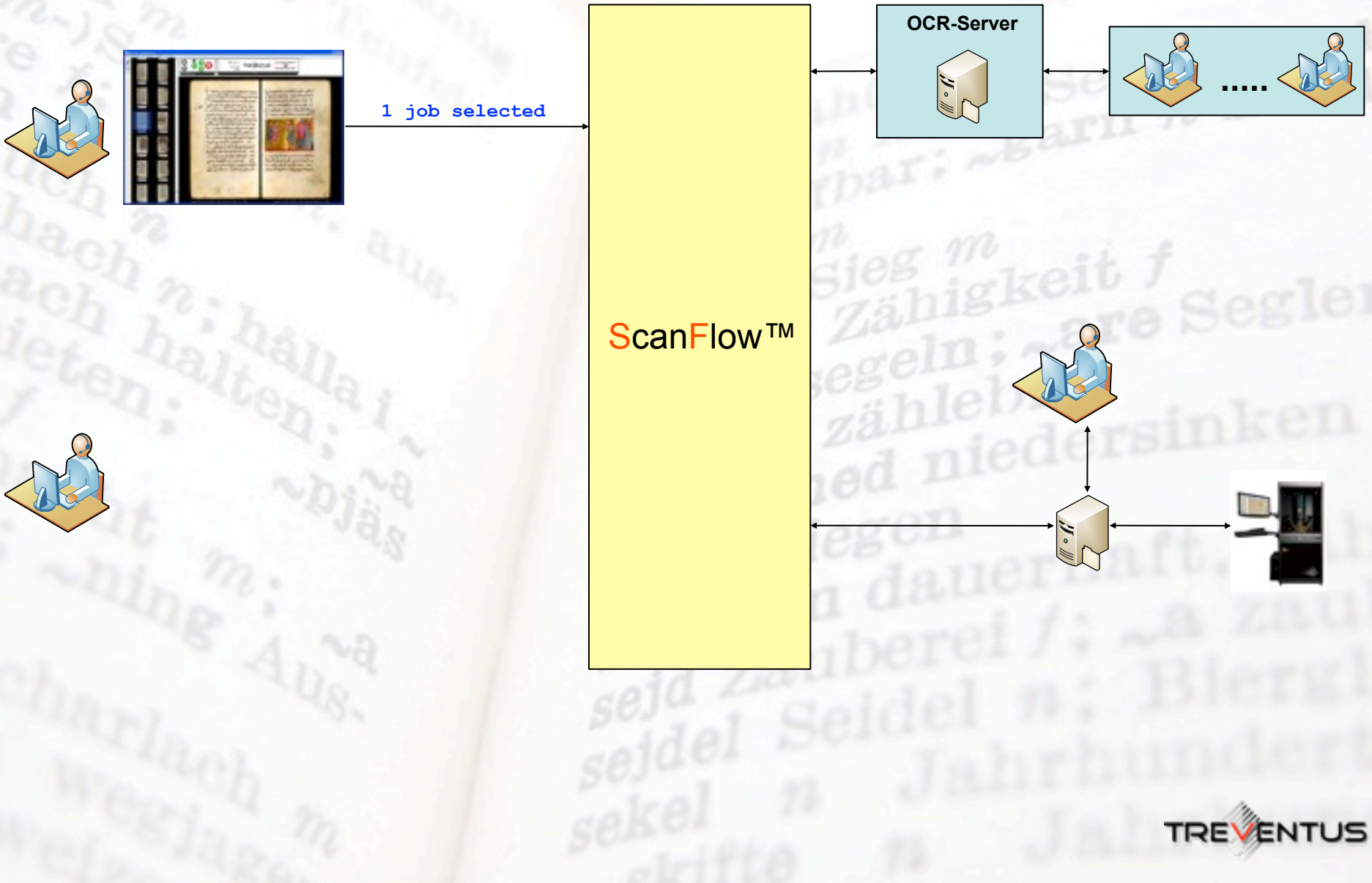


TREVENTUS

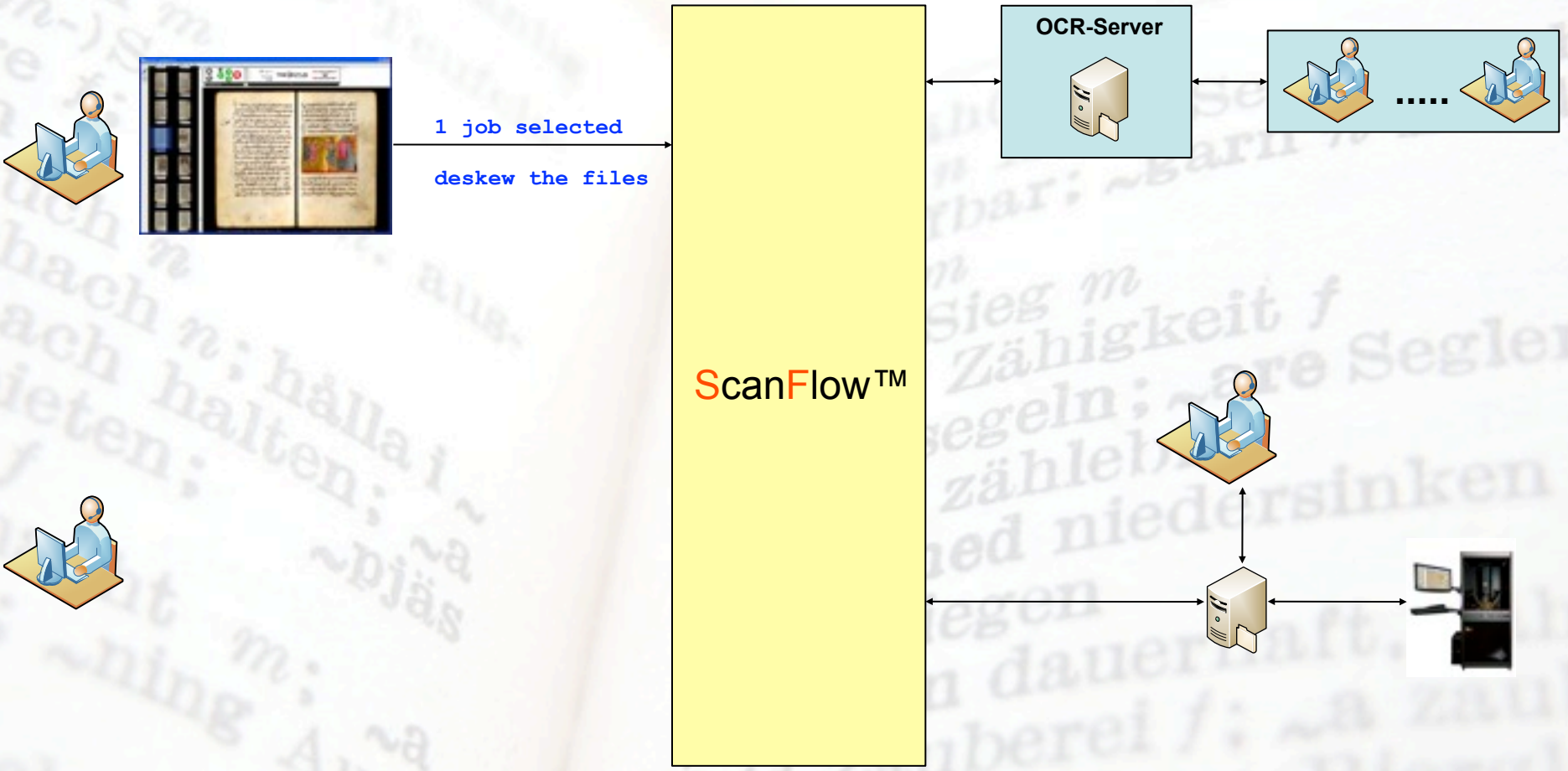
ScanFlow™: Example



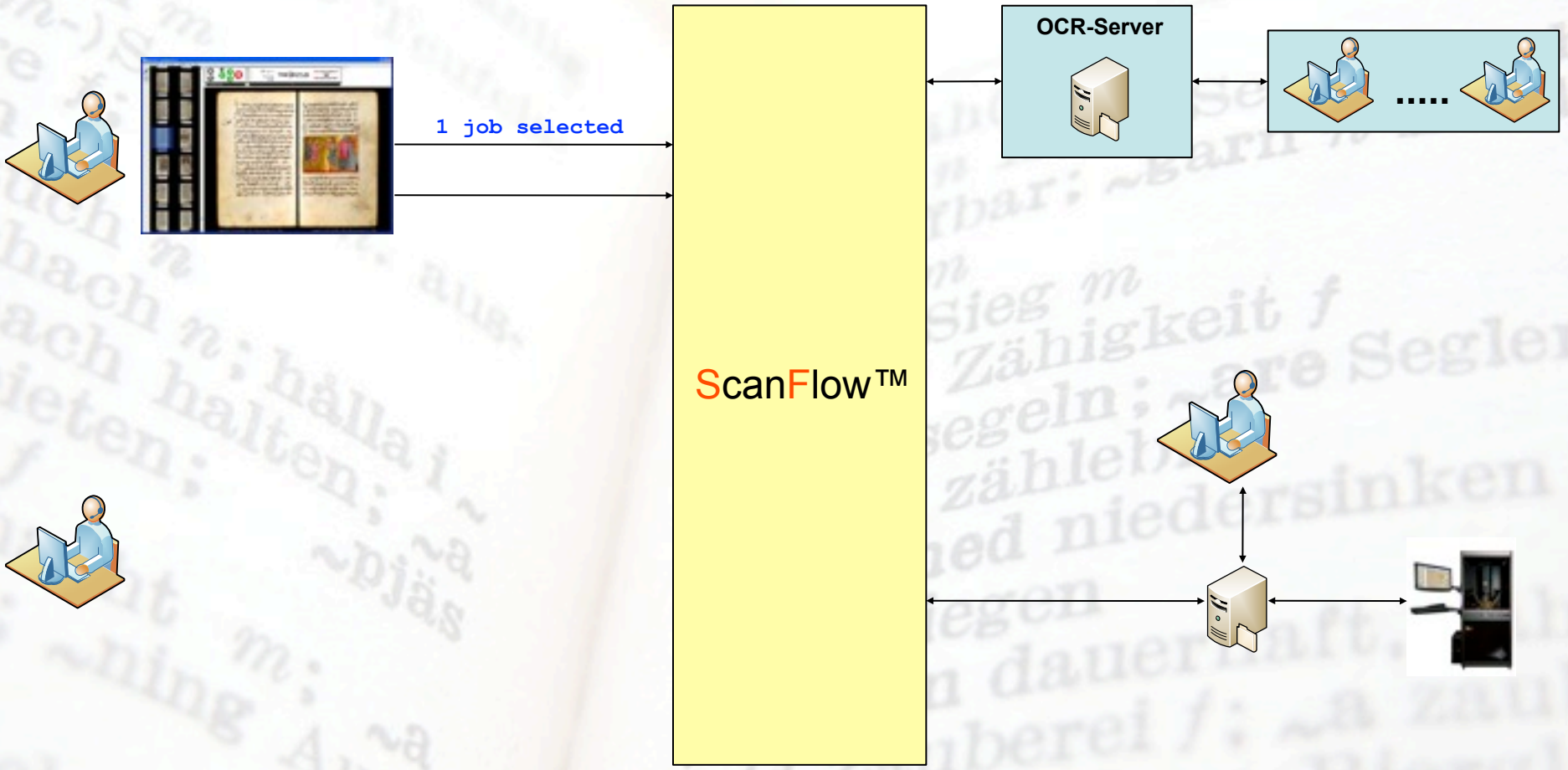
ScanFlow™: Example



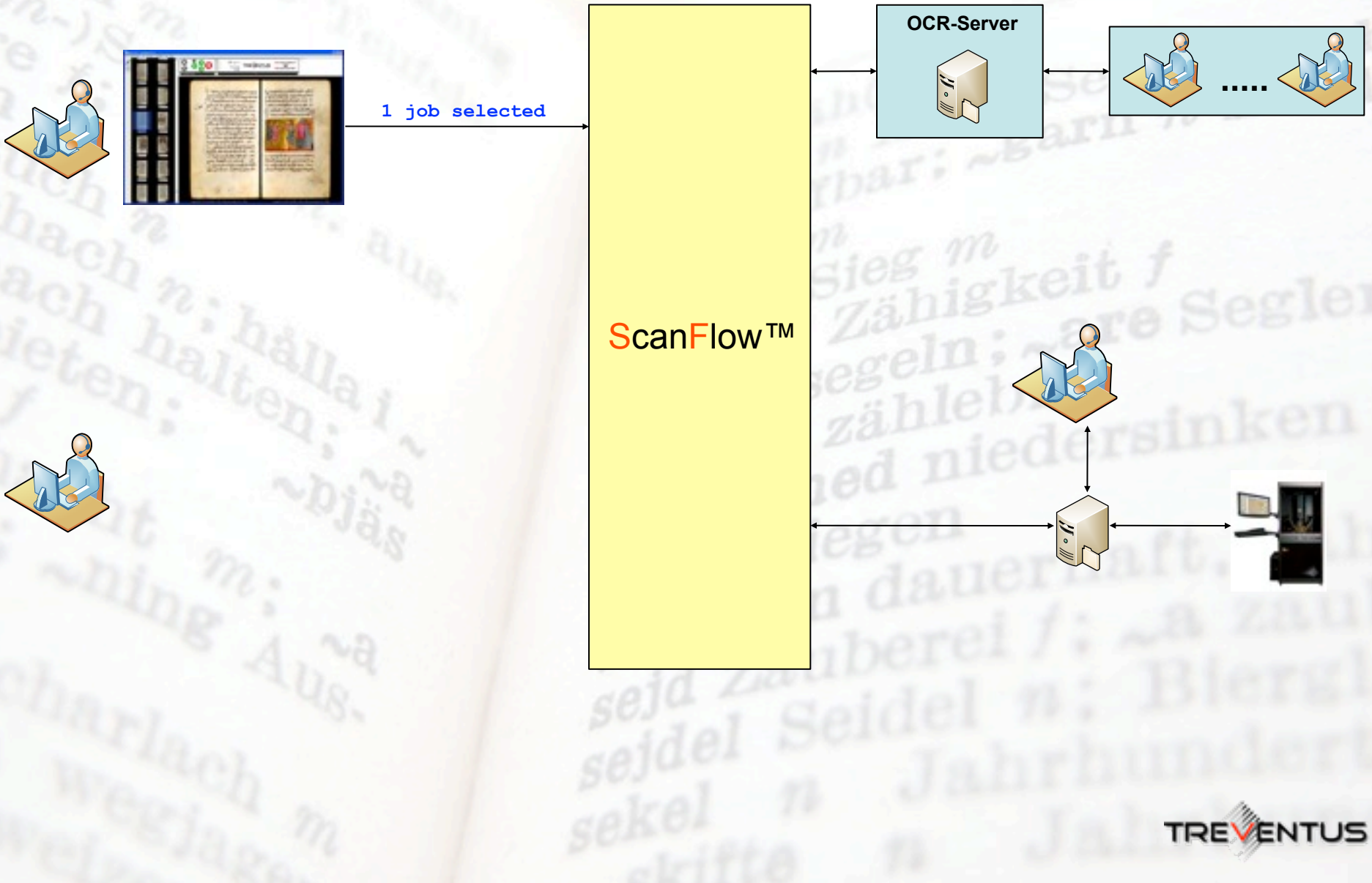
ScanFlow™: Example



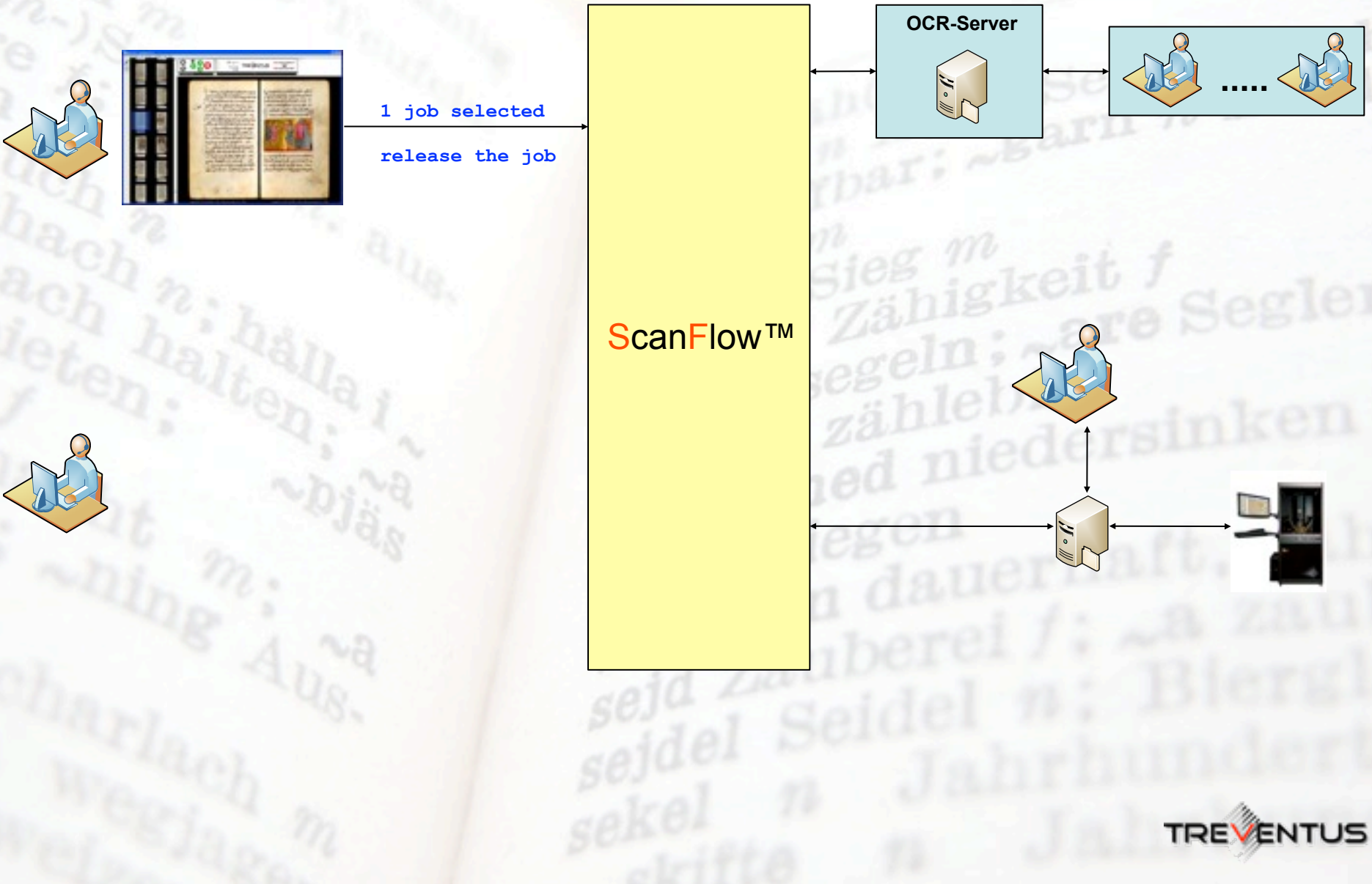
ScanFlow™: Example



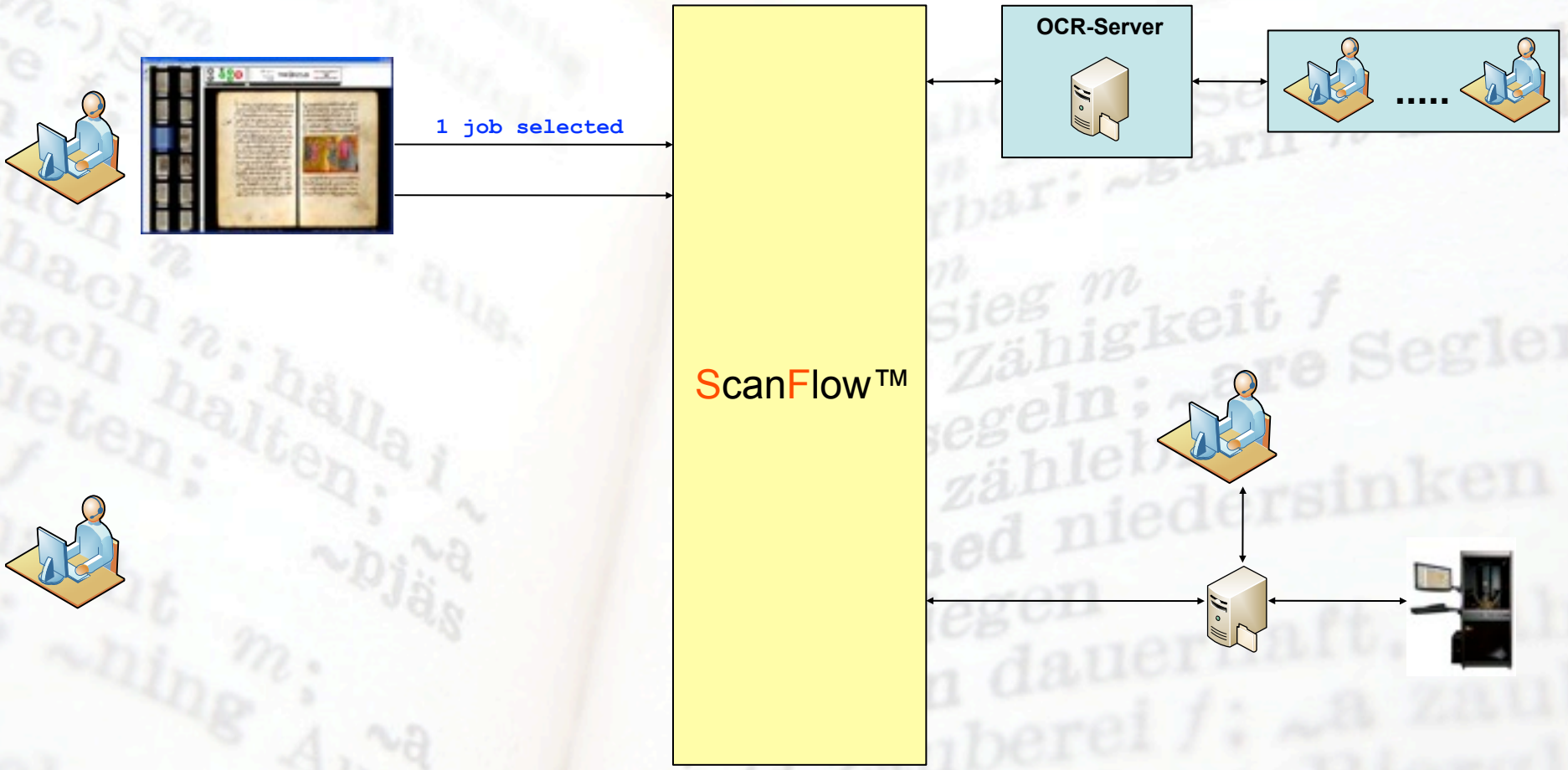
ScanFlow™: Example



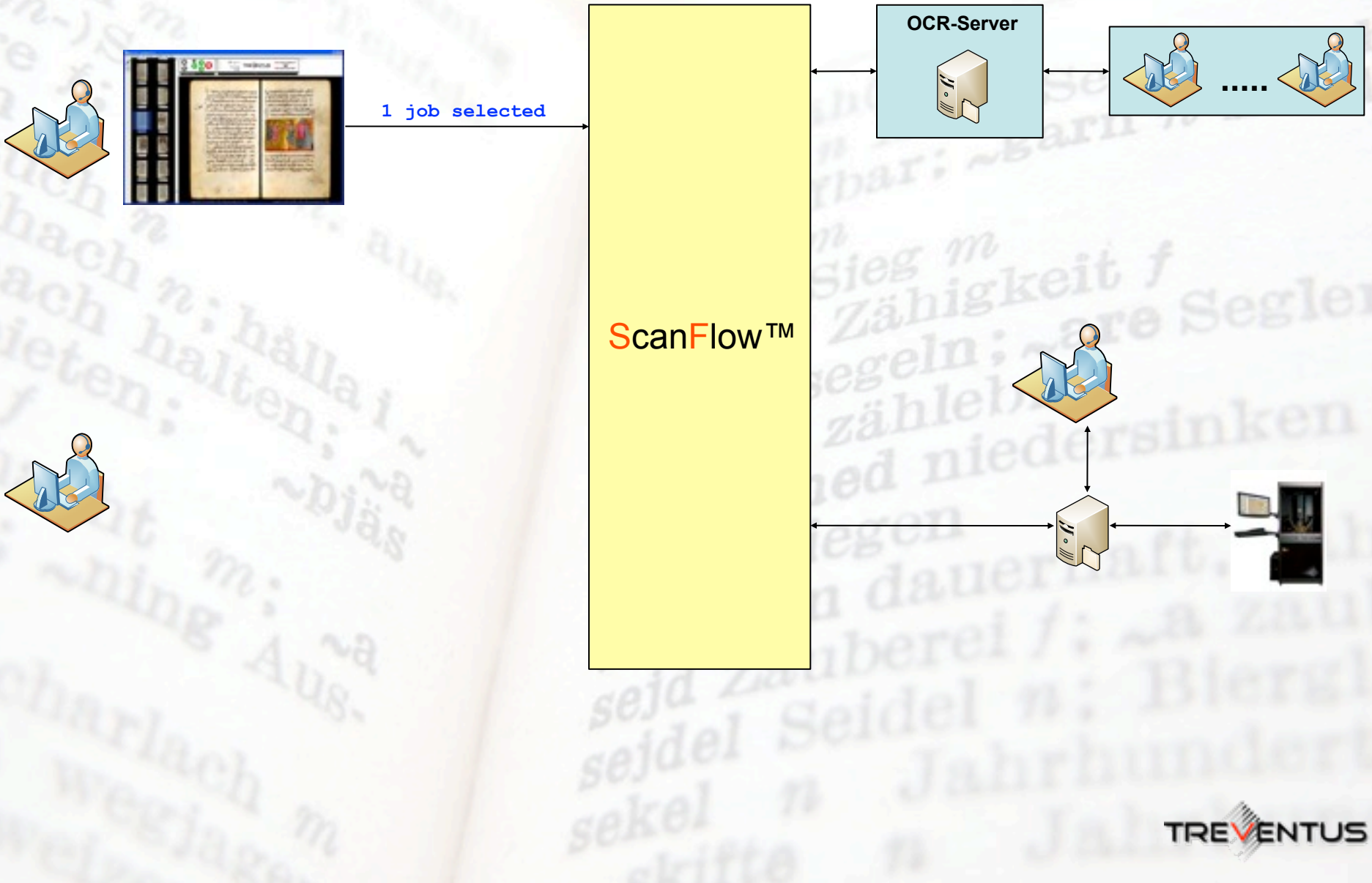
ScanFlow™: Example



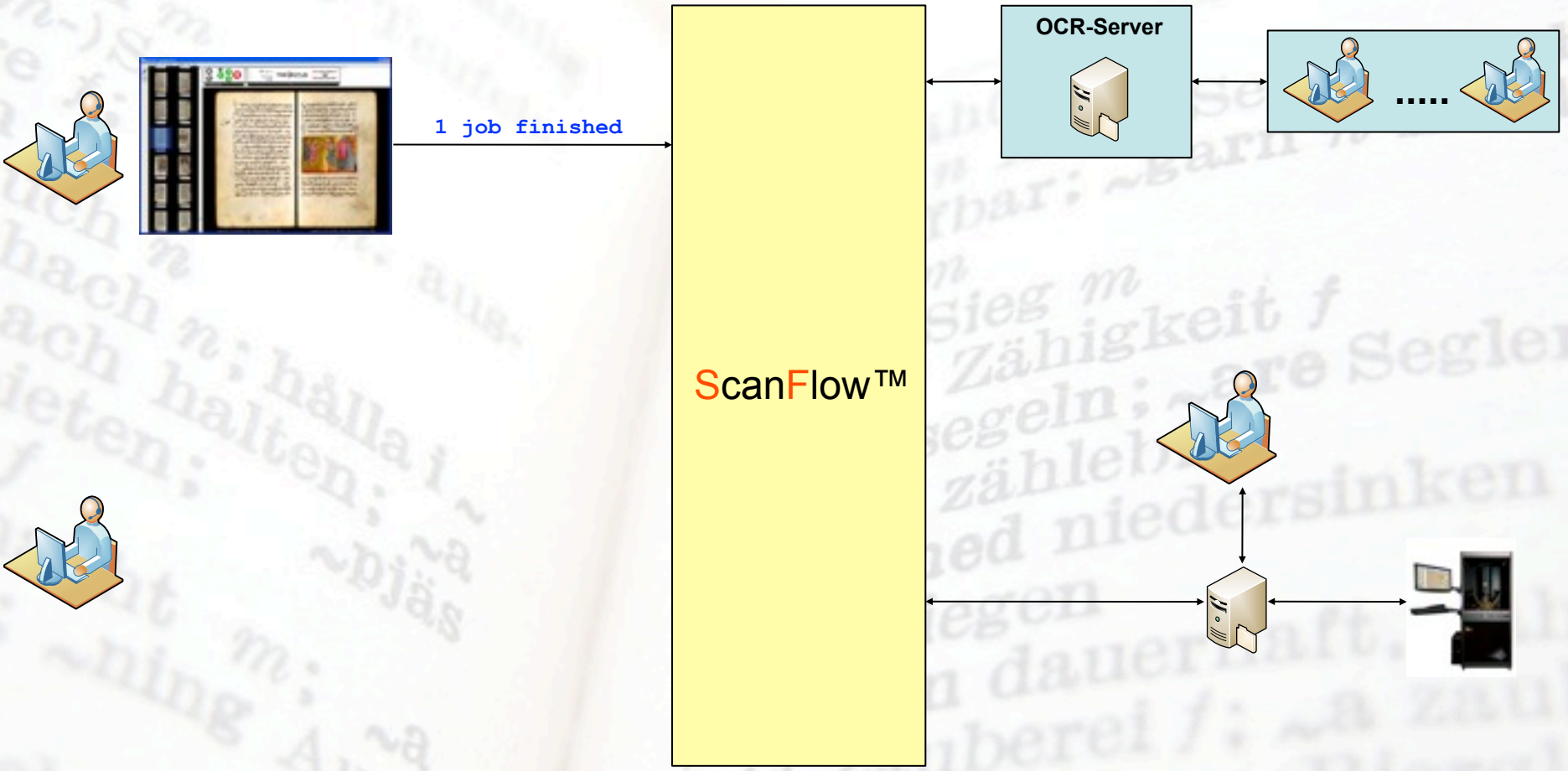
ScanFlow™: Example



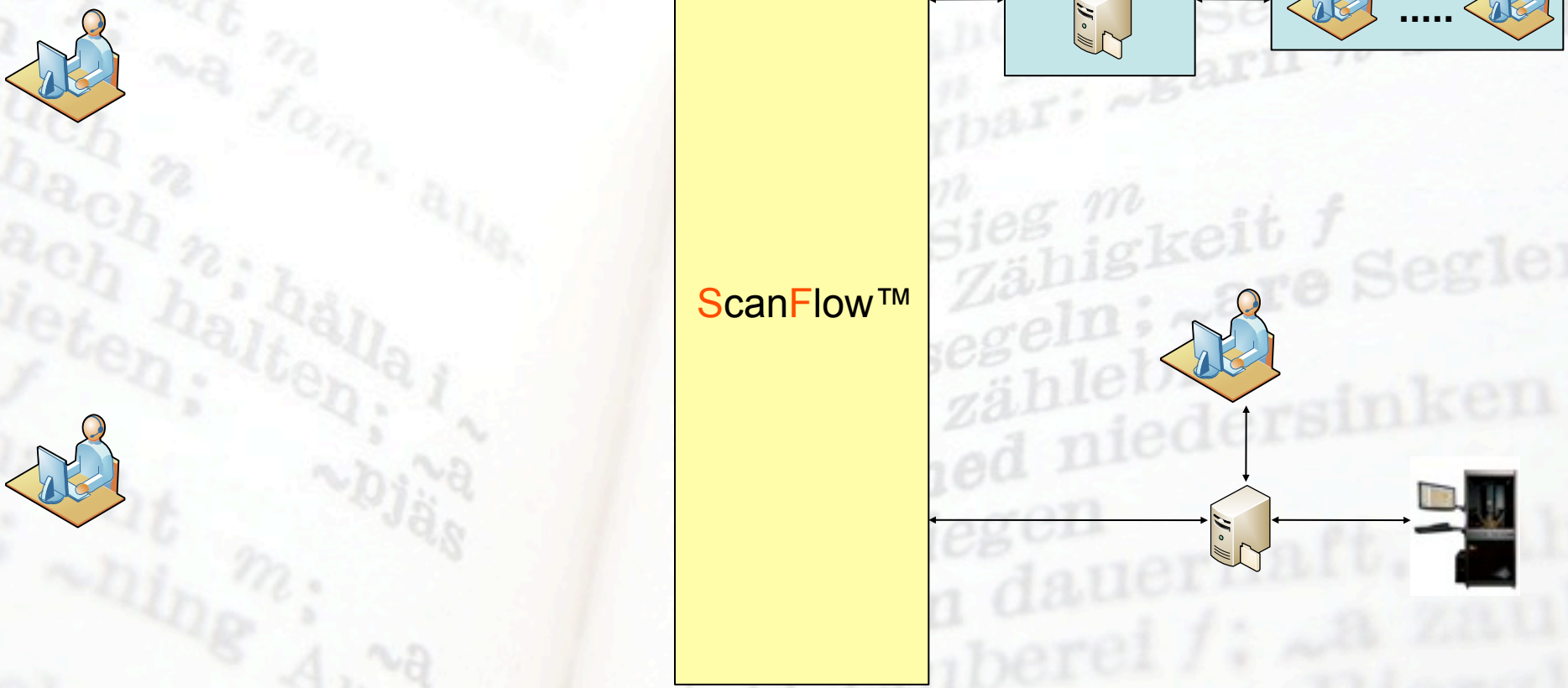
ScanFlow™: Example



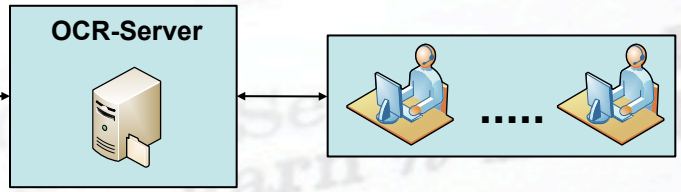
ScanFlow™: Example



ScanFlow™: Example

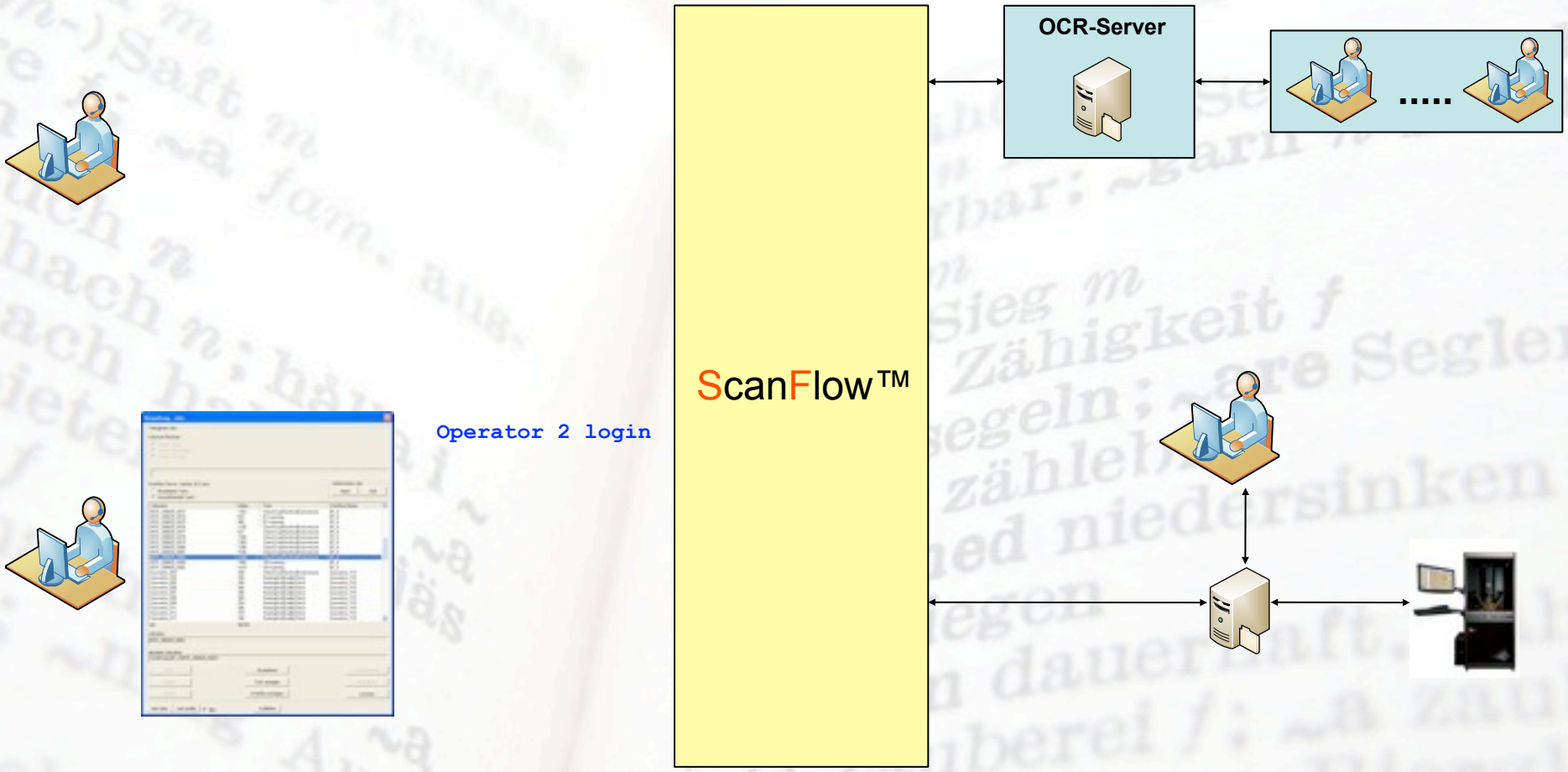


ScanFlow™: Example

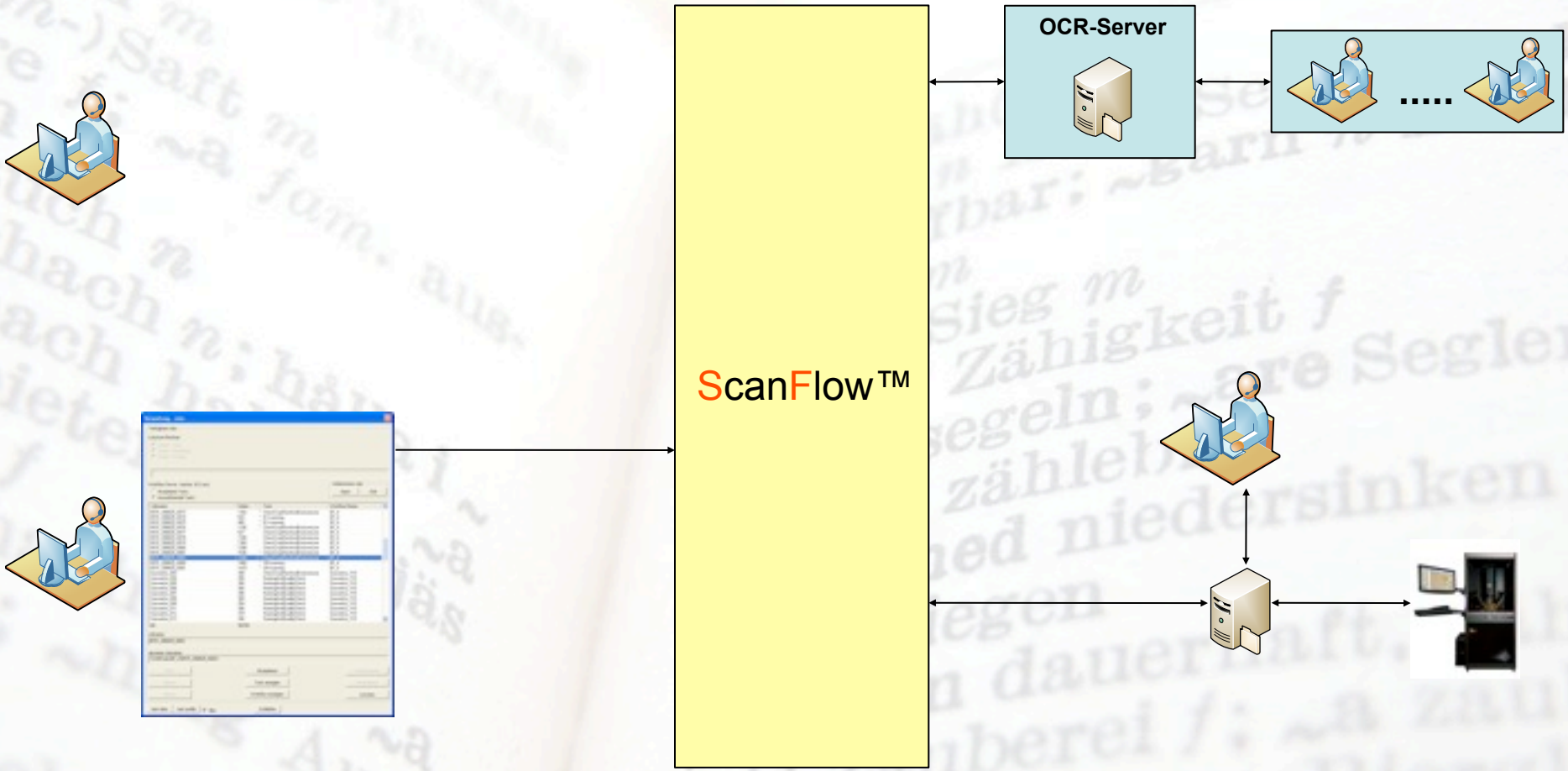


TREVENTUS

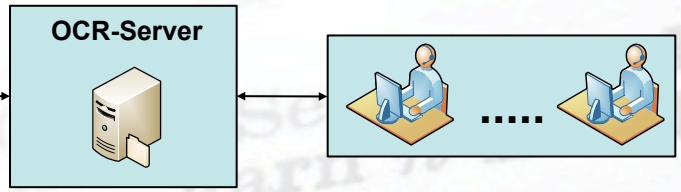
ScanFlow™: Example



ScanFlow™: Example



ScanFlow™: Example



TREVENTUS

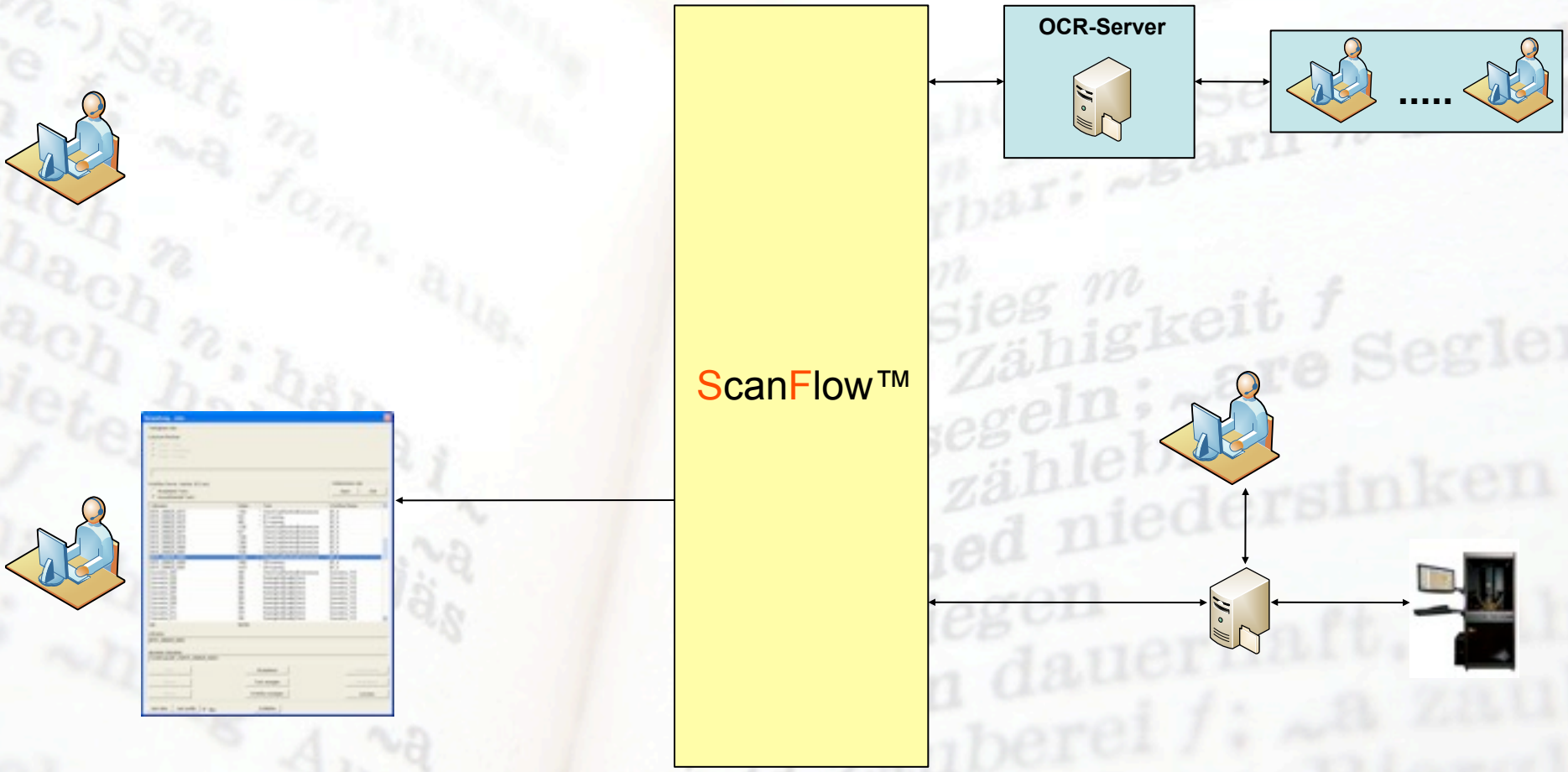
ScanFlow™: Example

List of all jobs

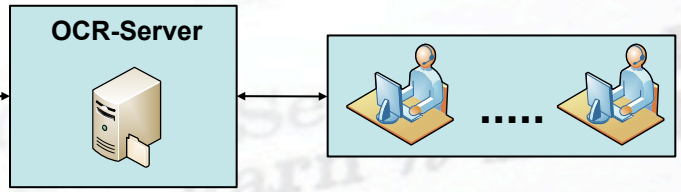
ScanFlow™

OCR-Server

ScanFlow™: Example

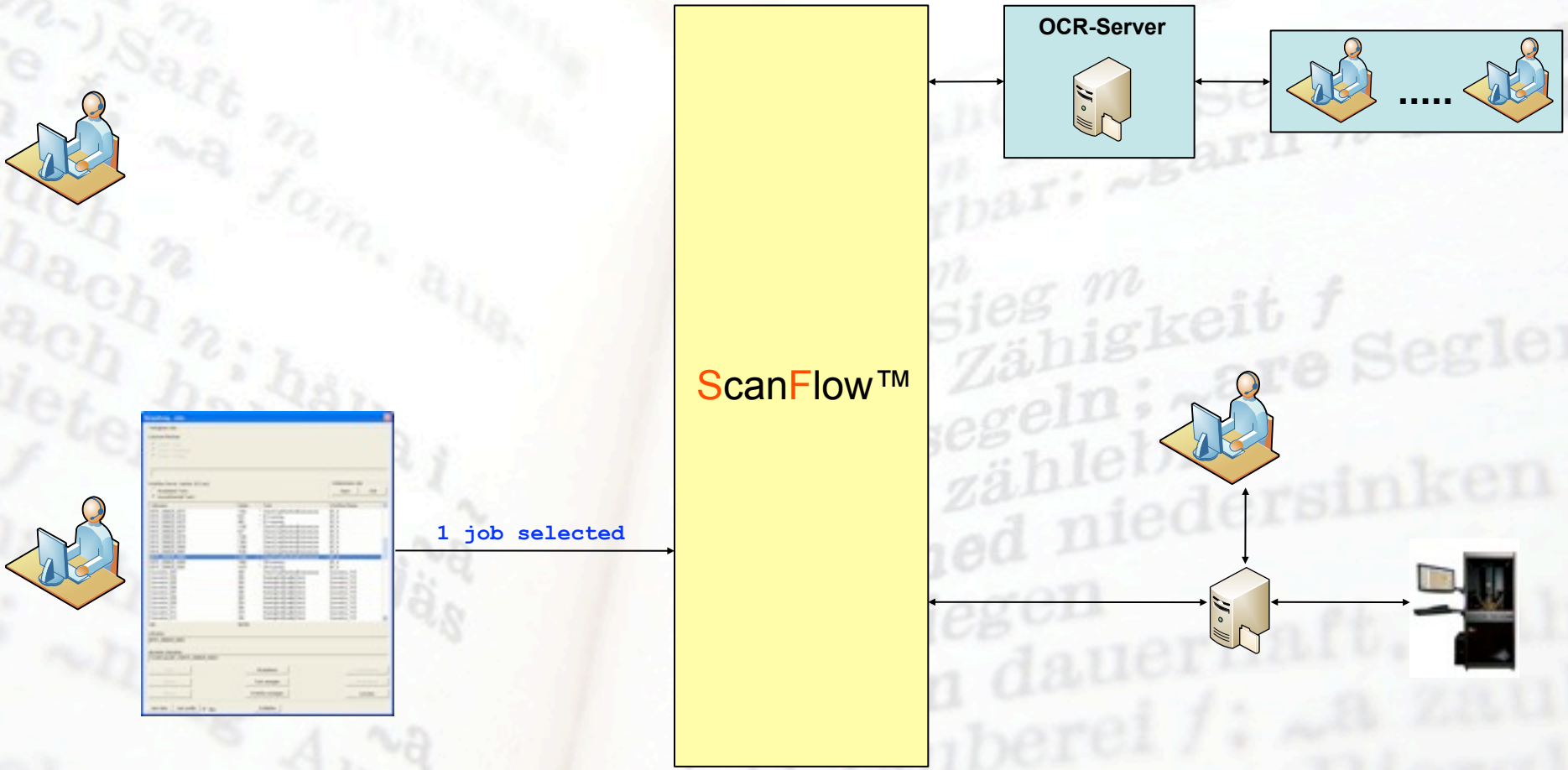


ScanFlow™: Example

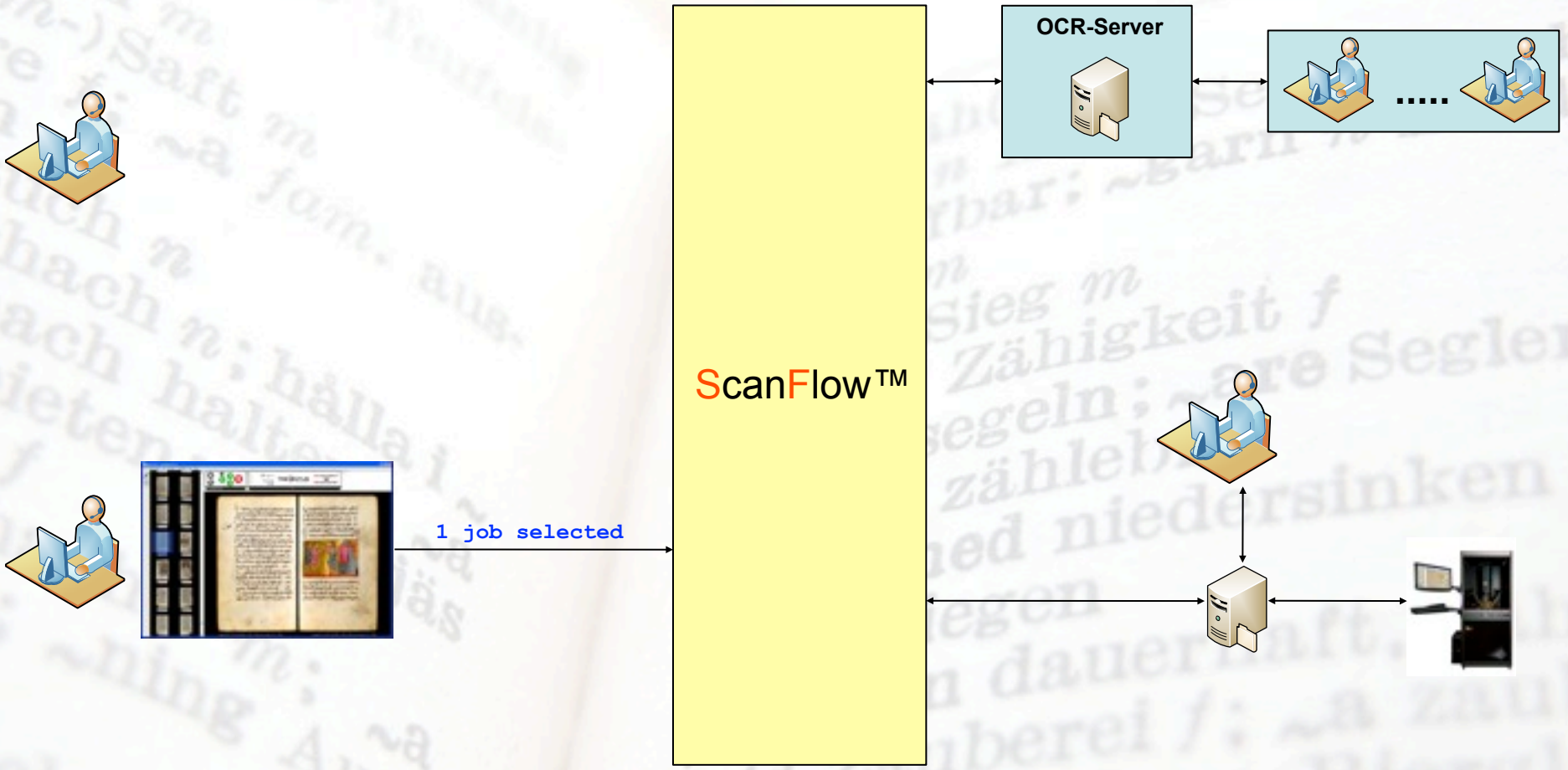


TREVENTUS

ScanFlow™: Example

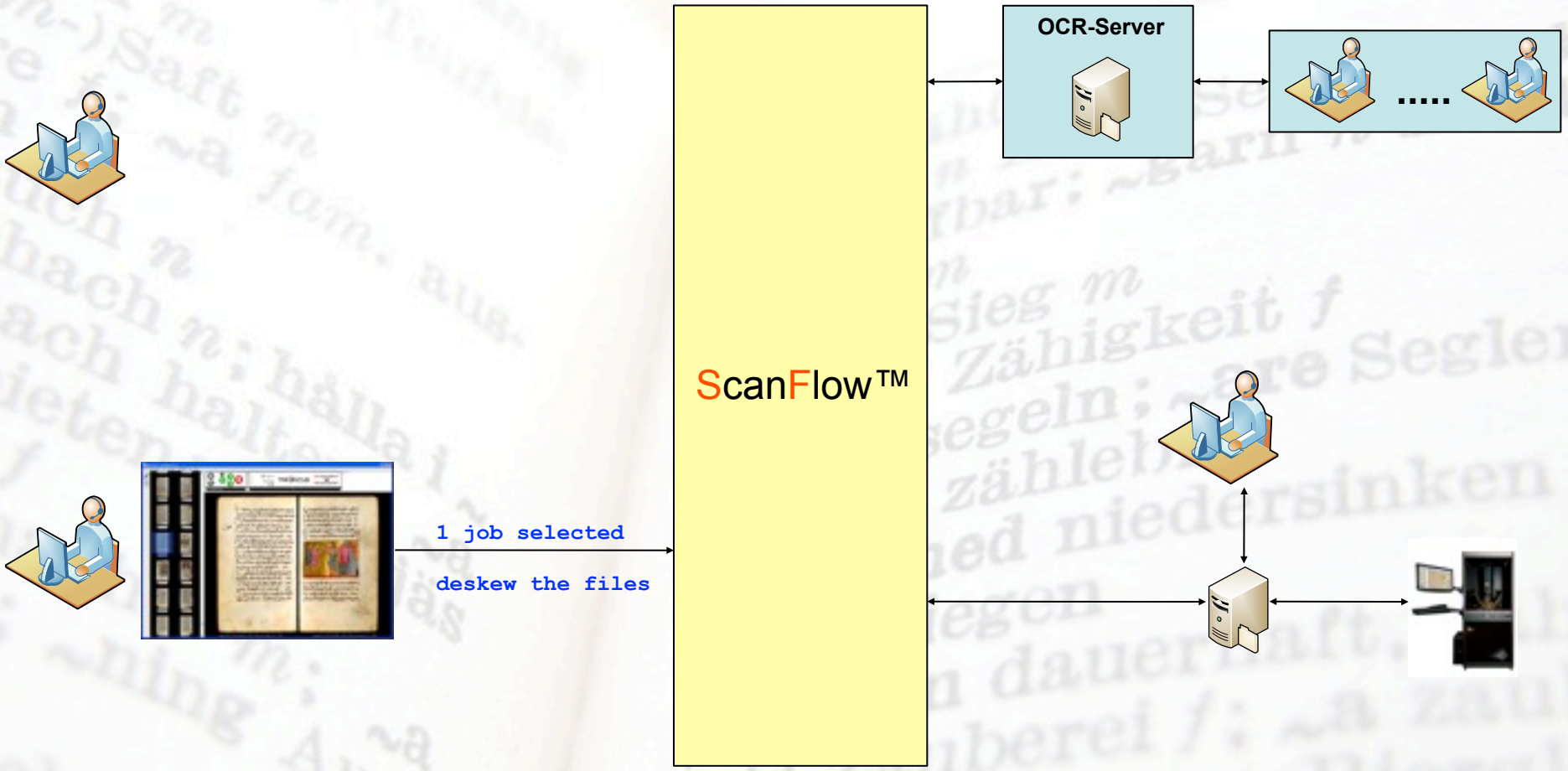


ScanFlow™: Example

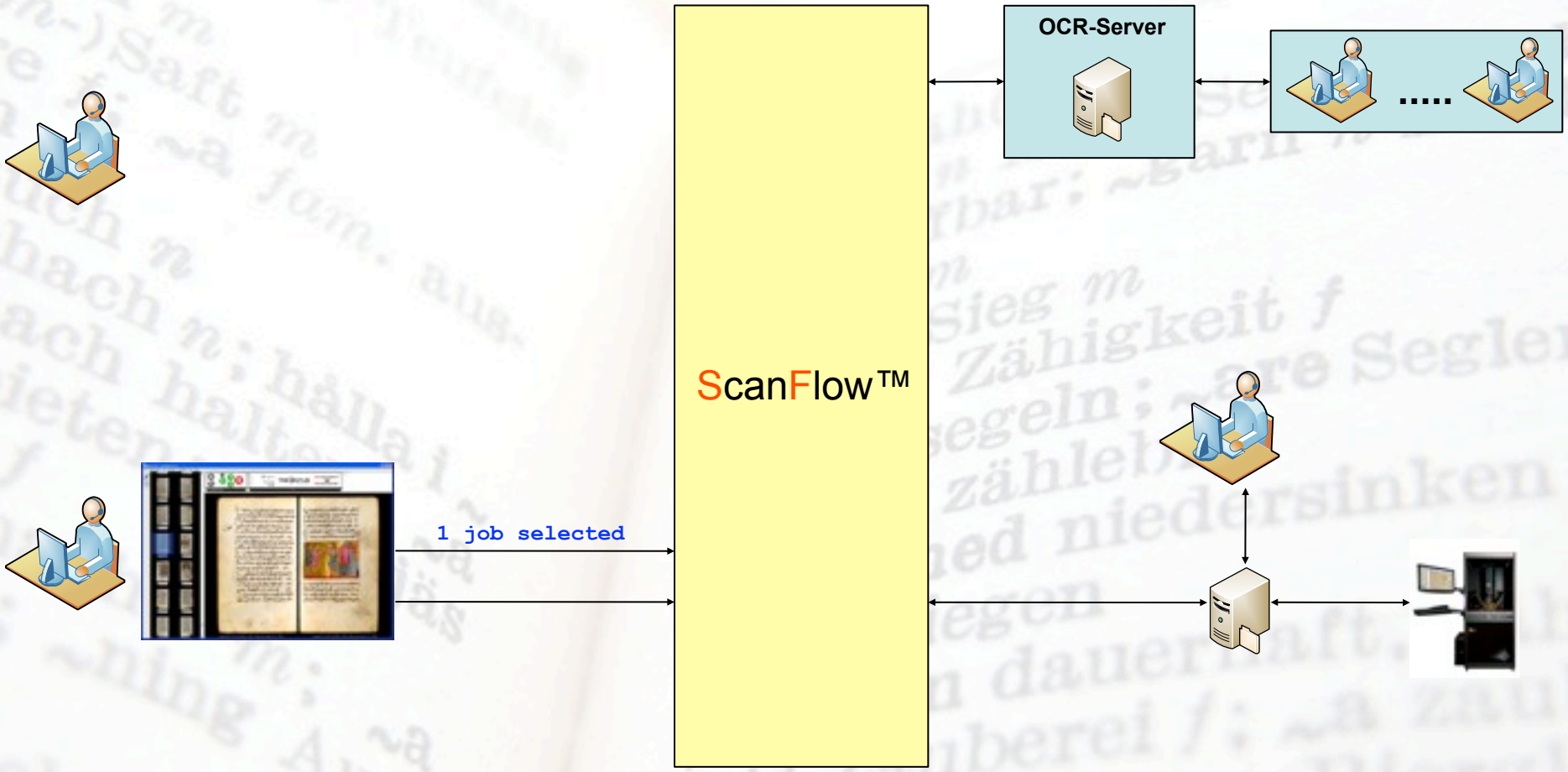


TREVENTUS

ScanFlow™: Example

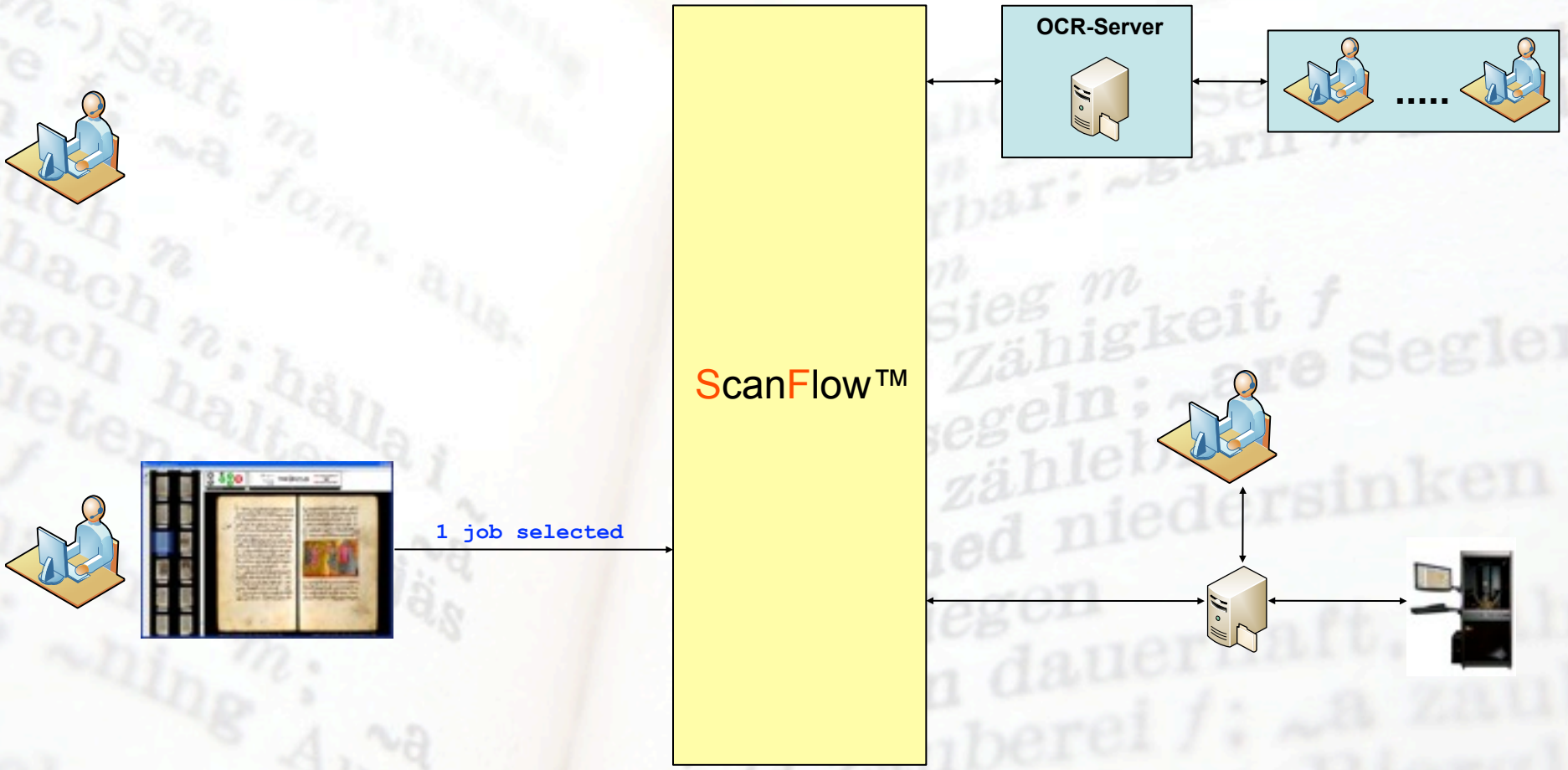


ScanFlow™: Example



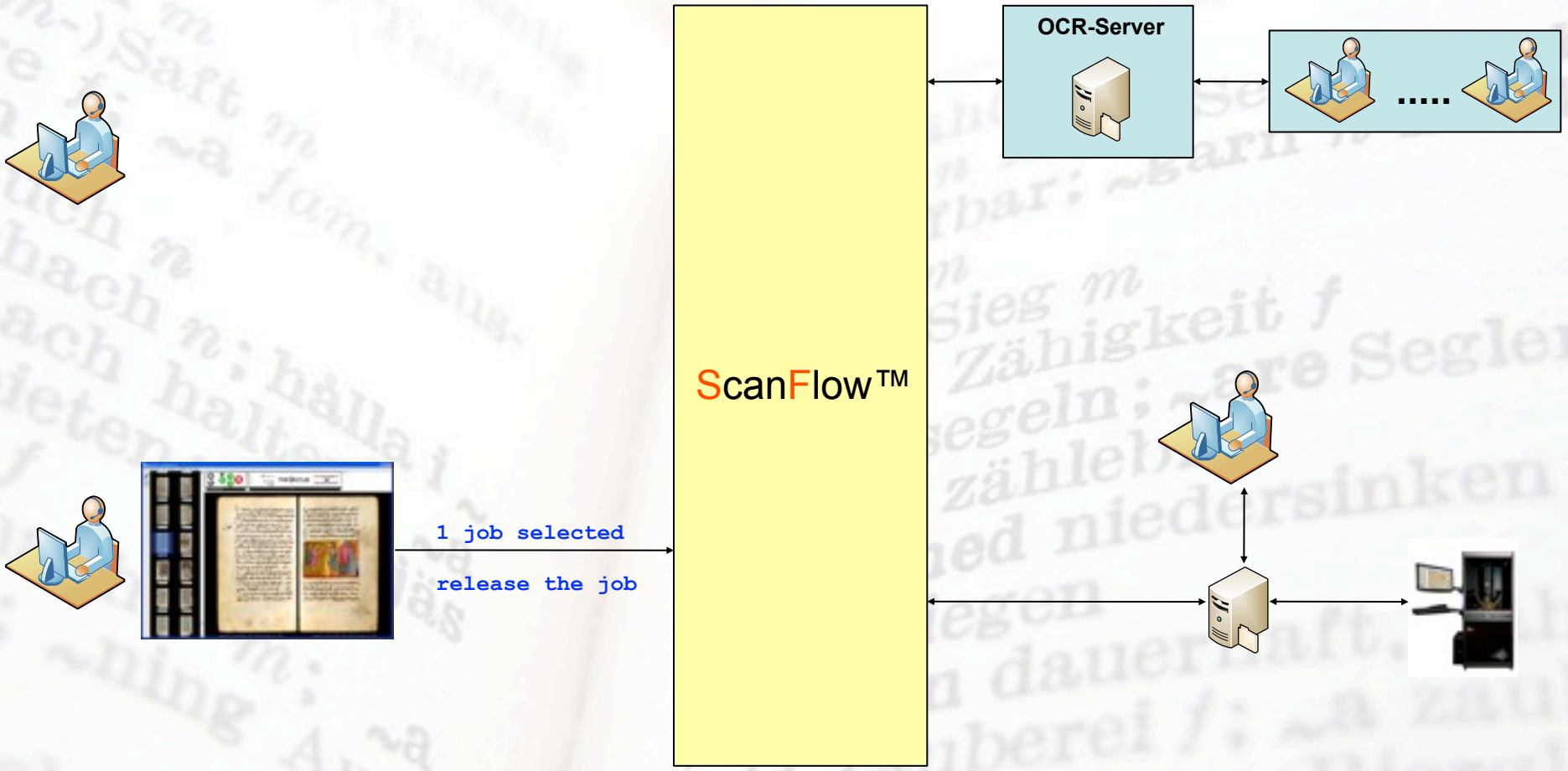
TREVENTUS

ScanFlow™: Example



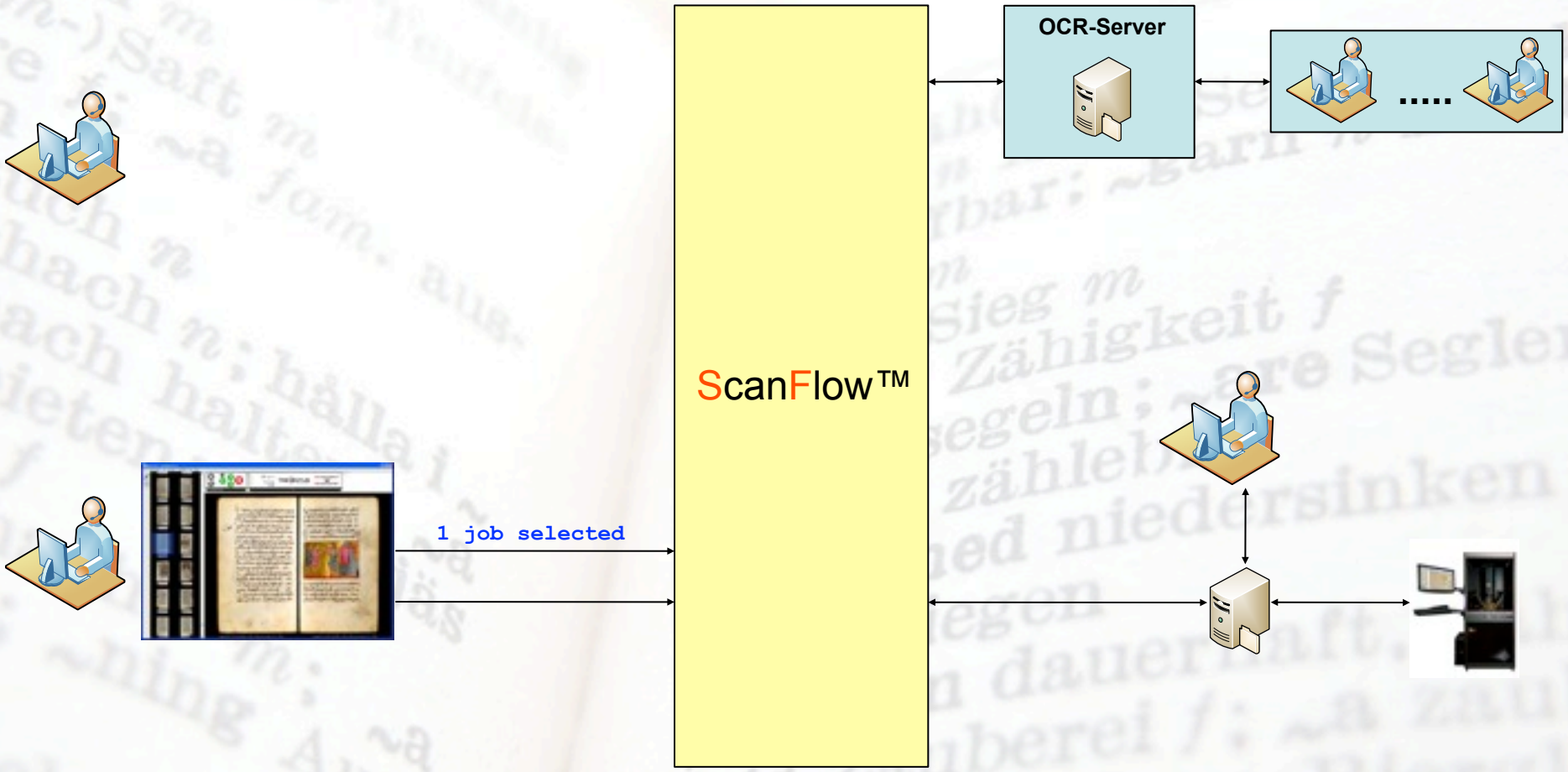
TREVENTUS

ScanFlow™: Example



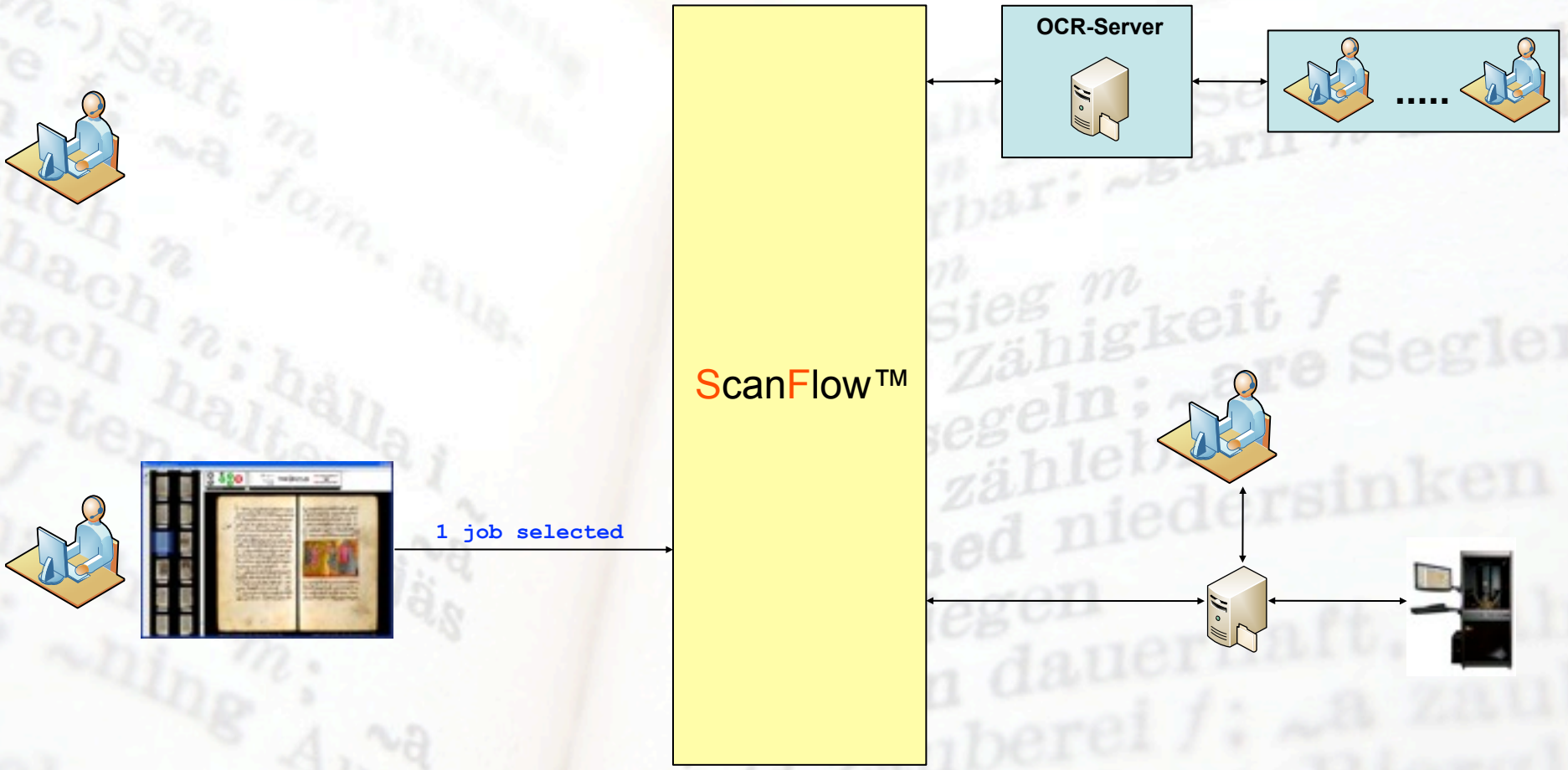
TREVENTUS

ScanFlow™: Example



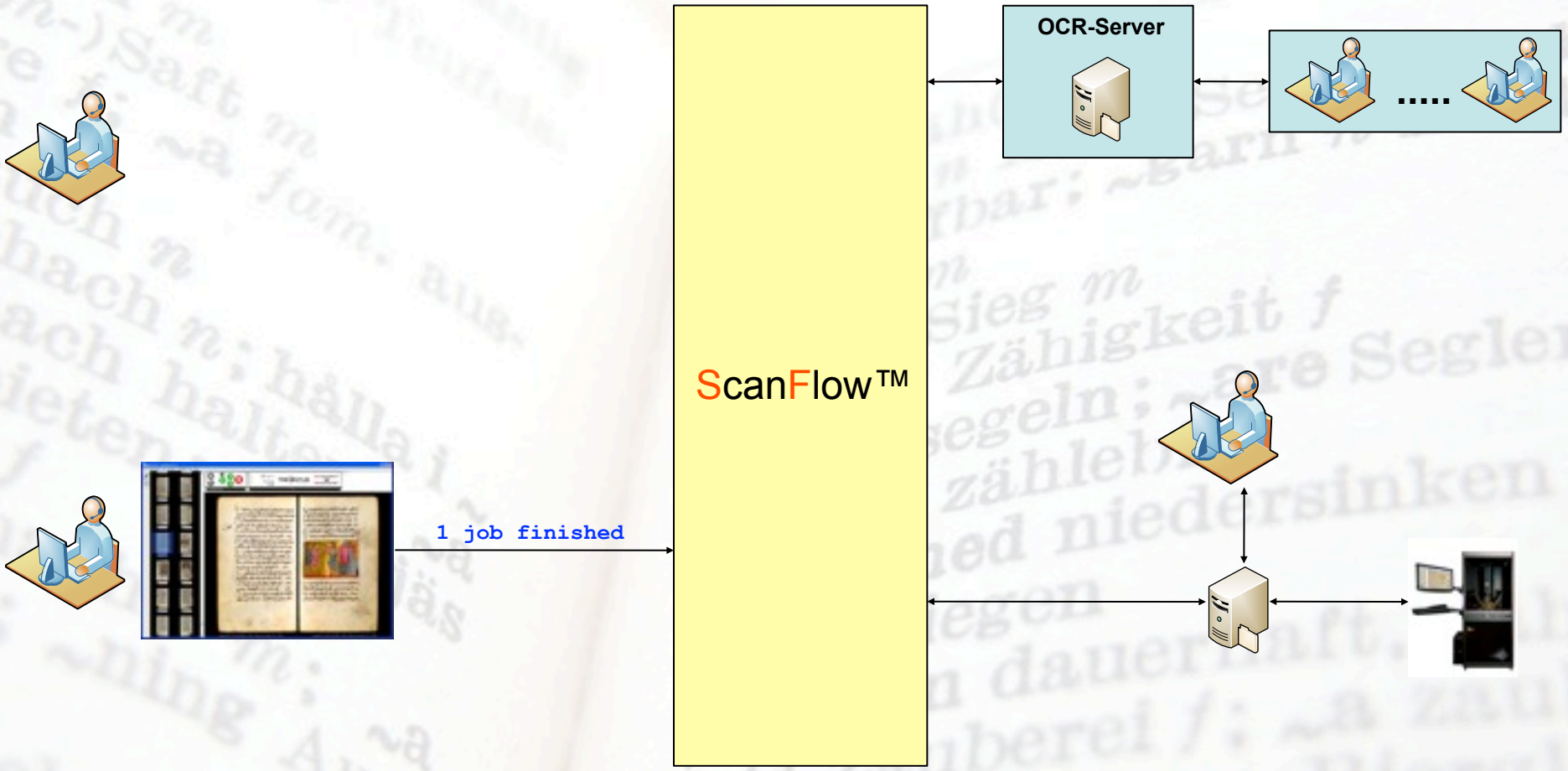
TREVENTUS

ScanFlow™: Example



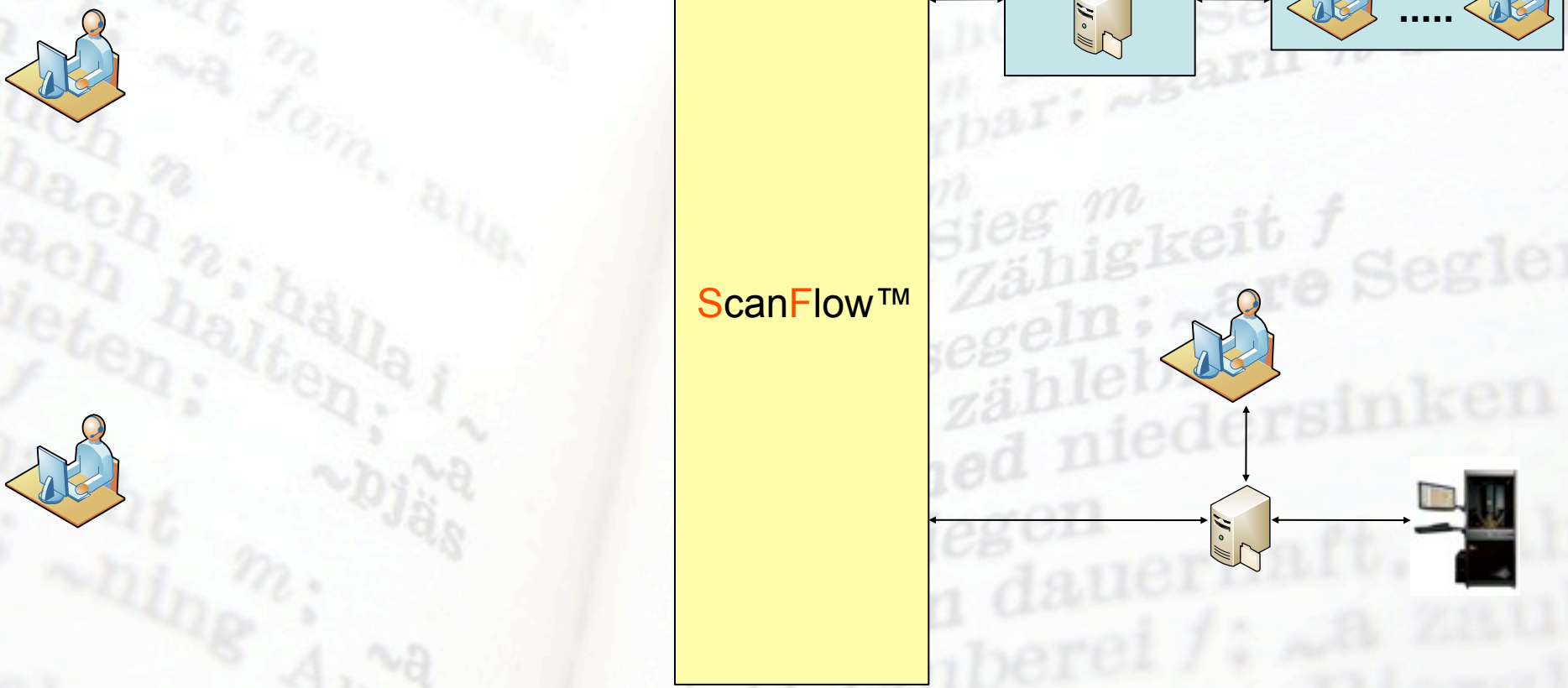
TREVENTUS

ScanFlow™: Example

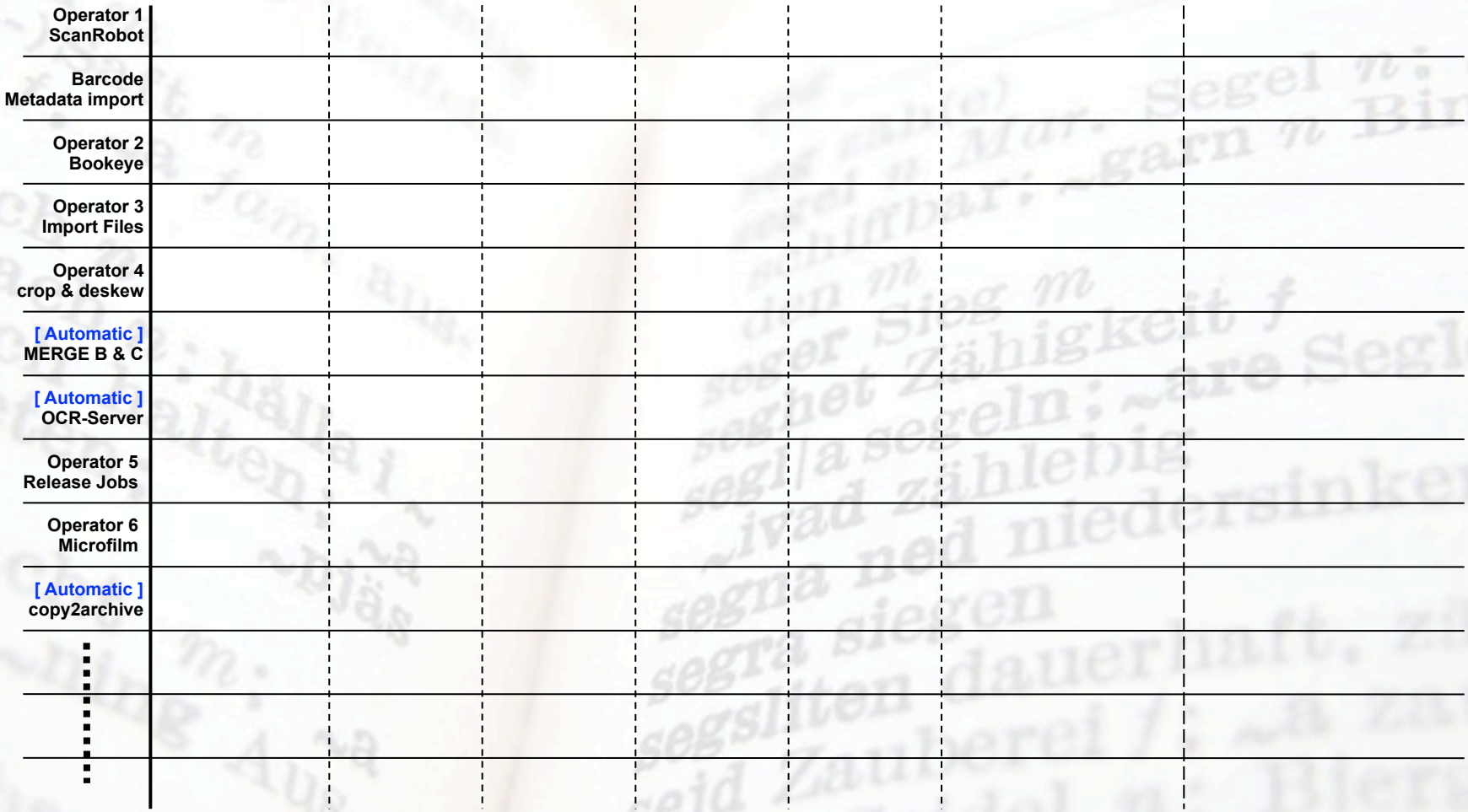


TREVENTUS

ScanFlow™: Example



ScanFlow™: Jobs in the timetable

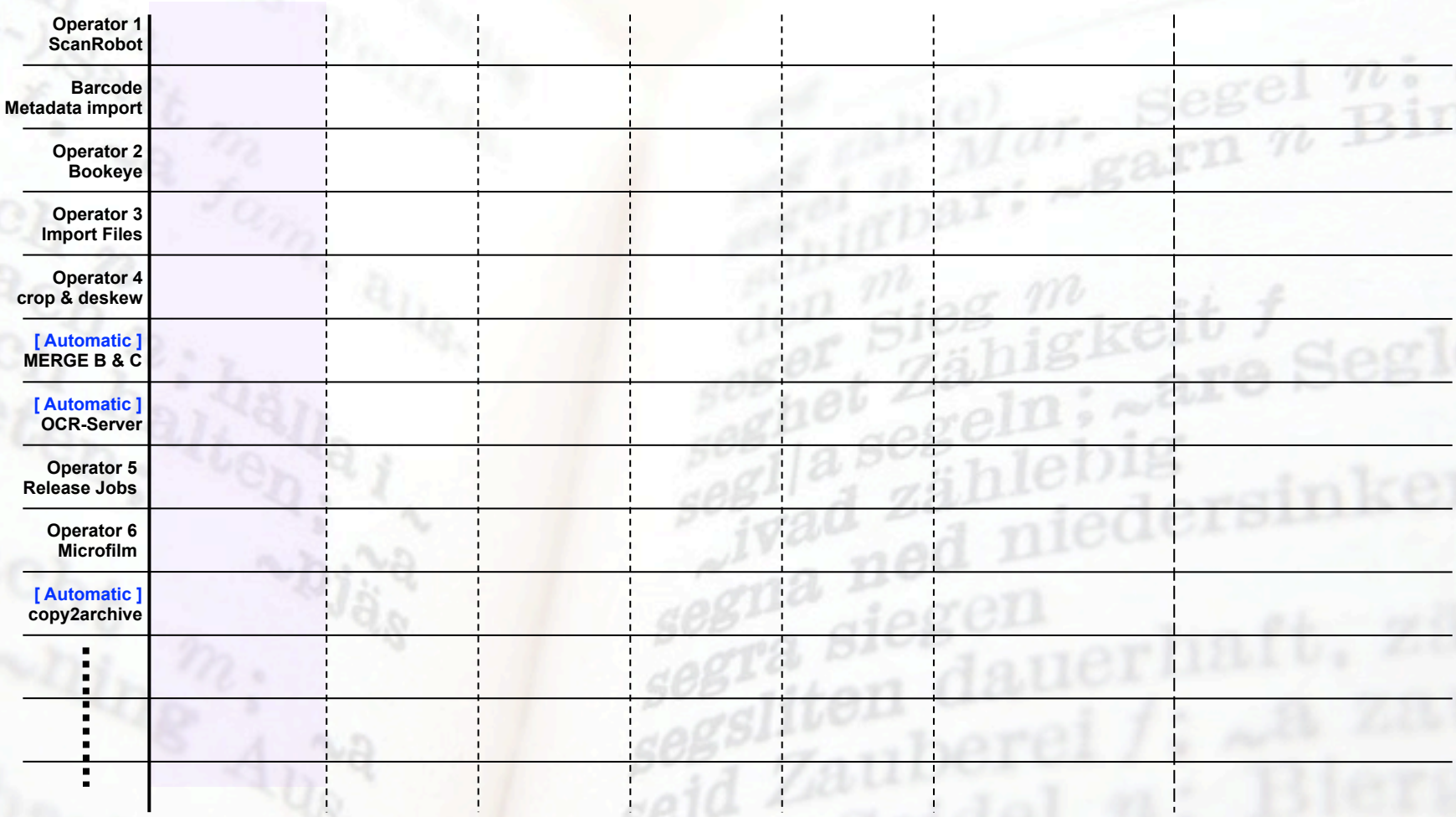


ScanFlow™

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

TREVENTUS

ScanFlow™: Jobs in the timetable

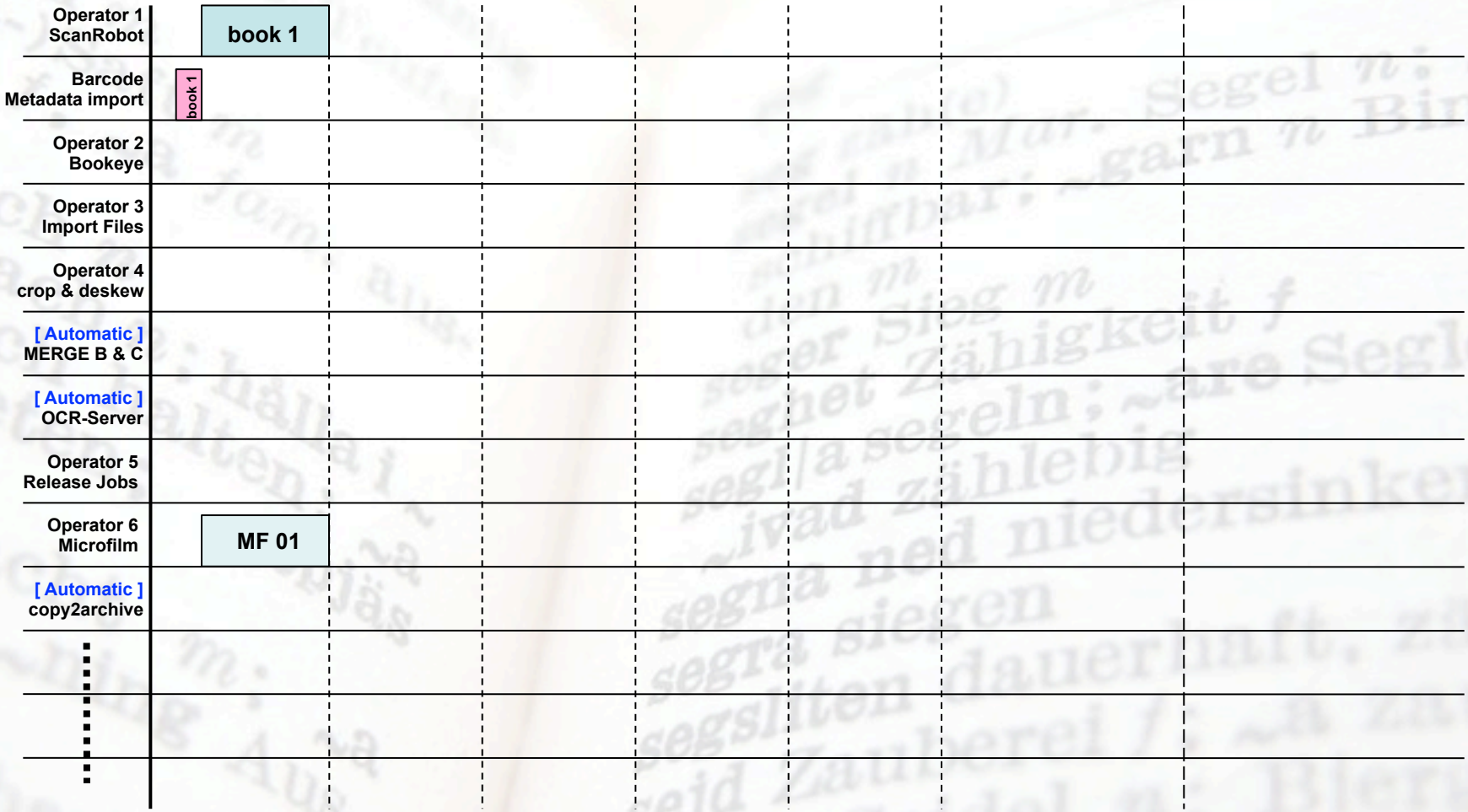


ScanFlow™

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

TREVENTUS

ScanFlow™: Jobs in the timetable

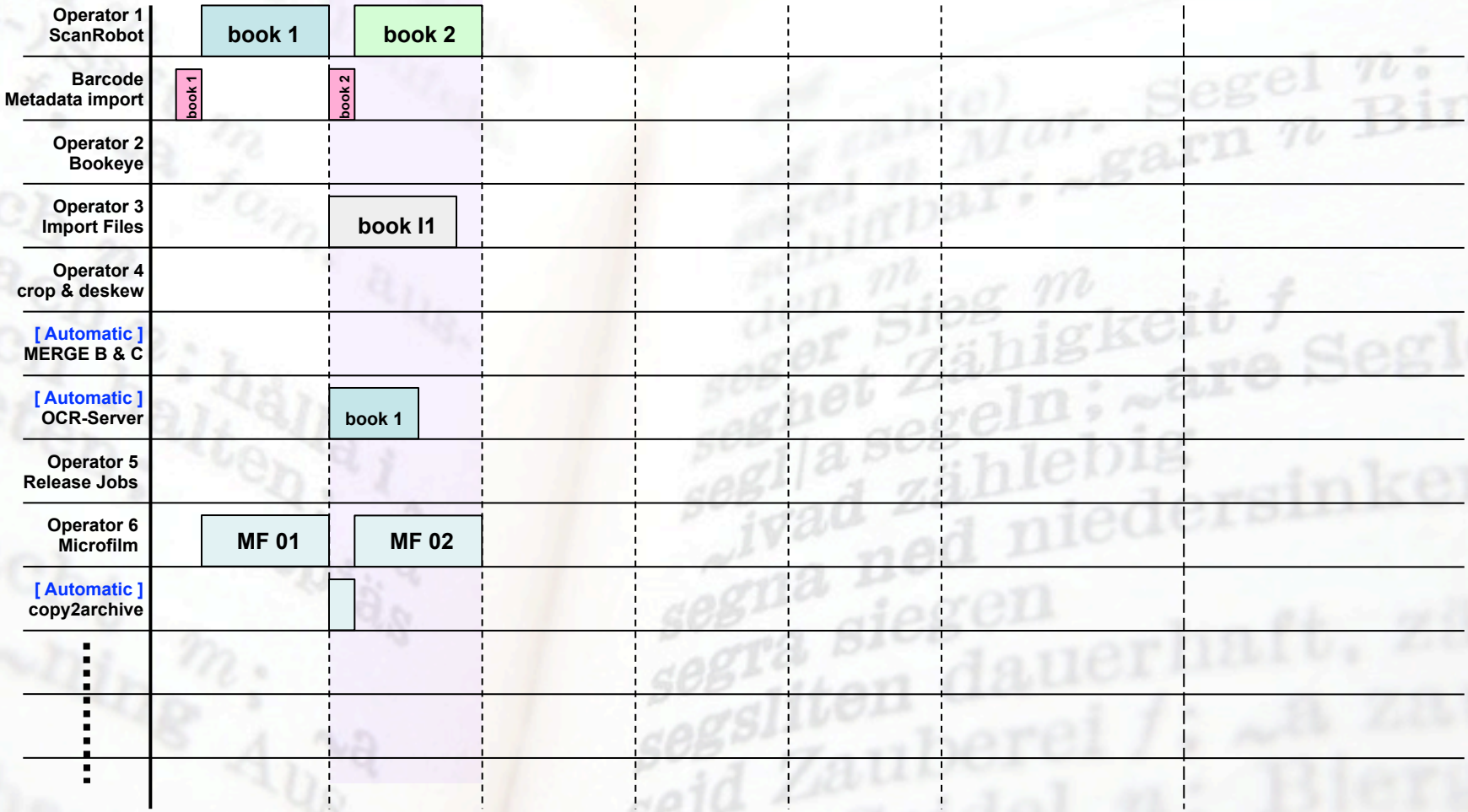


ScanFlow™

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

TREVENTUS

ScanFlow™: Jobs in the timetable

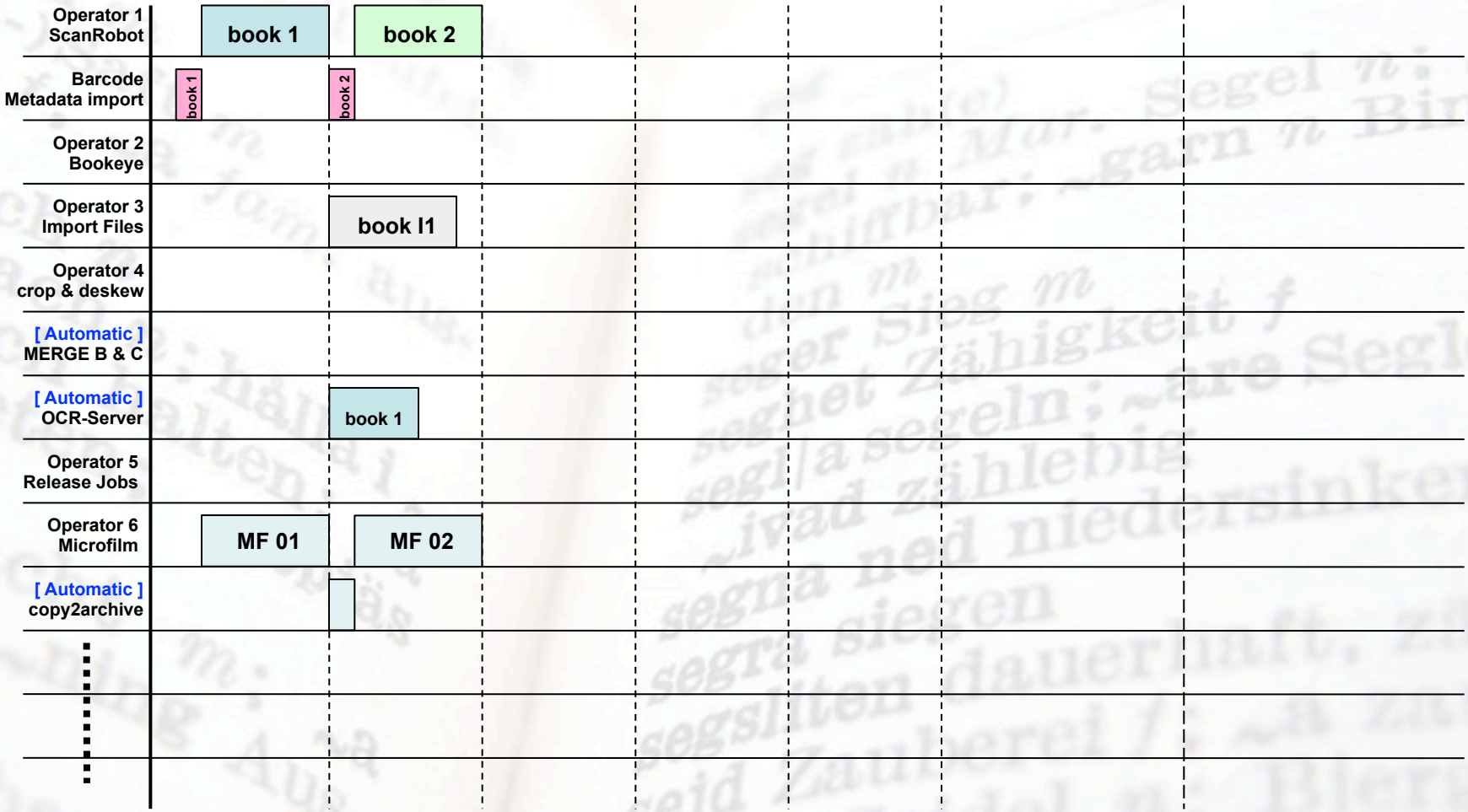


ScanFlow™

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

TREVENTUS

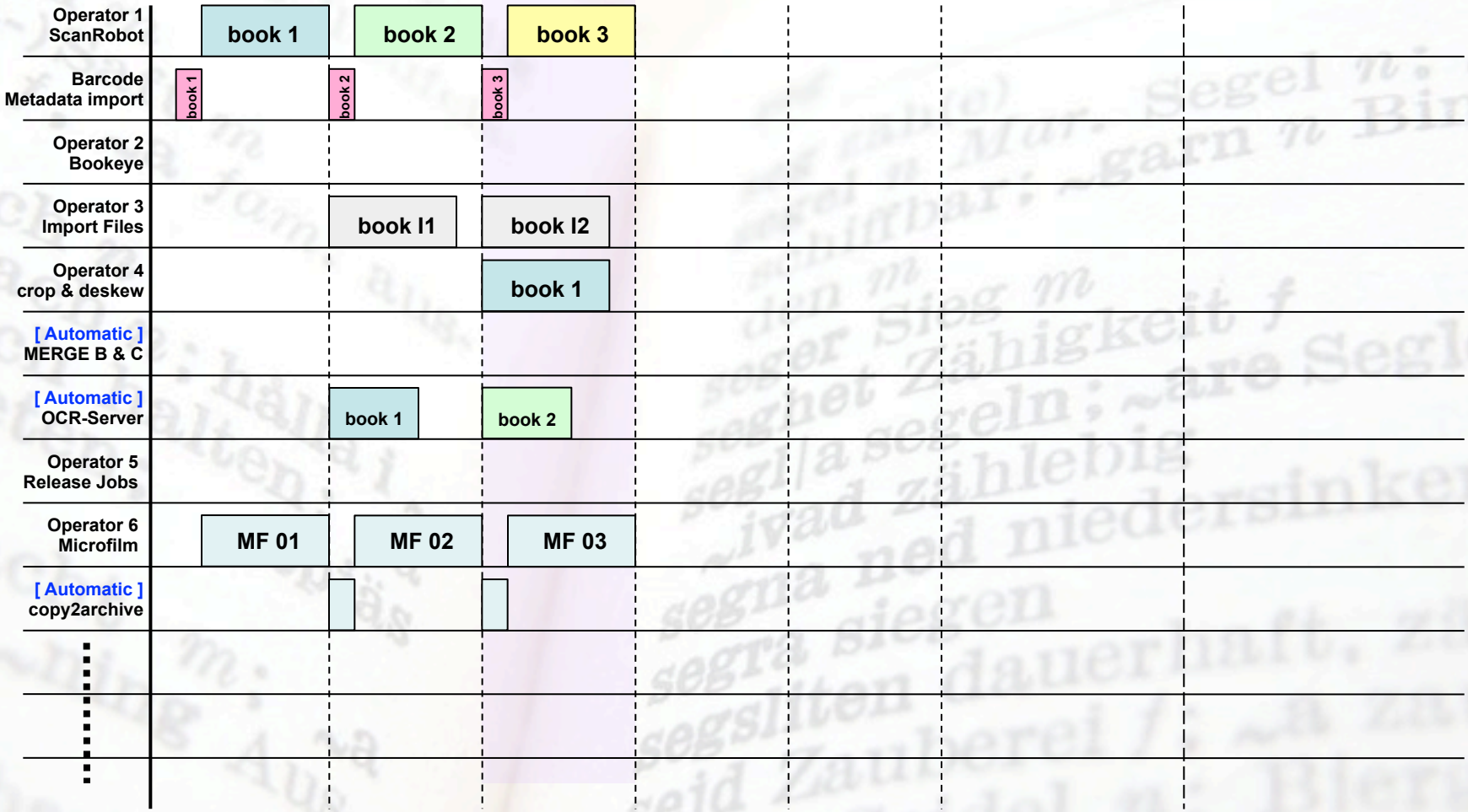
ScanFlow™: Jobs in the timetable



ScanFlow™

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

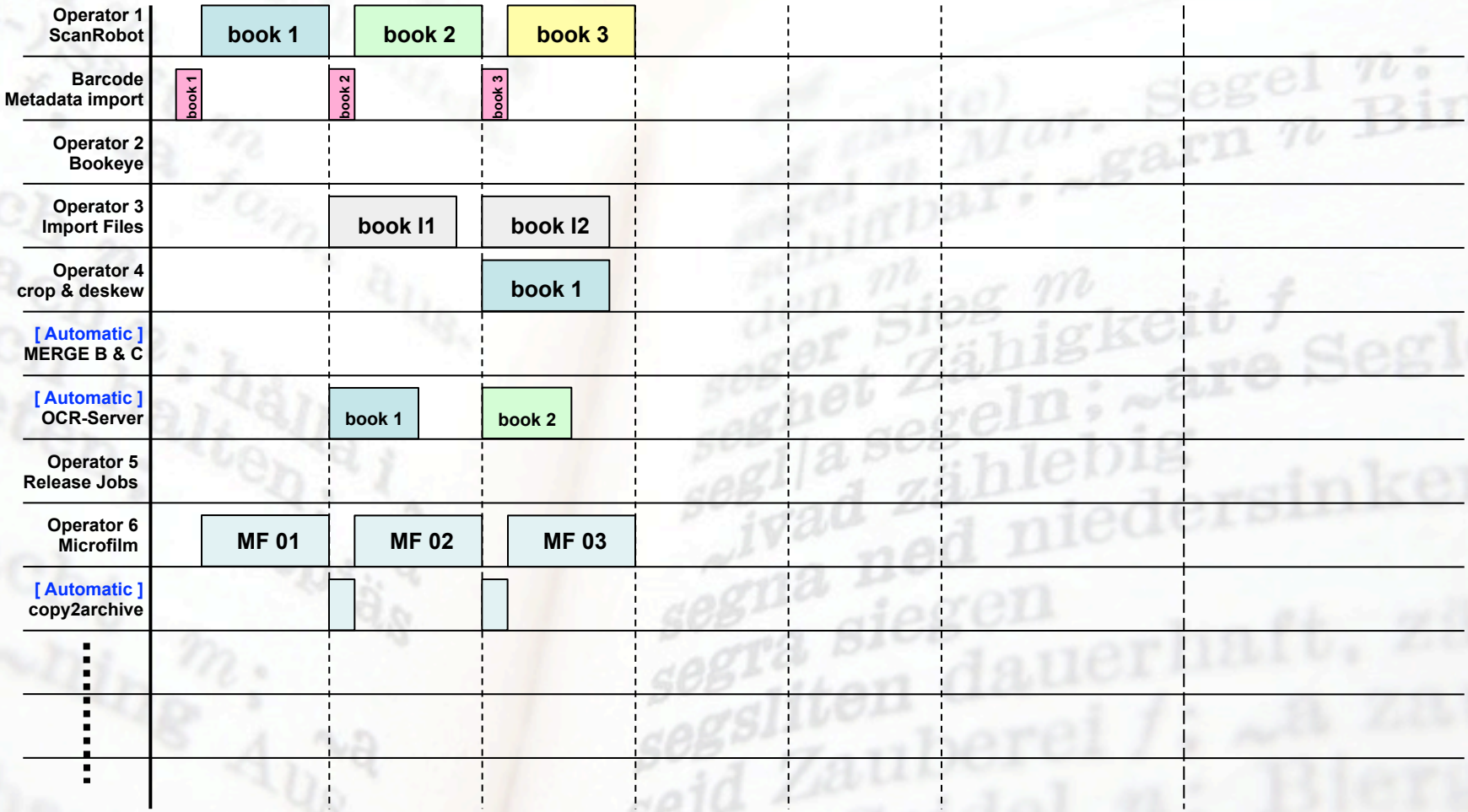
ScanFlow™: Jobs in the timetable



ScanFlow™

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

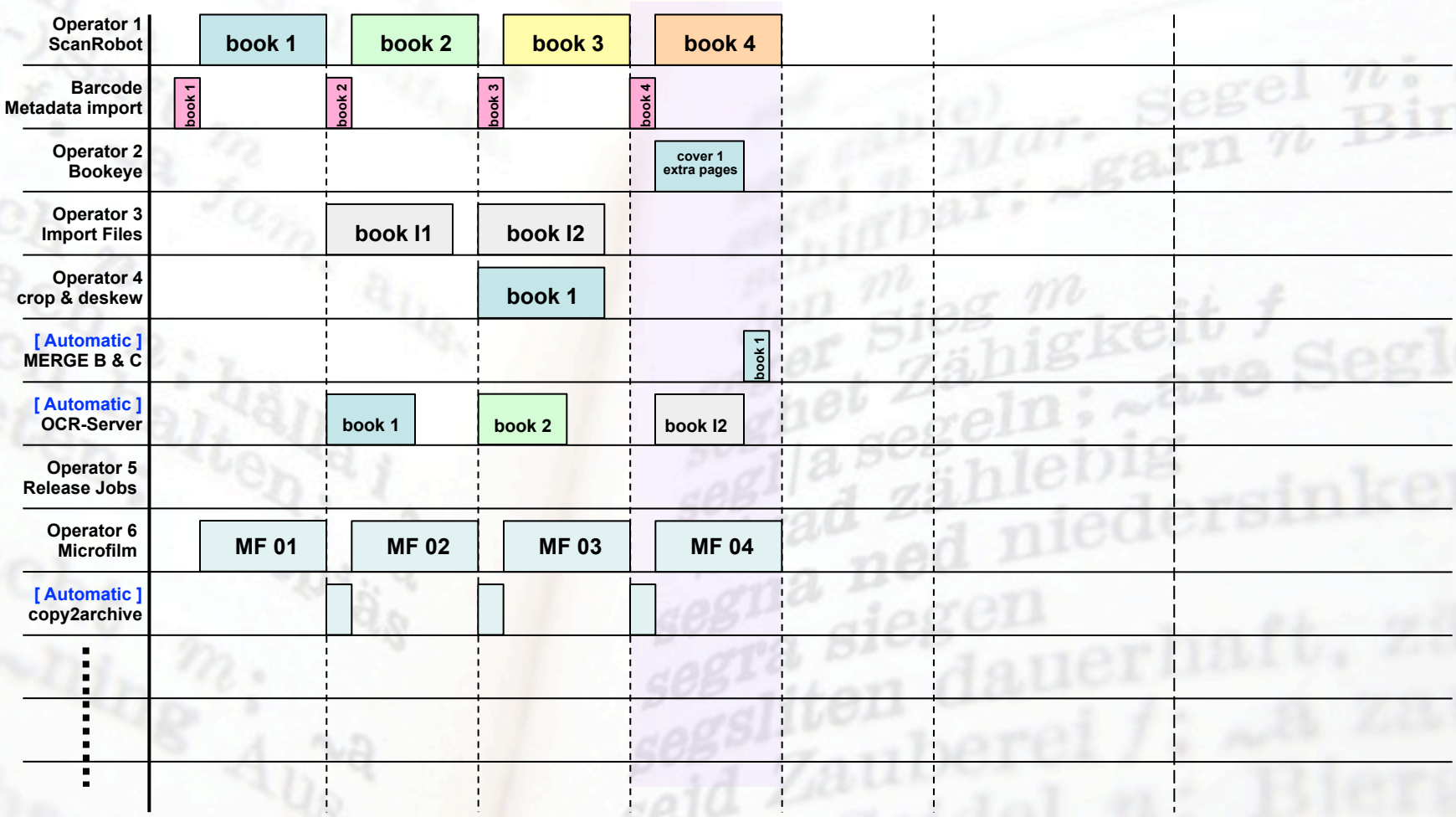
ScanFlow™: Jobs in the timetable



ScanFlow™

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

ScanFlow™: Jobs in the timetable

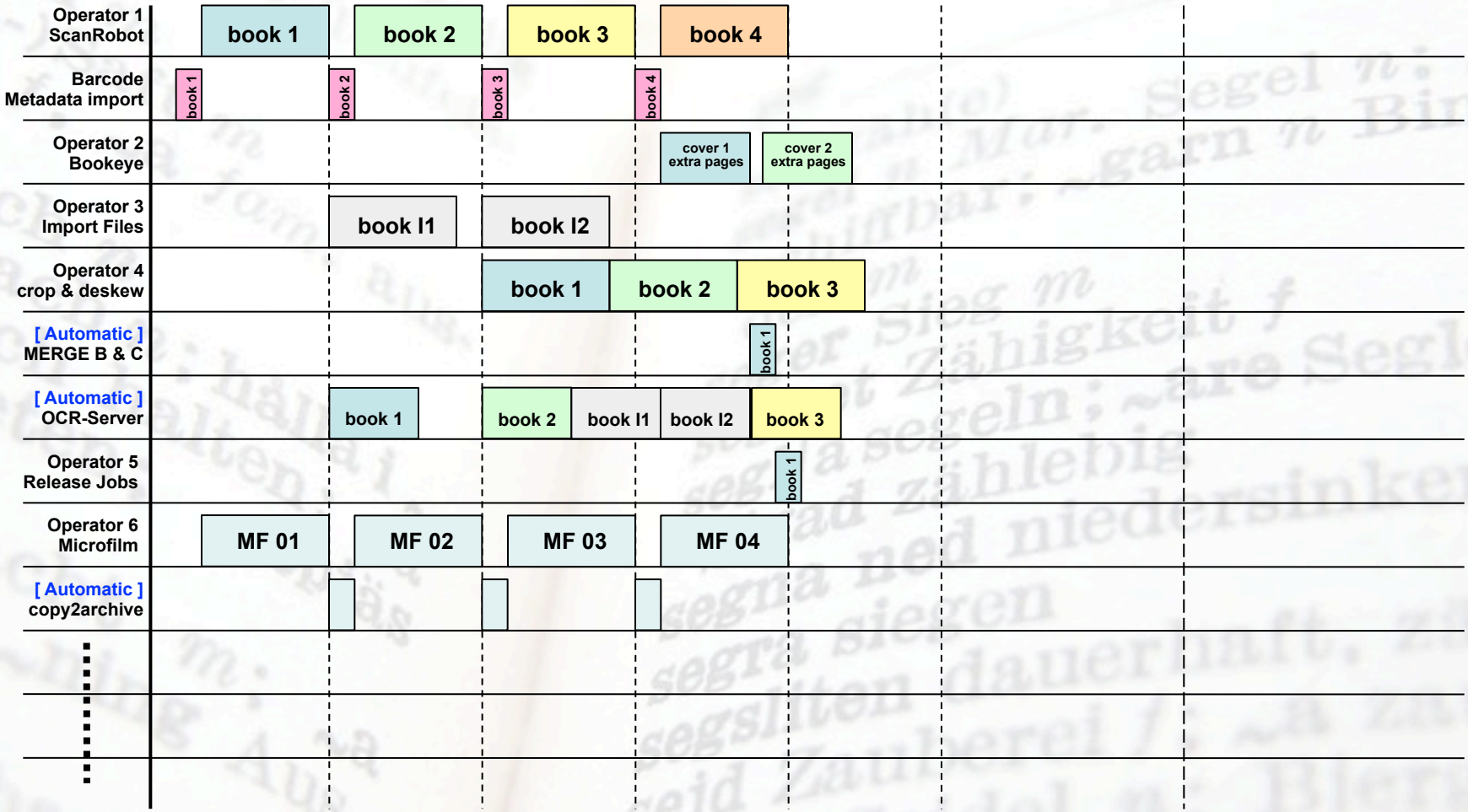


ScanFlow™

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

TREVENTUS

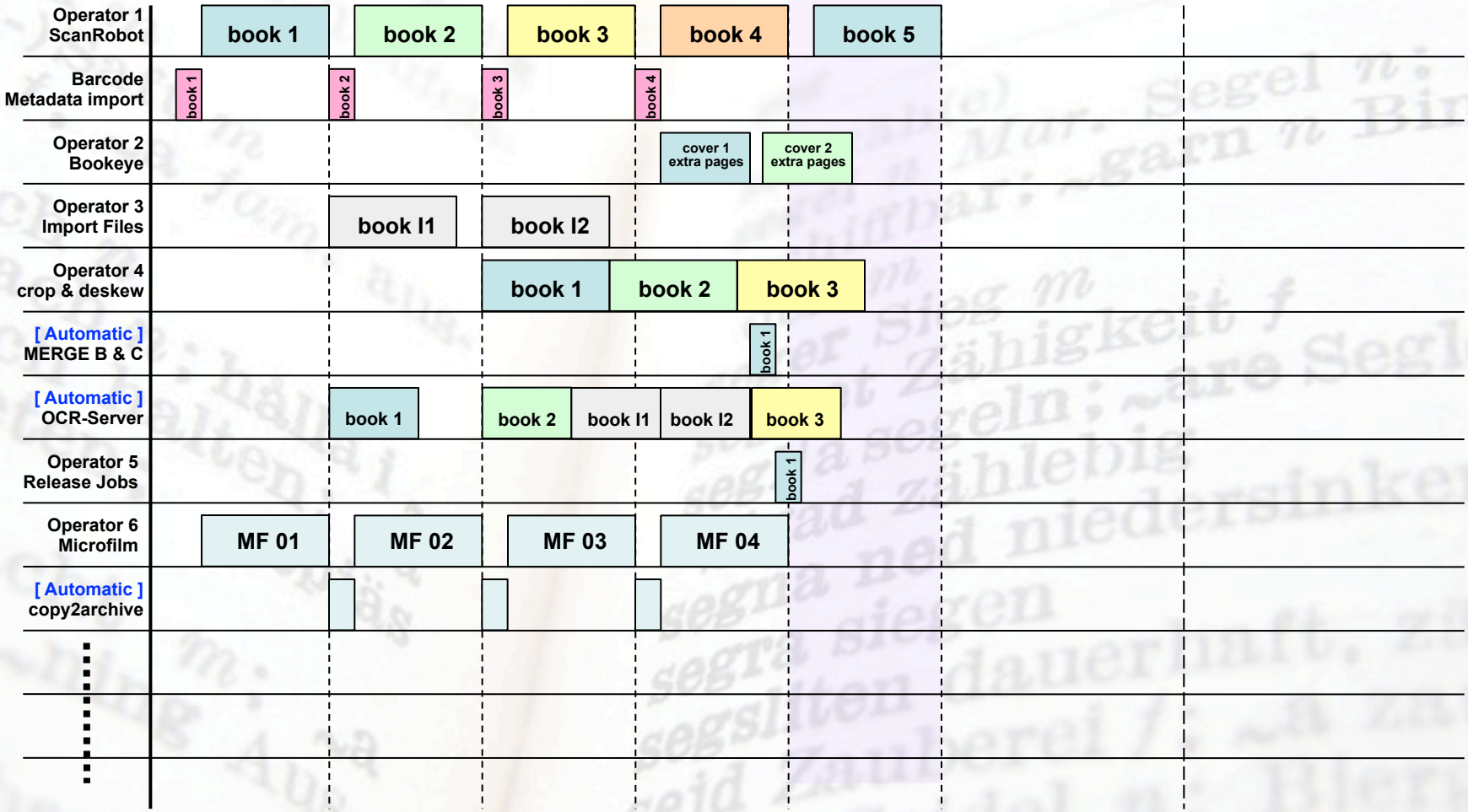
ScanFlow™: Jobs in the timetable



ScanFlow™

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

ScanFlow™: Jobs in the timetable

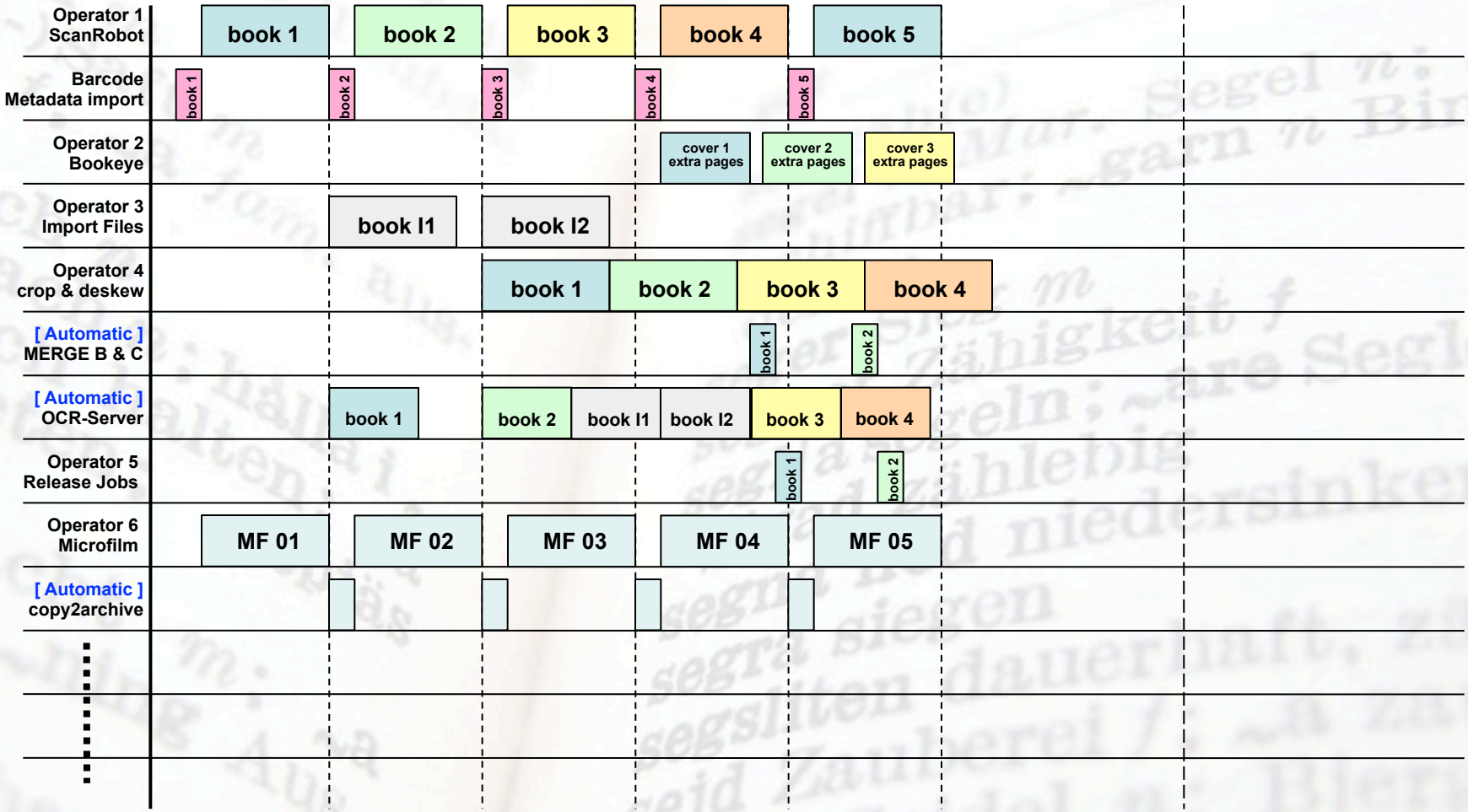


ScanFlow™

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

TREVENTUS

ScanFlow™: Jobs in the timetable

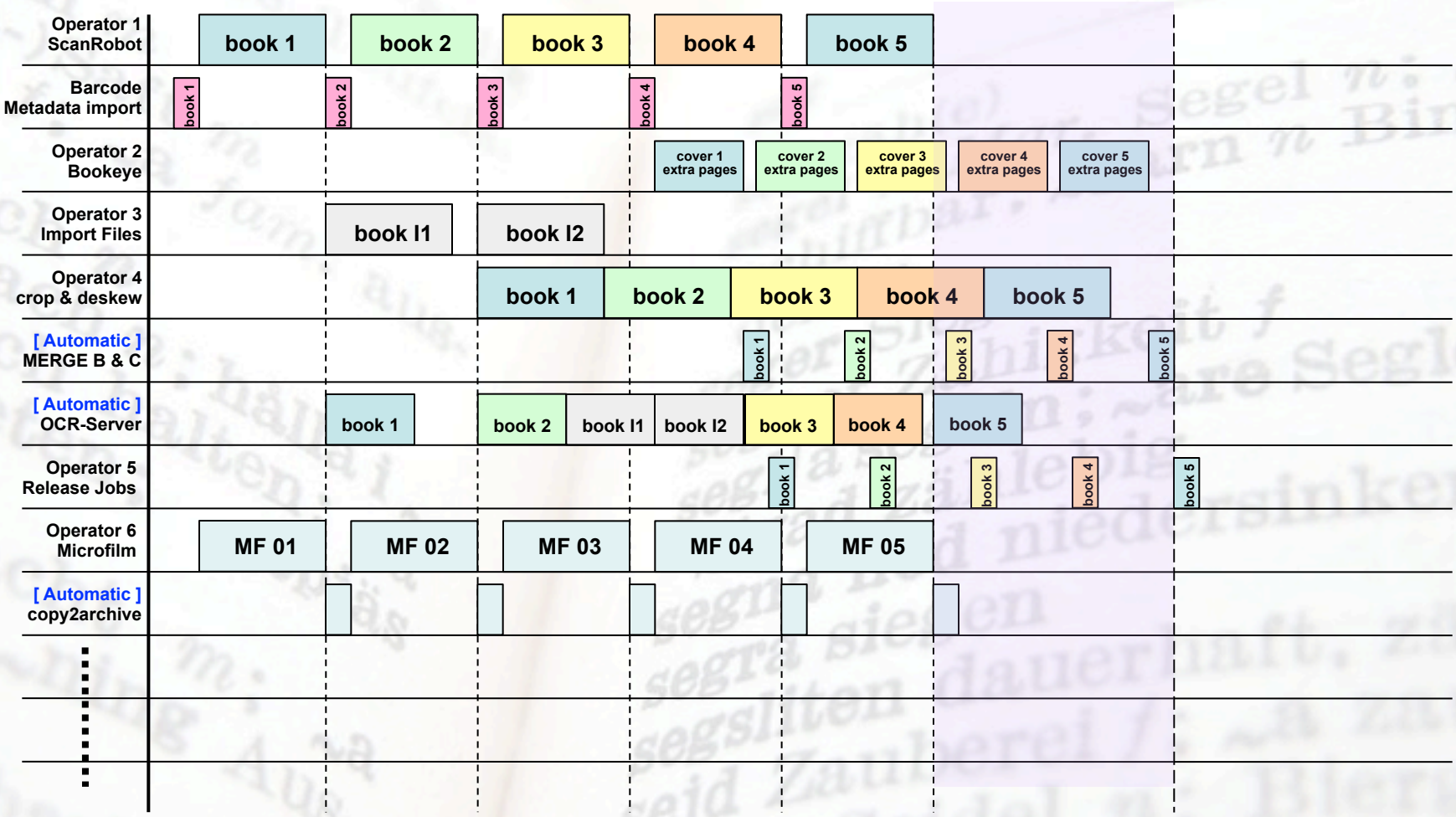


ScanFlow™

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

TREVENTUS

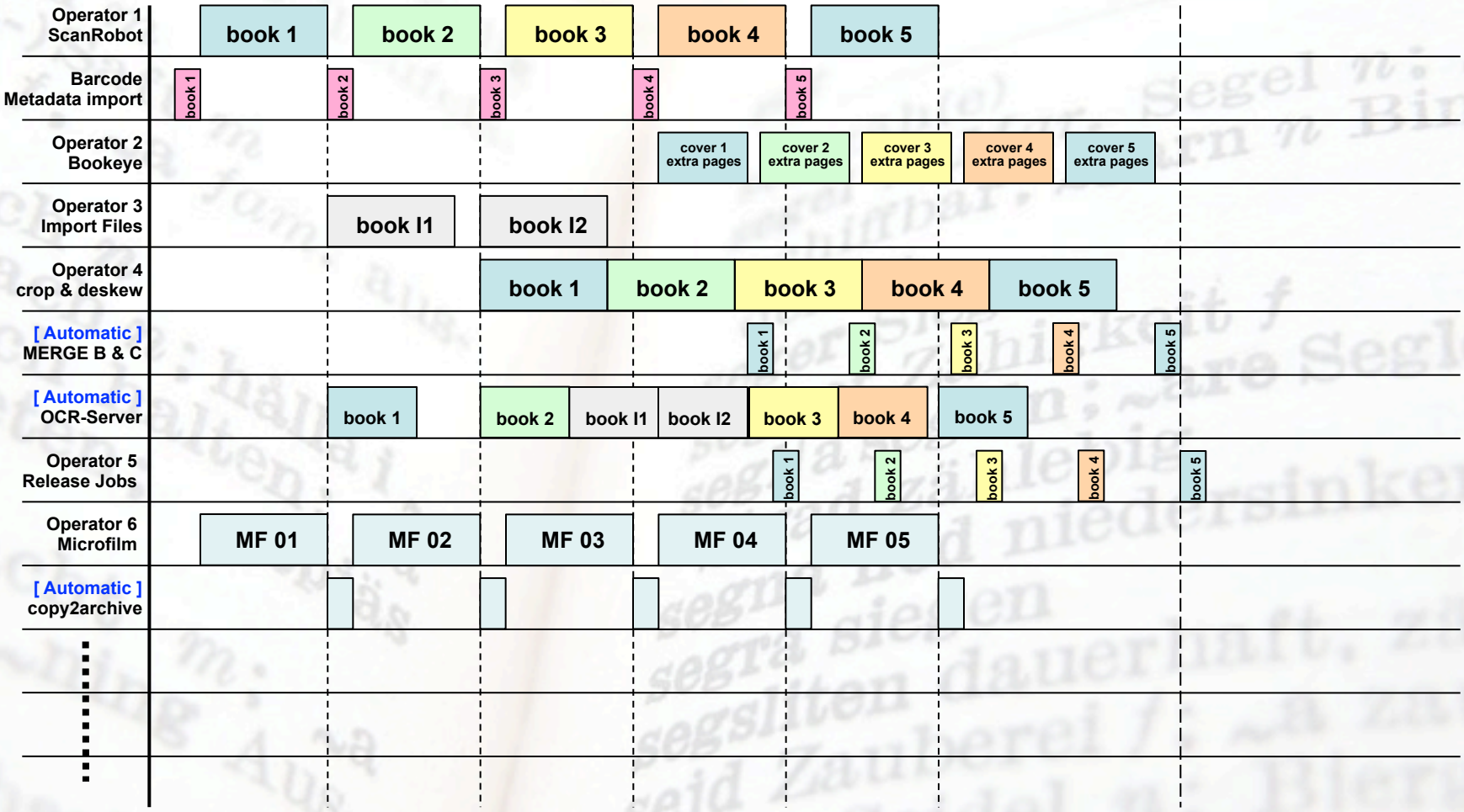
ScanFlow™: Jobs in the timetable



ScanFlow™

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

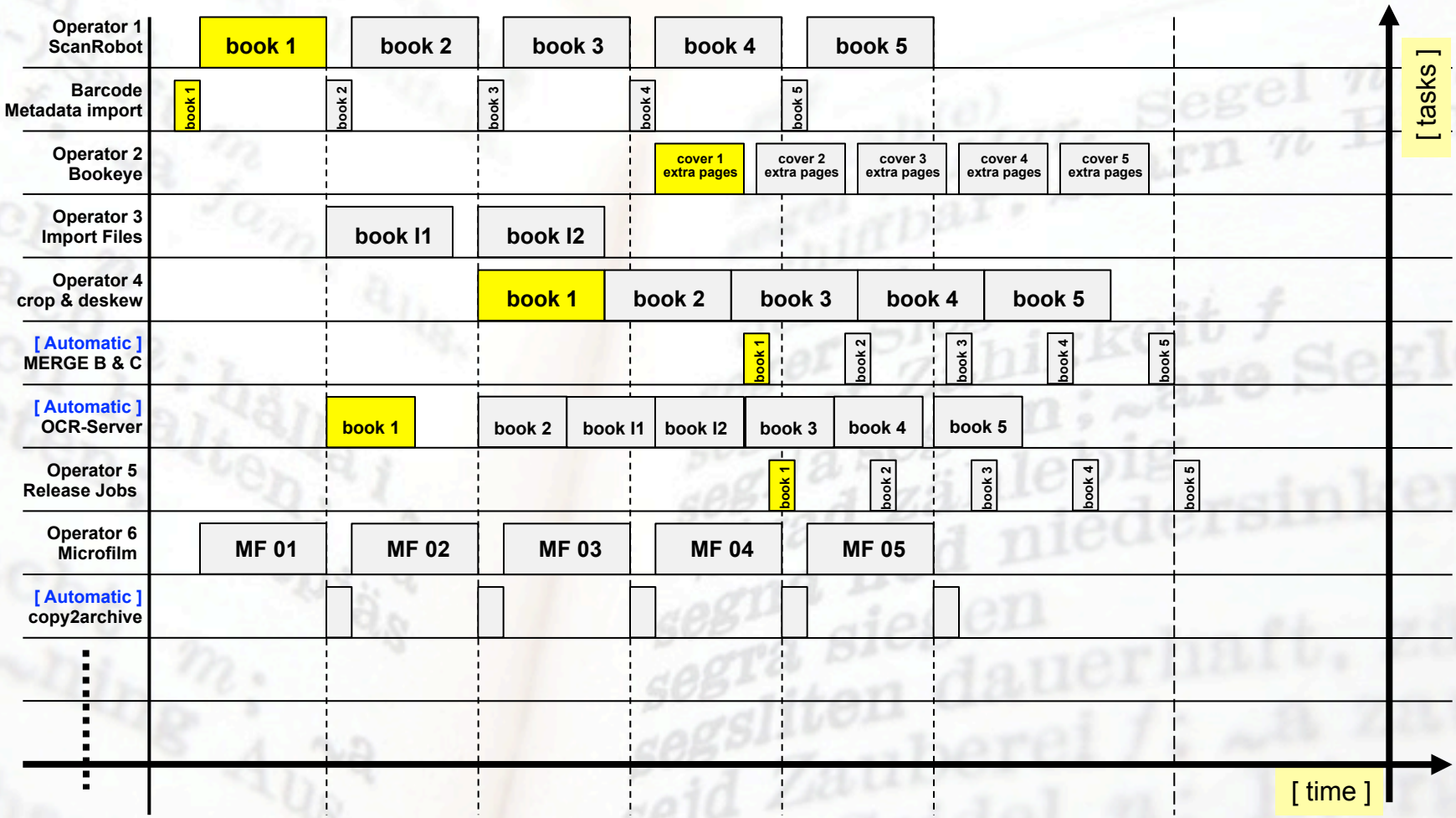
ScanFlow™: Jobs in the timetable



ScanFlow™

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

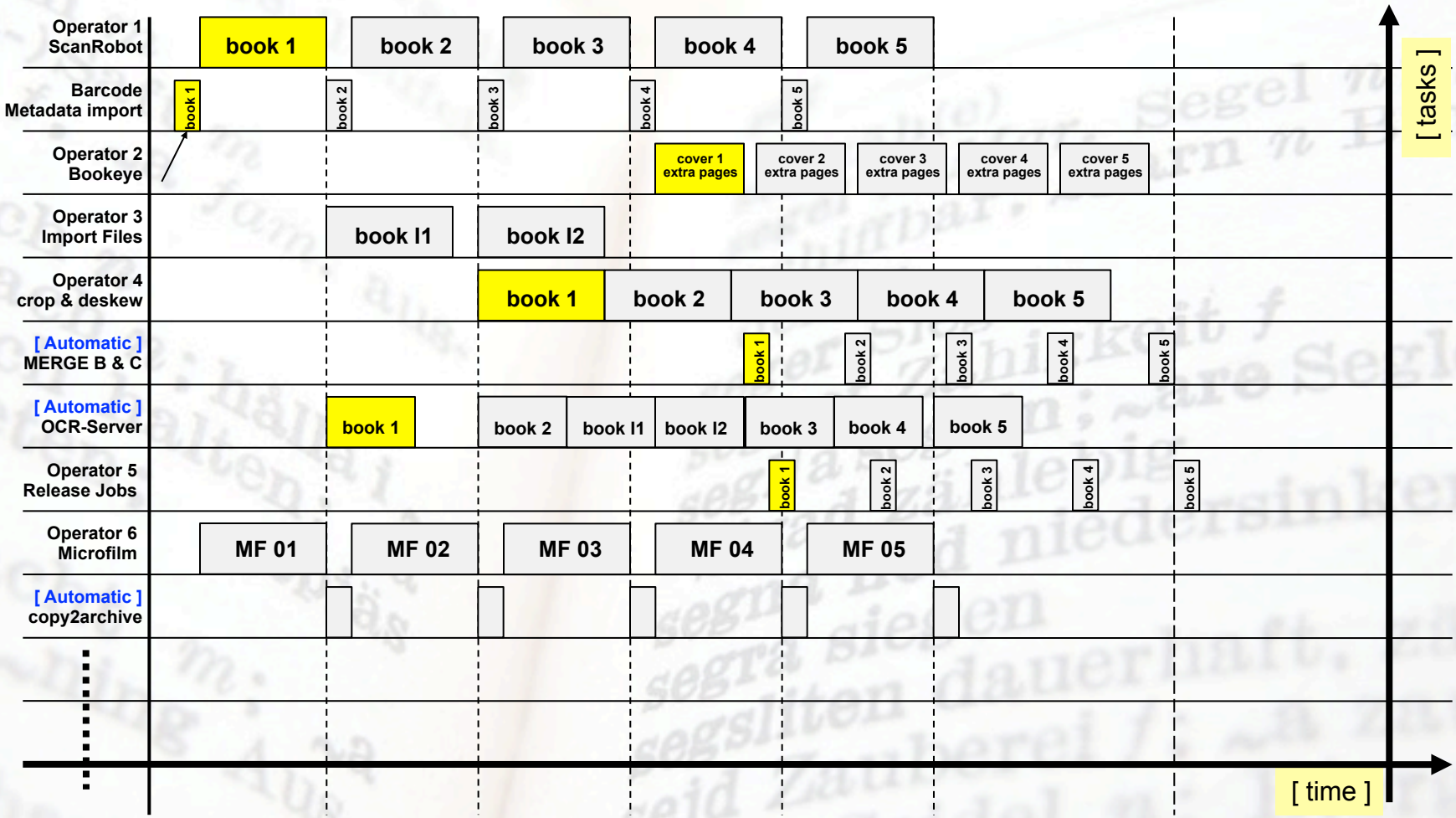
ScanFlow™: Jobs in the timetable



ScanFlow™

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

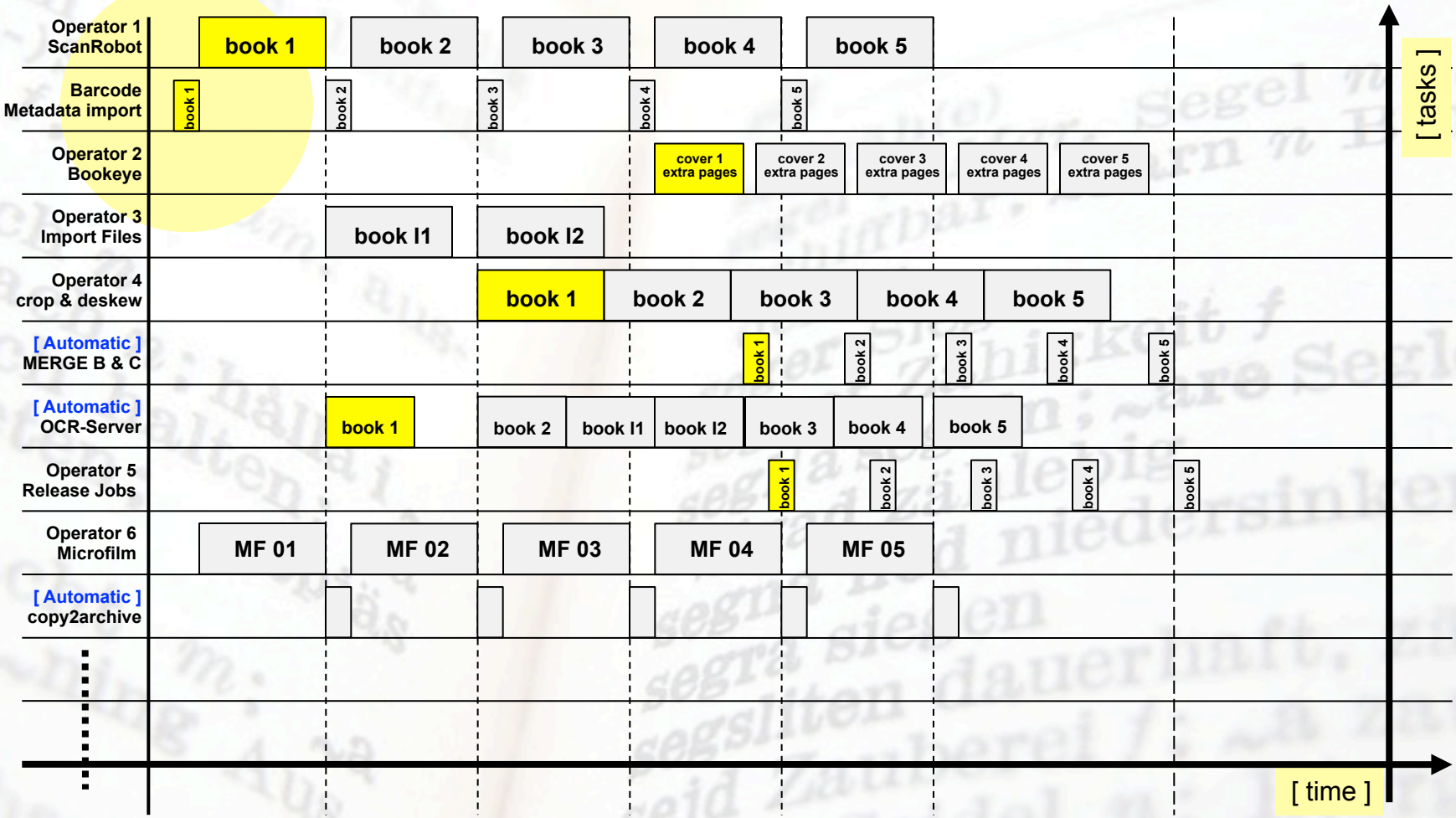
ScanFlow™: Jobs in the timetable



ScanFlow™

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

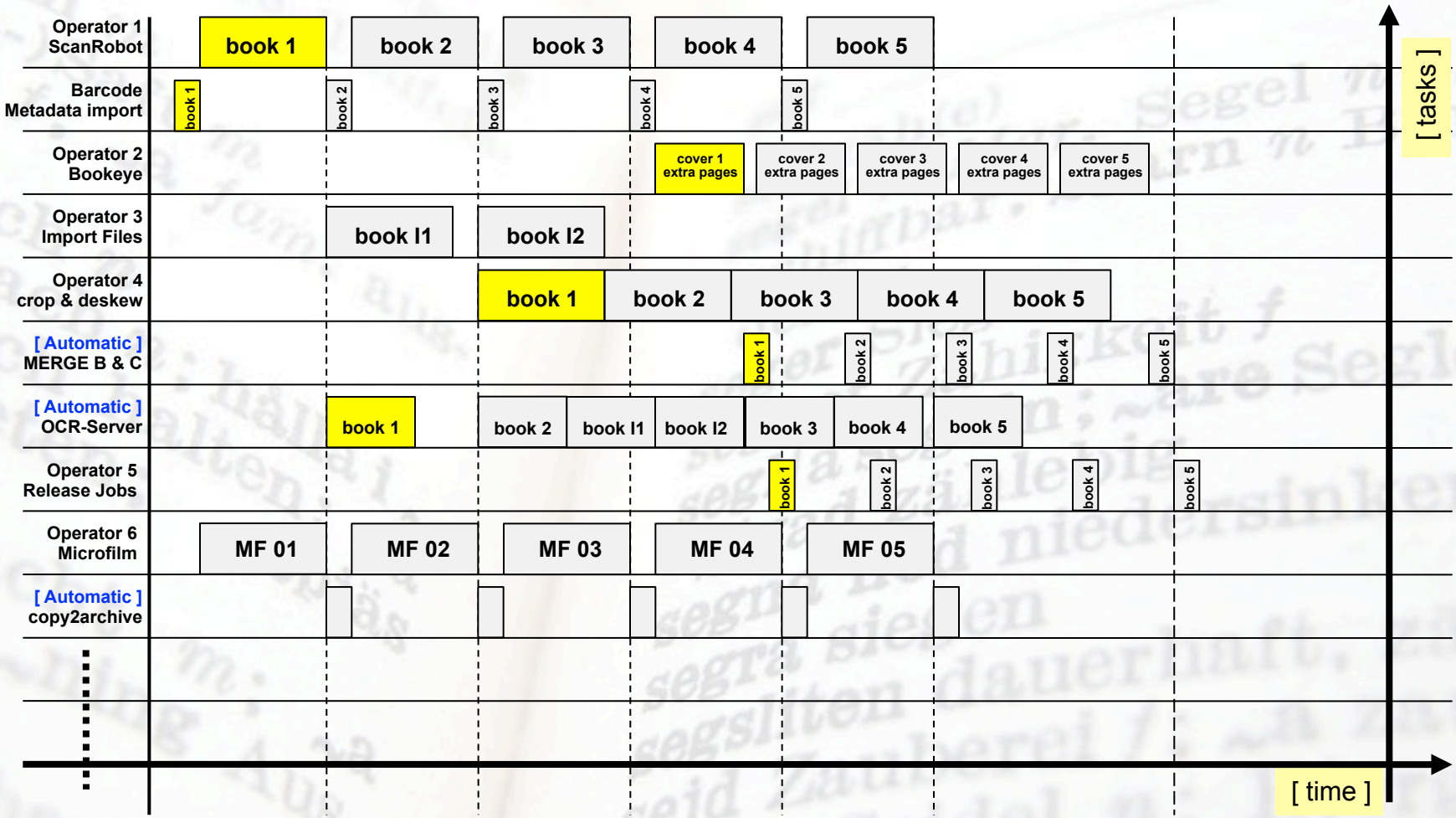
ScanFlow™: Jobs in the timetable



ScanFlow™

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

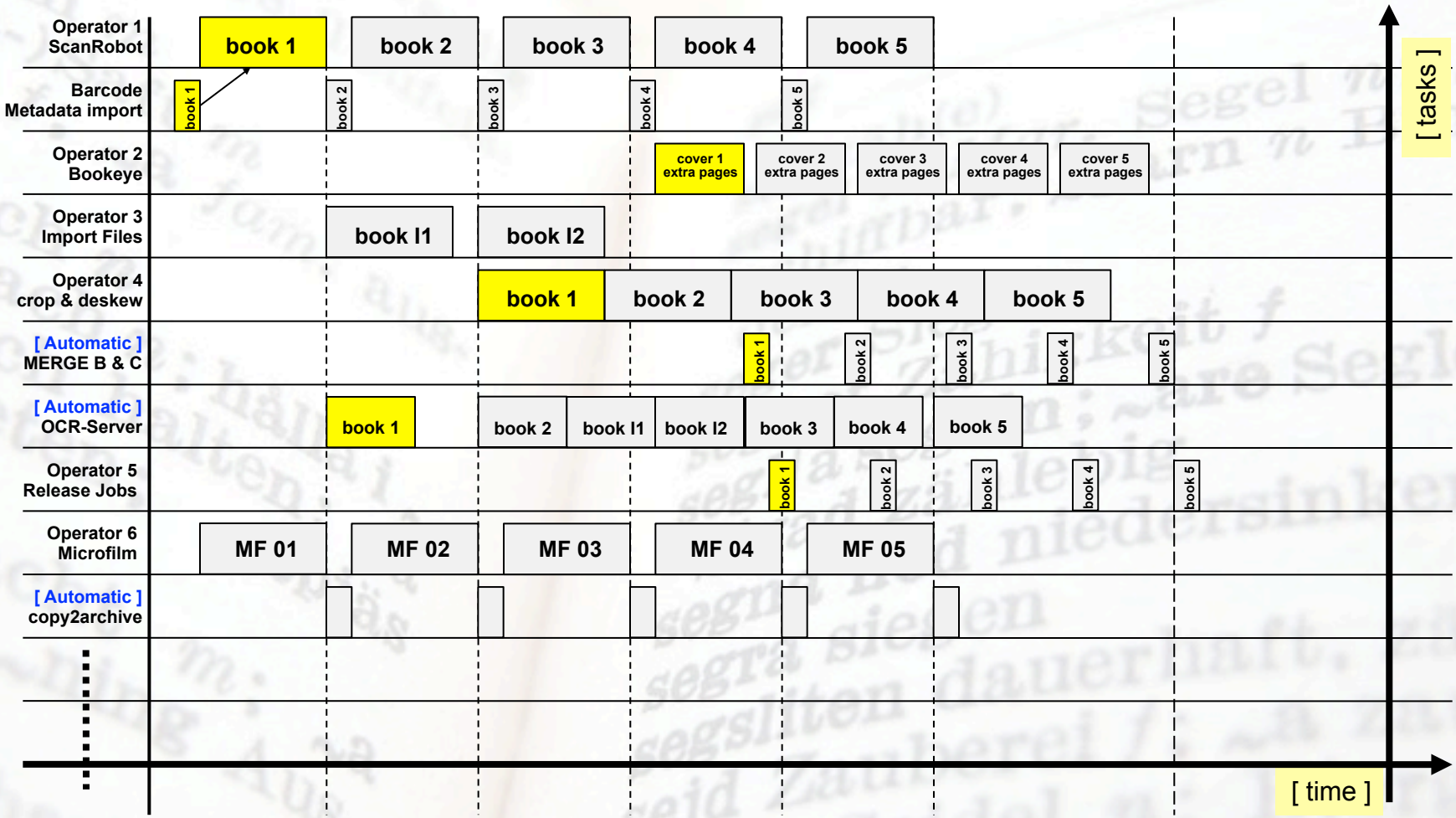
ScanFlow™: Jobs in the timetable



ScanFlow™

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

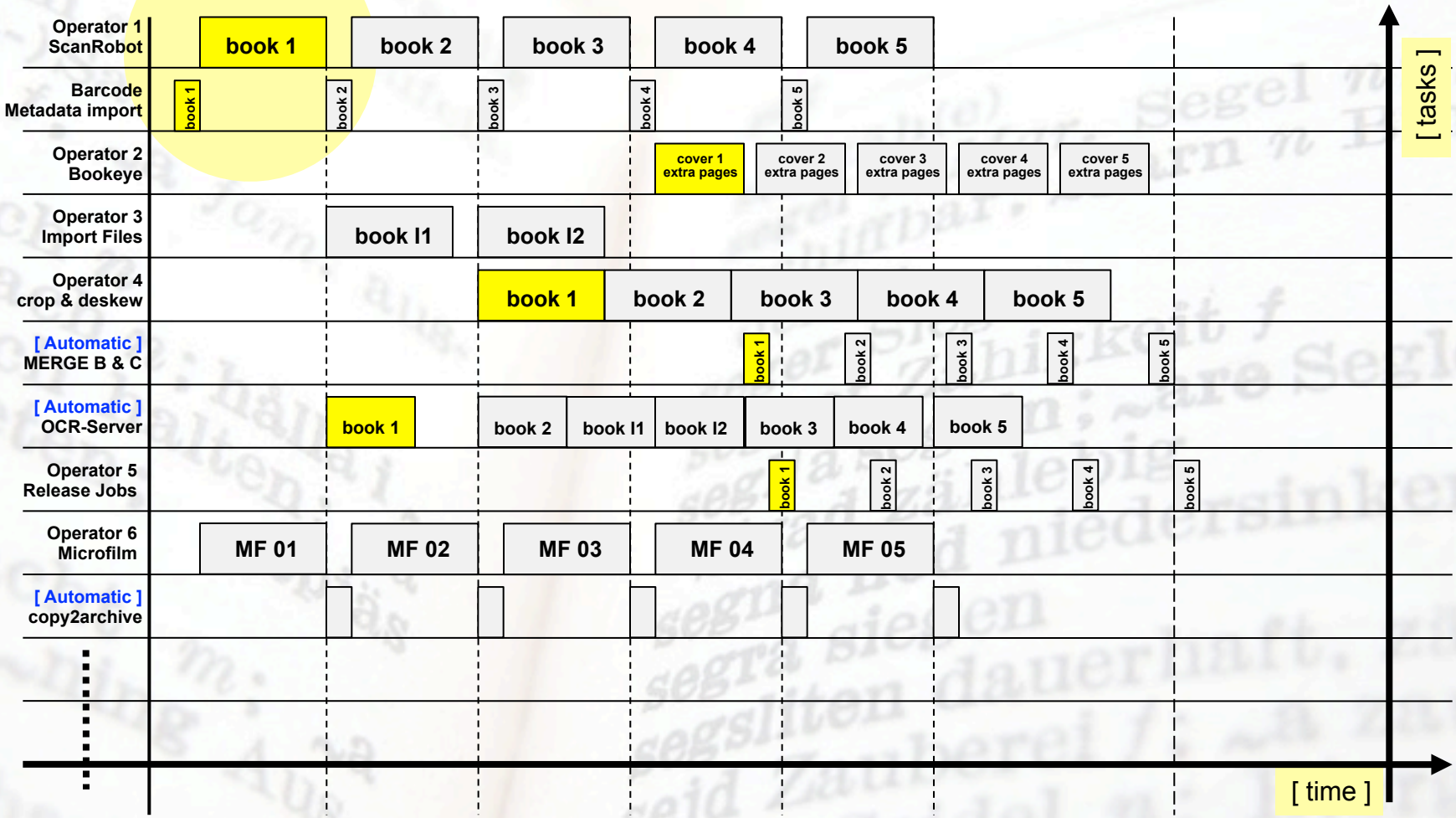
ScanFlow™: Jobs in the timetable



ScanFlow™

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

ScanFlow™: Jobs in the timetable

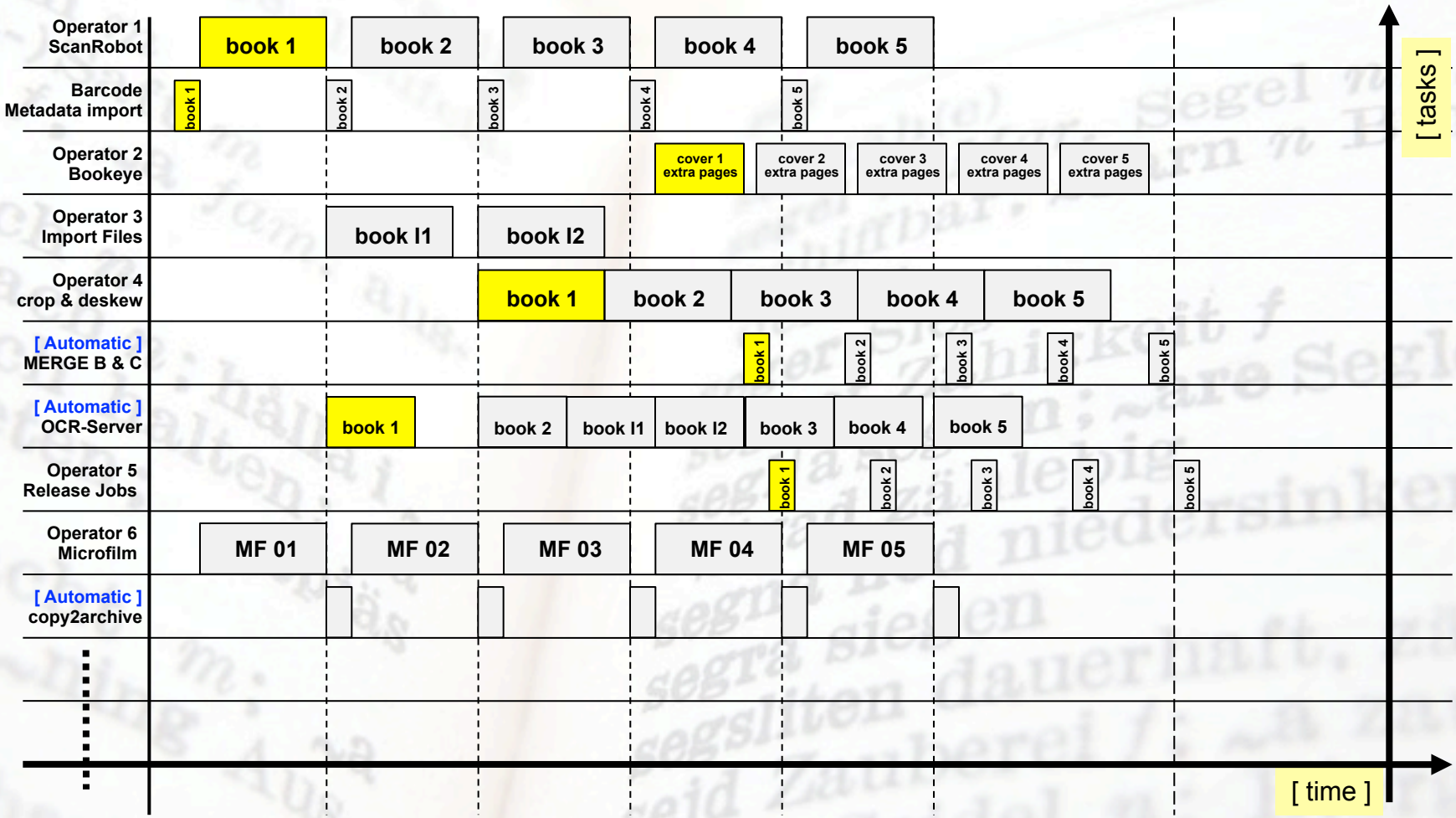


ScanFlow™

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

TREVENTUS

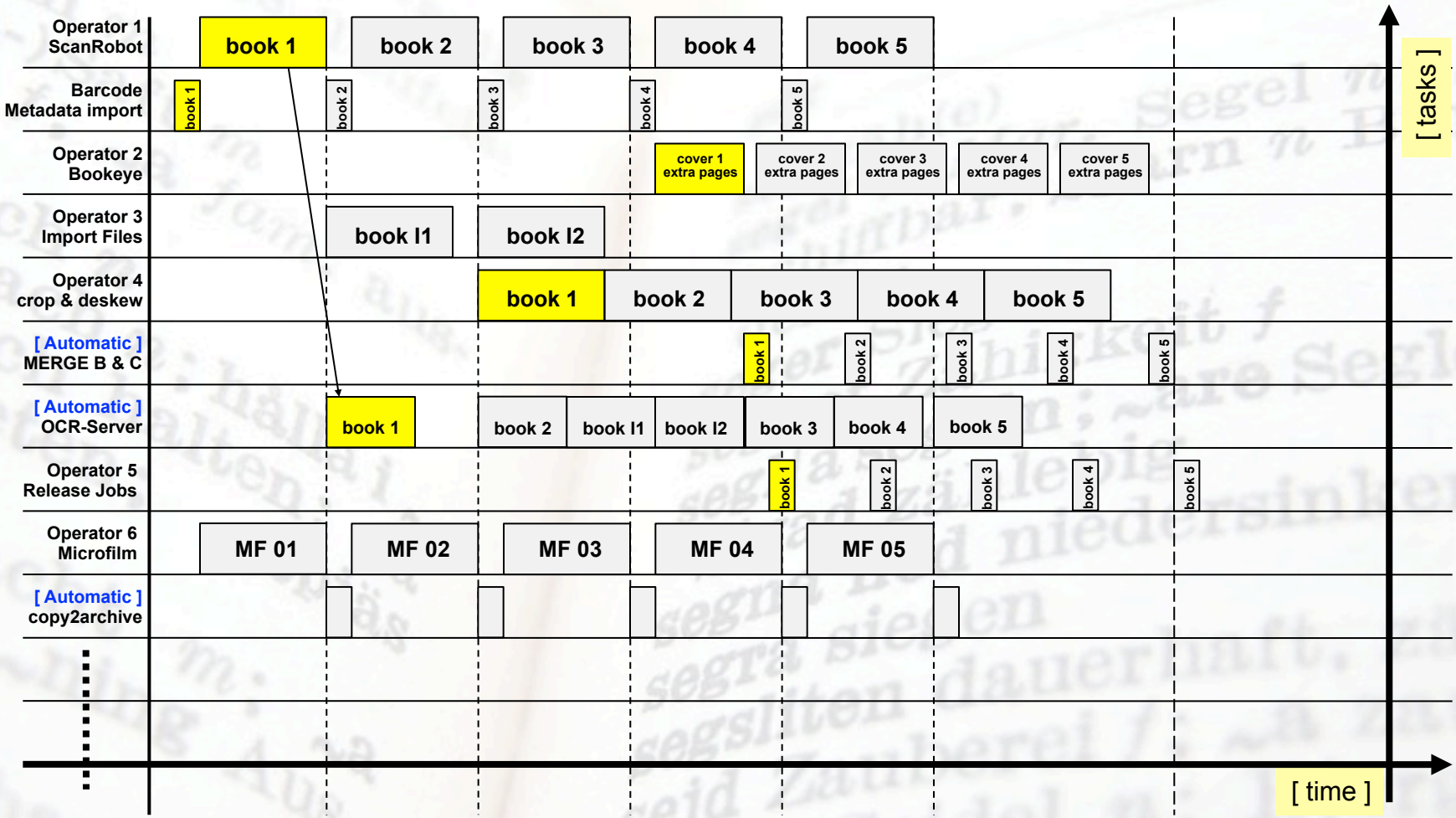
ScanFlow™: Jobs in the timetable



ScanFlow™

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

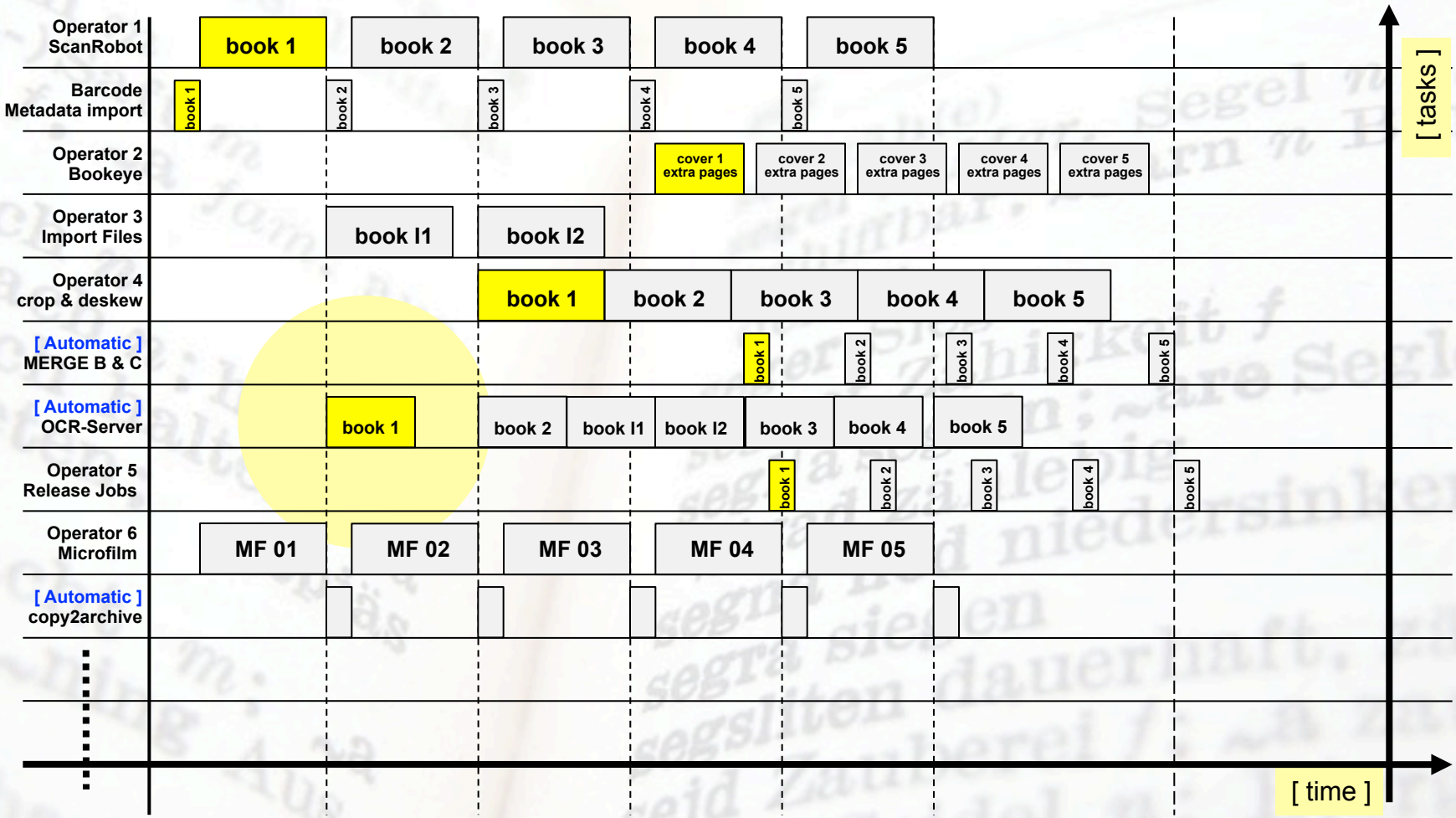
ScanFlow™: Jobs in the timetable



ScanFlow™

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

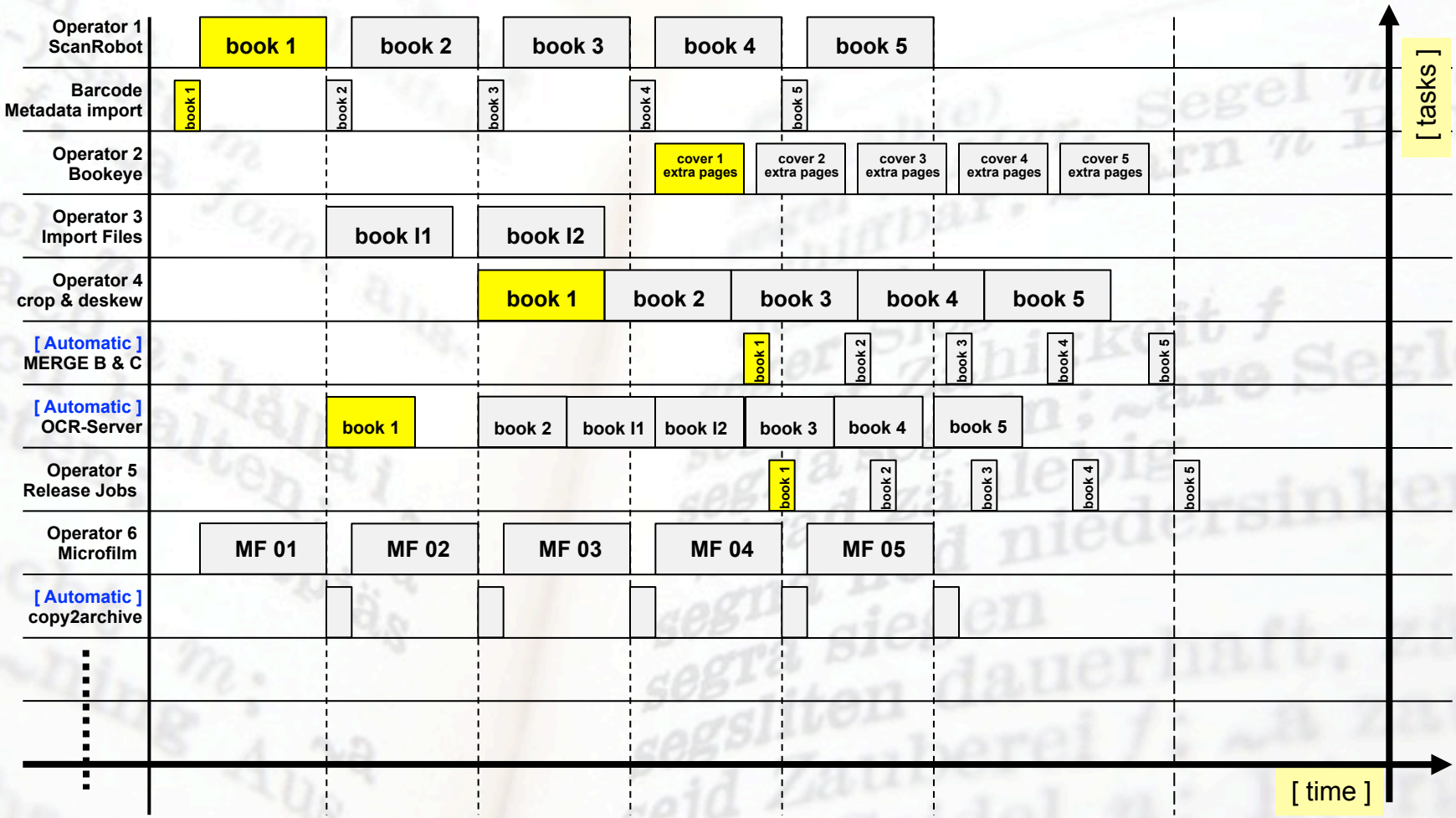
ScanFlow™: Jobs in the timetable



ScanFlow™

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

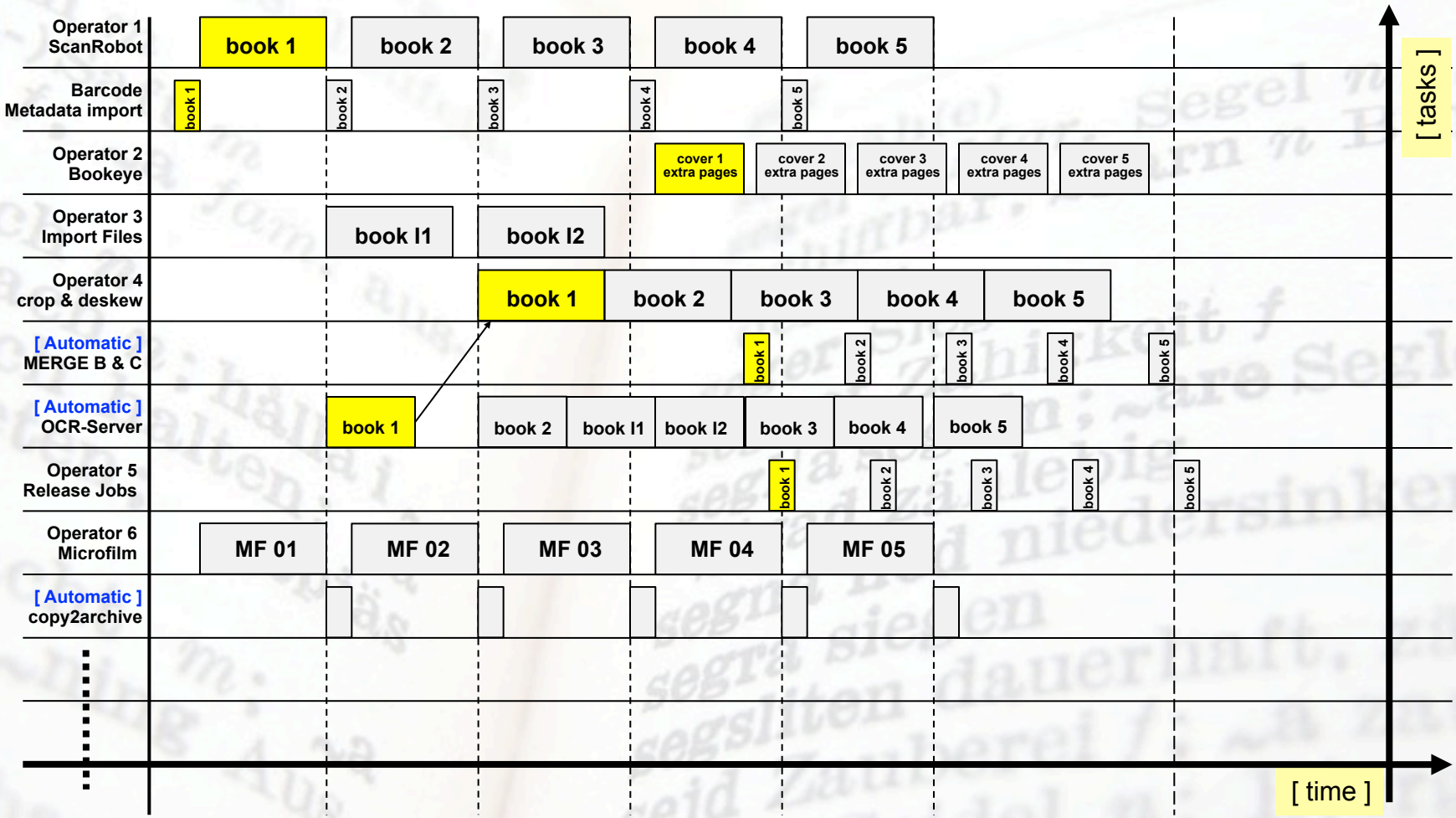
ScanFlow™: Jobs in the timetable



ScanFlow™

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

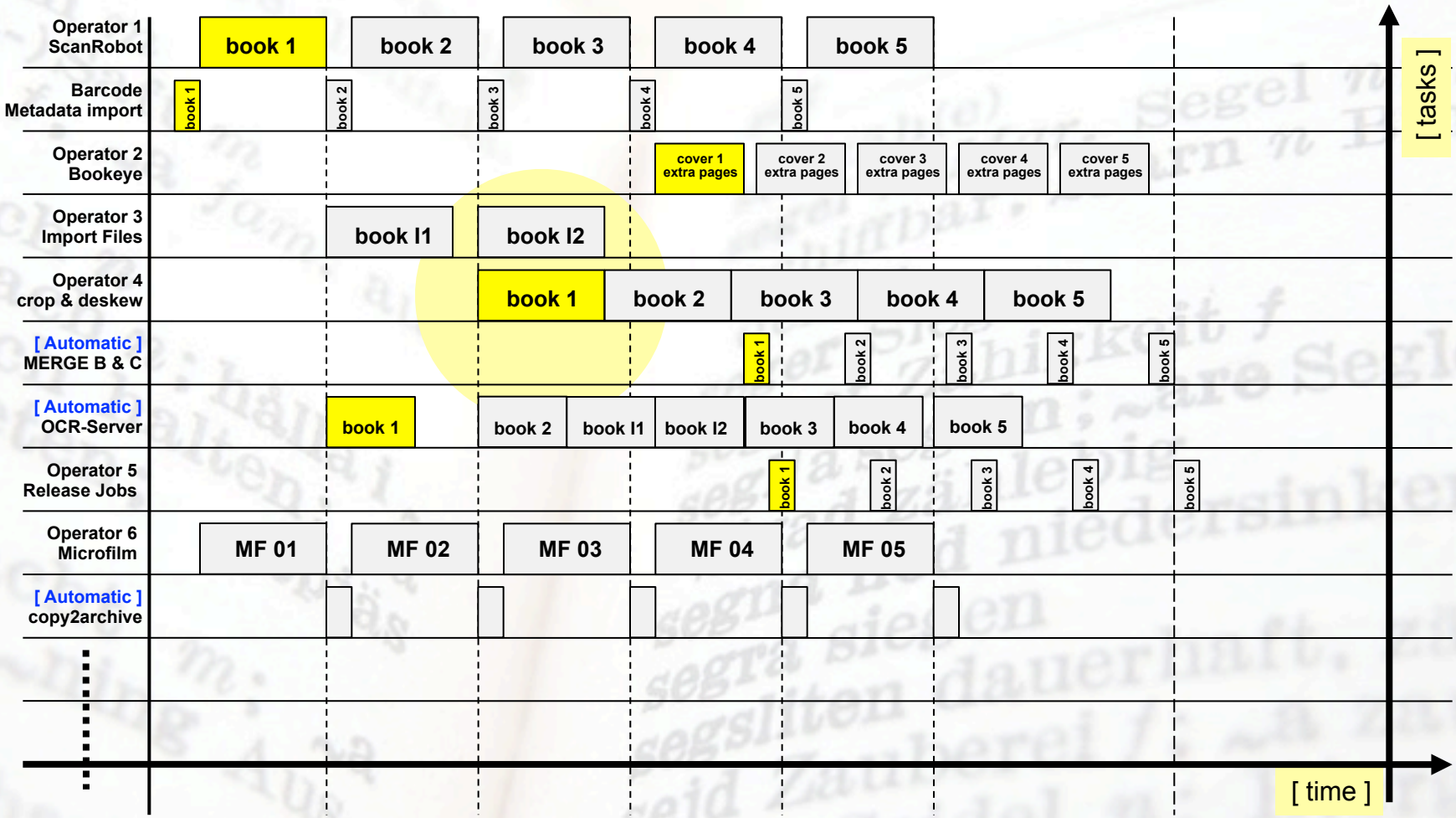
ScanFlow™: Jobs in the timetable



ScanFlow™

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

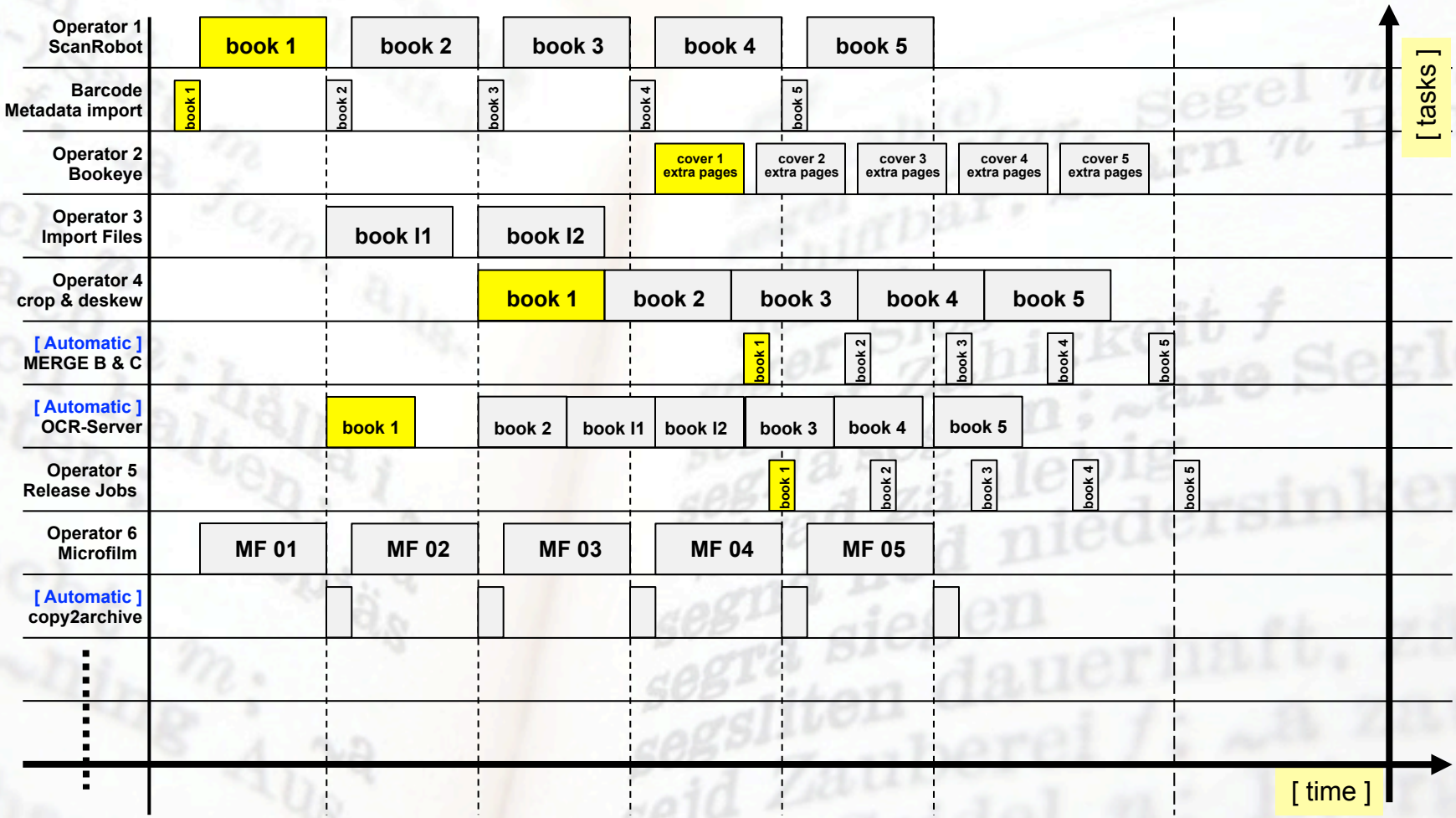
ScanFlow™: Jobs in the timetable



ScanFlow™

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

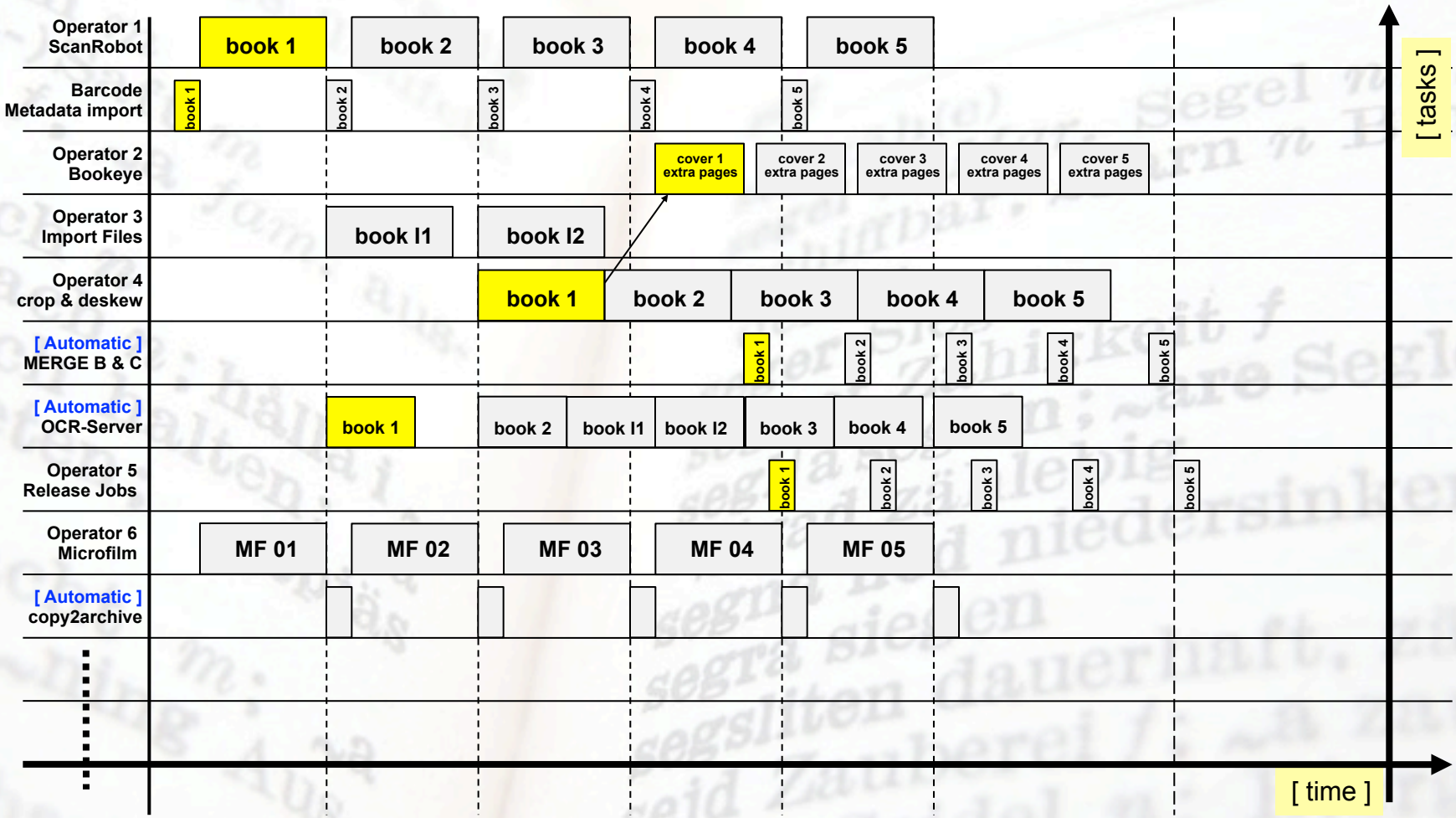
ScanFlow™: Jobs in the timetable



ScanFlow™

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

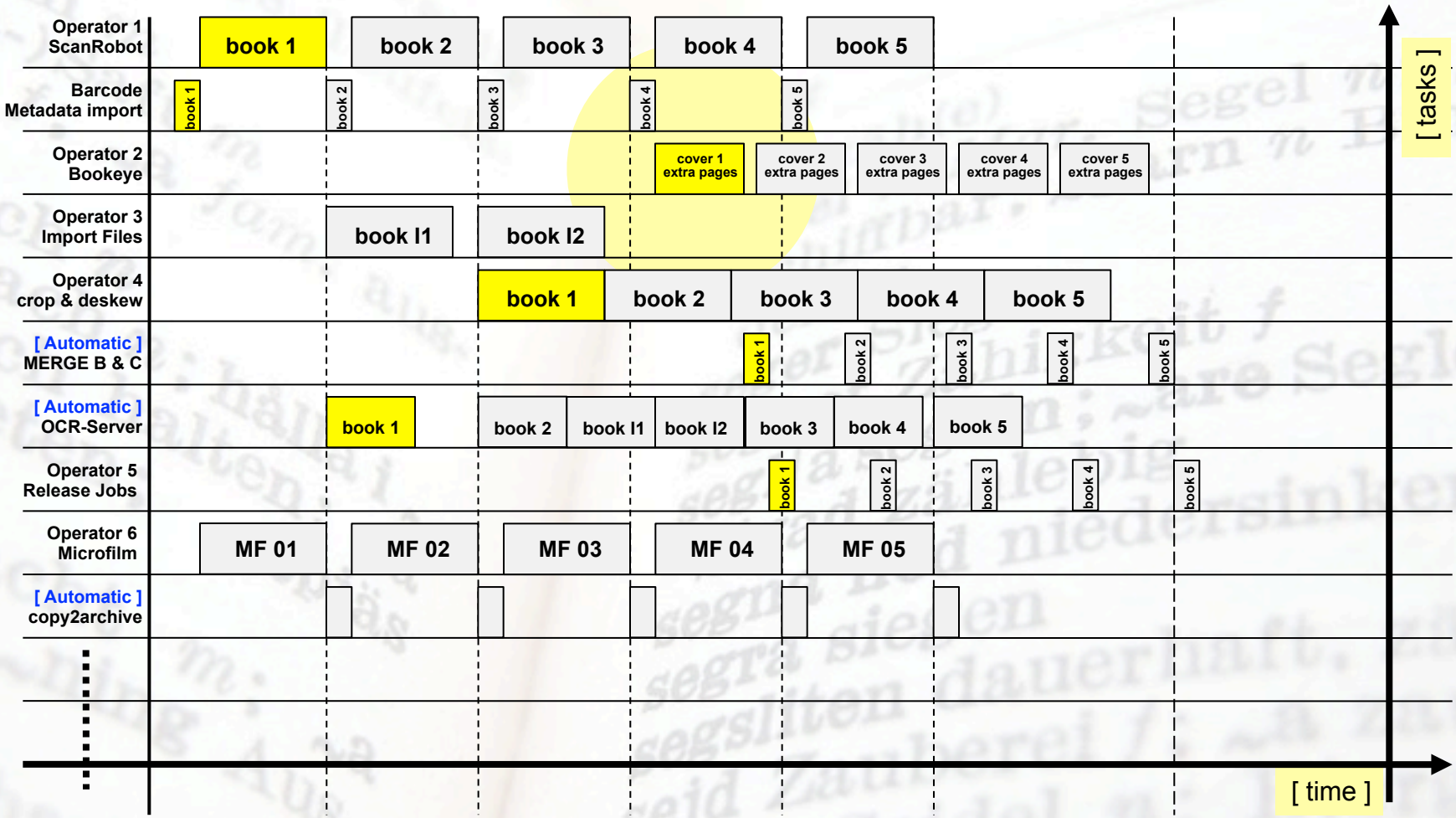
ScanFlow™: Jobs in the timetable



ScanFlow™

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

ScanFlow™: Jobs in the timetable

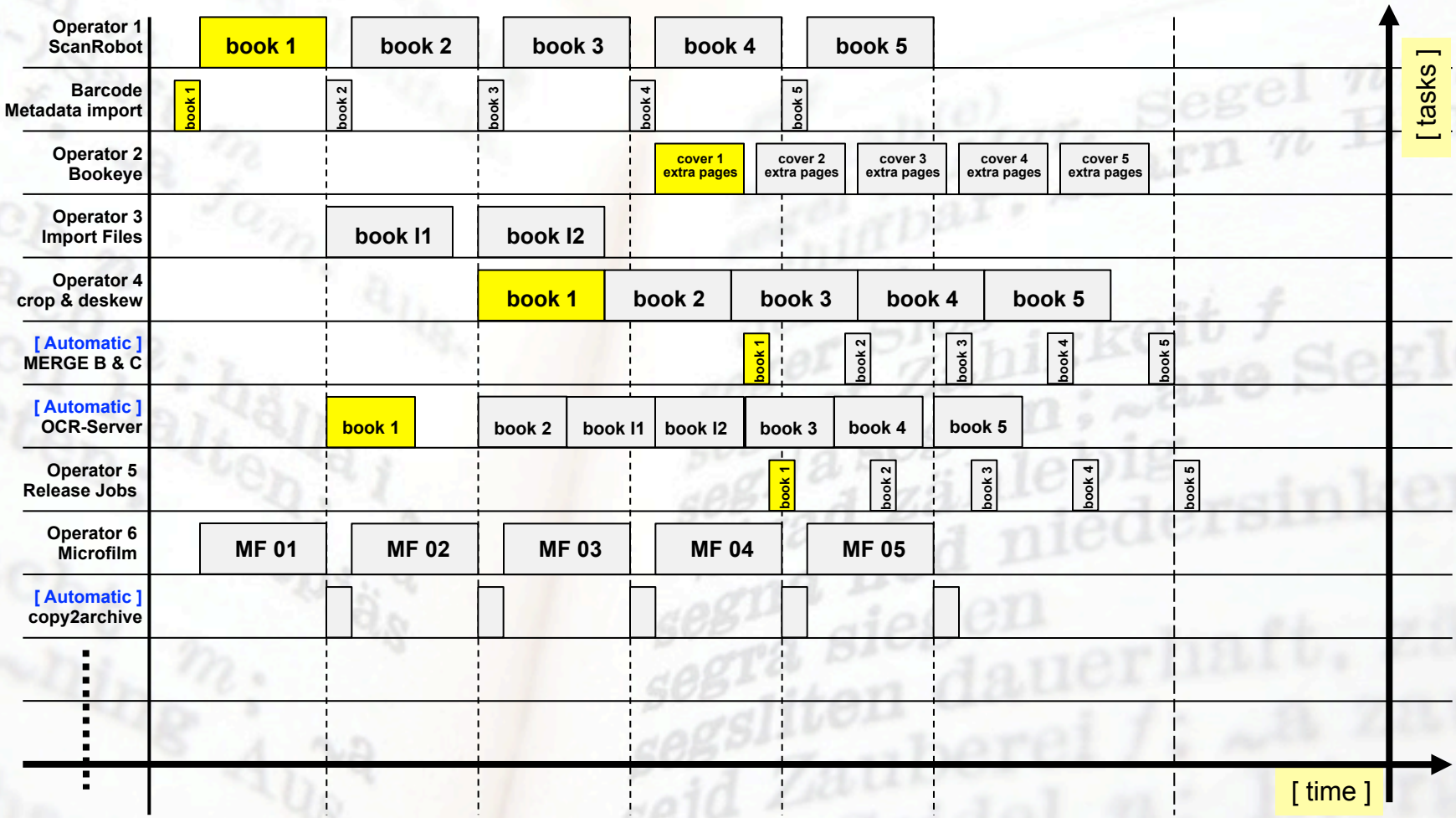


ScanFlow™

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

TREVENTUS

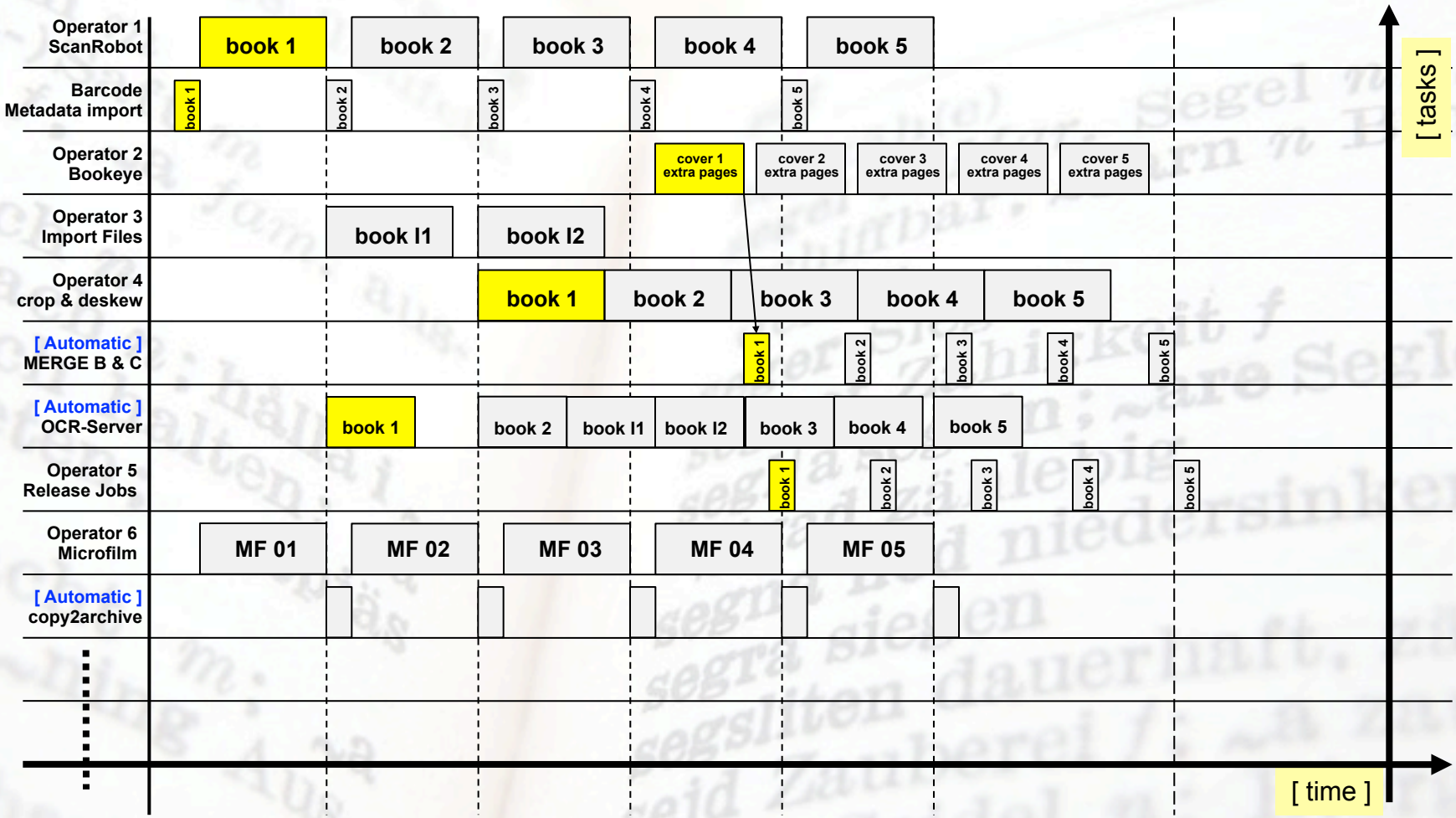
ScanFlow™: Jobs in the timetable



ScanFlow™

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

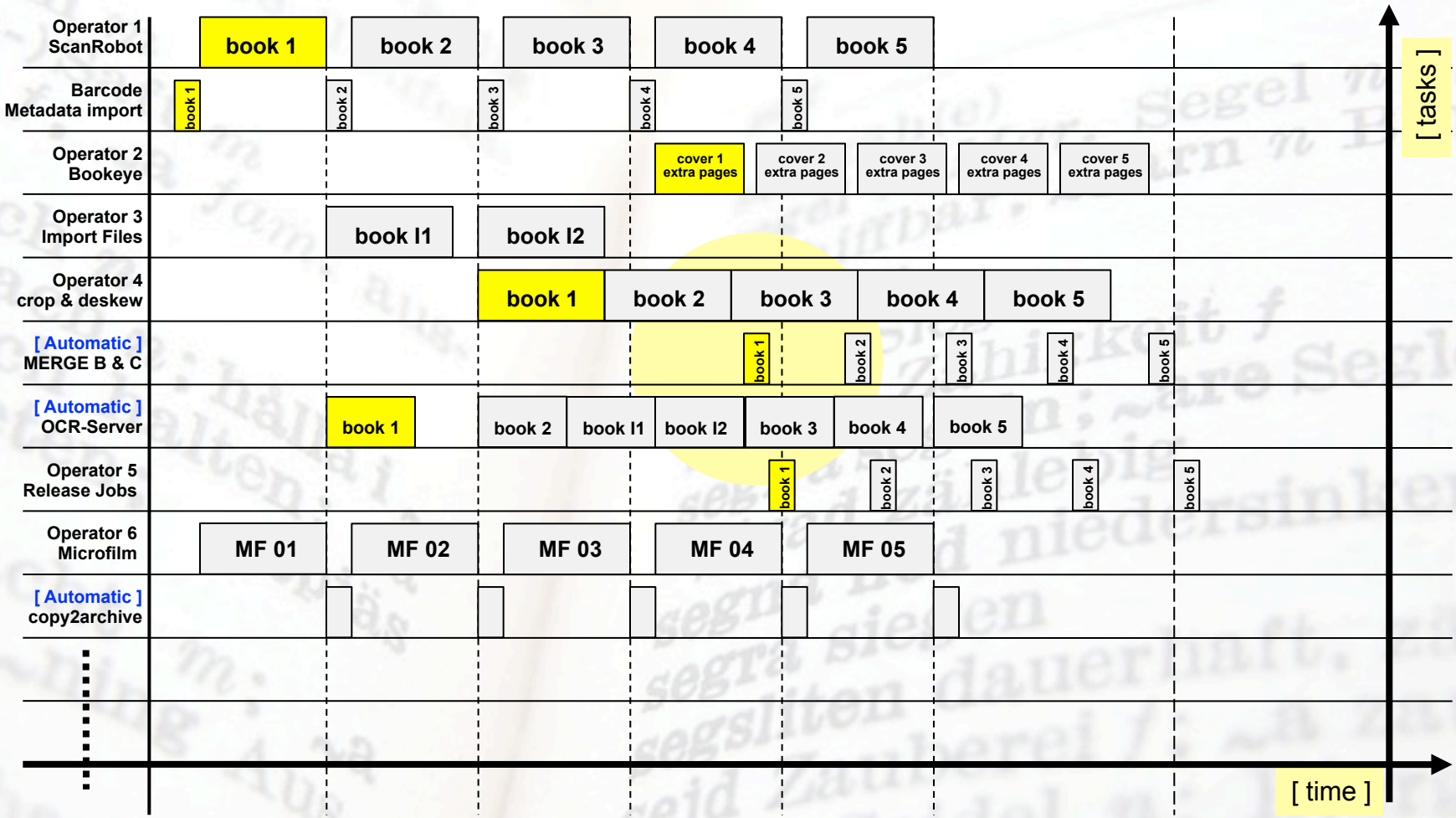
ScanFlow™: Jobs in the timetable



ScanFlow™

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

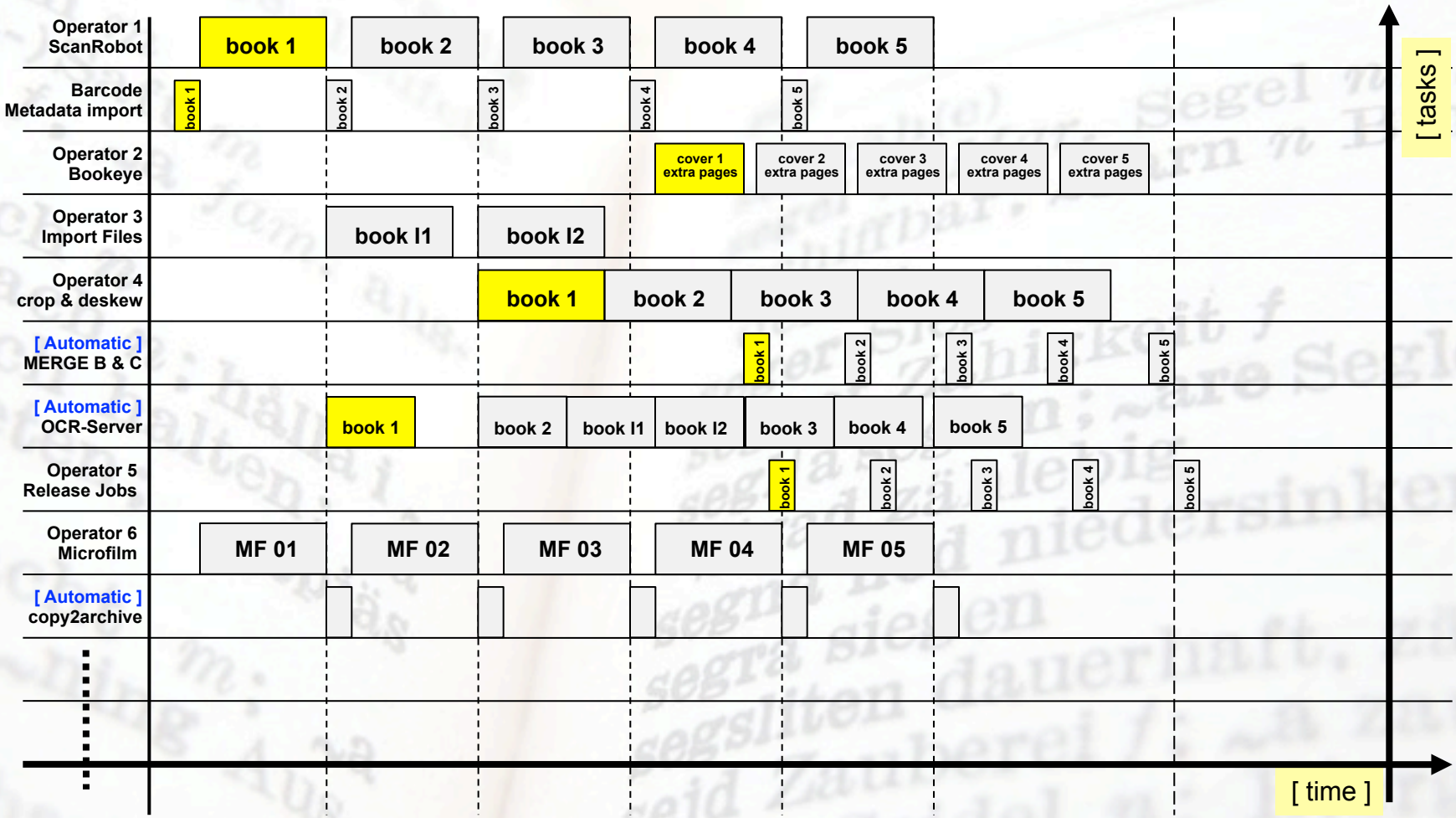
ScanFlow™: Jobs in the timetable



ScanFlow™

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

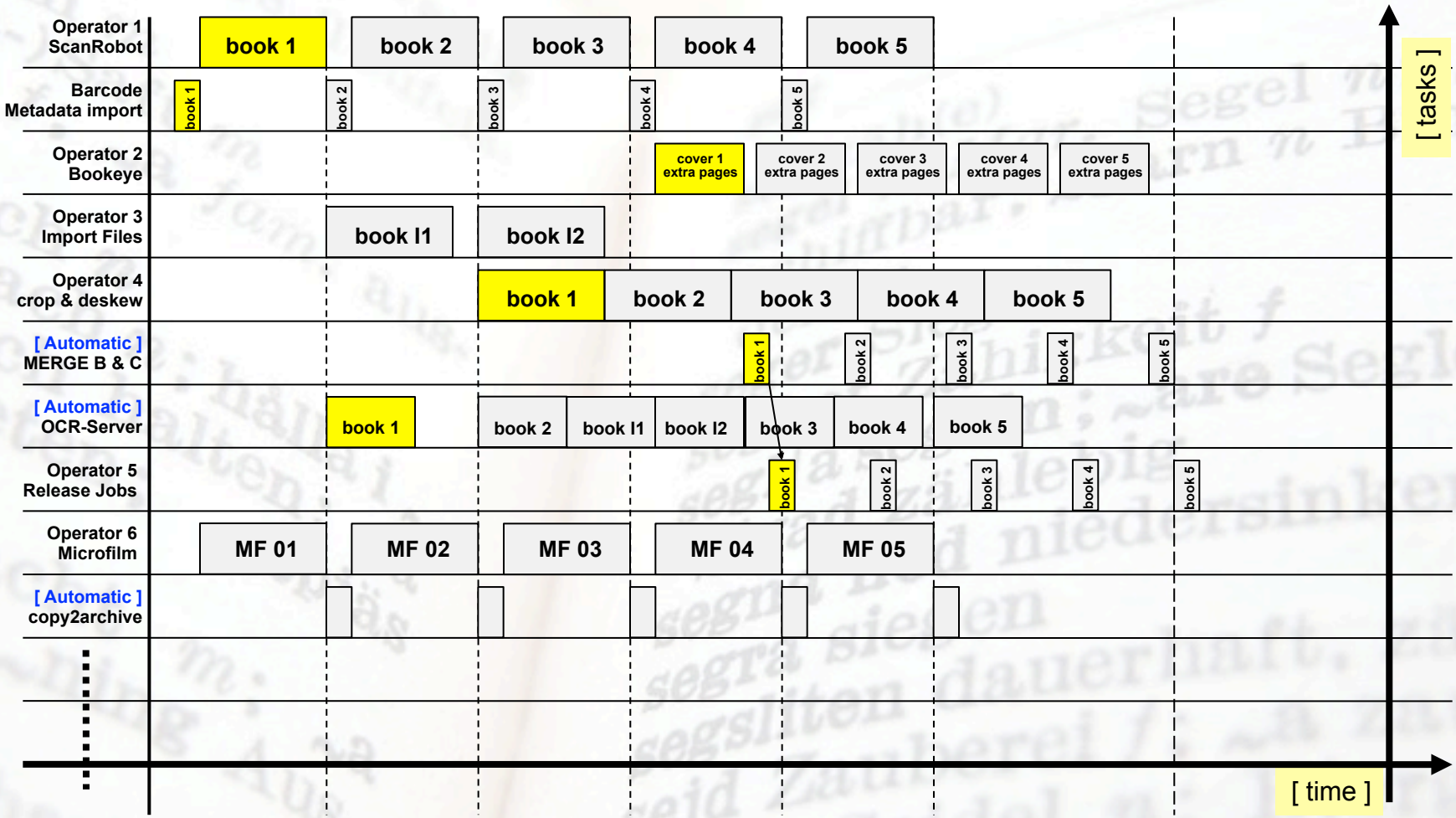
ScanFlow™: Jobs in the timetable



ScanFlow™

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

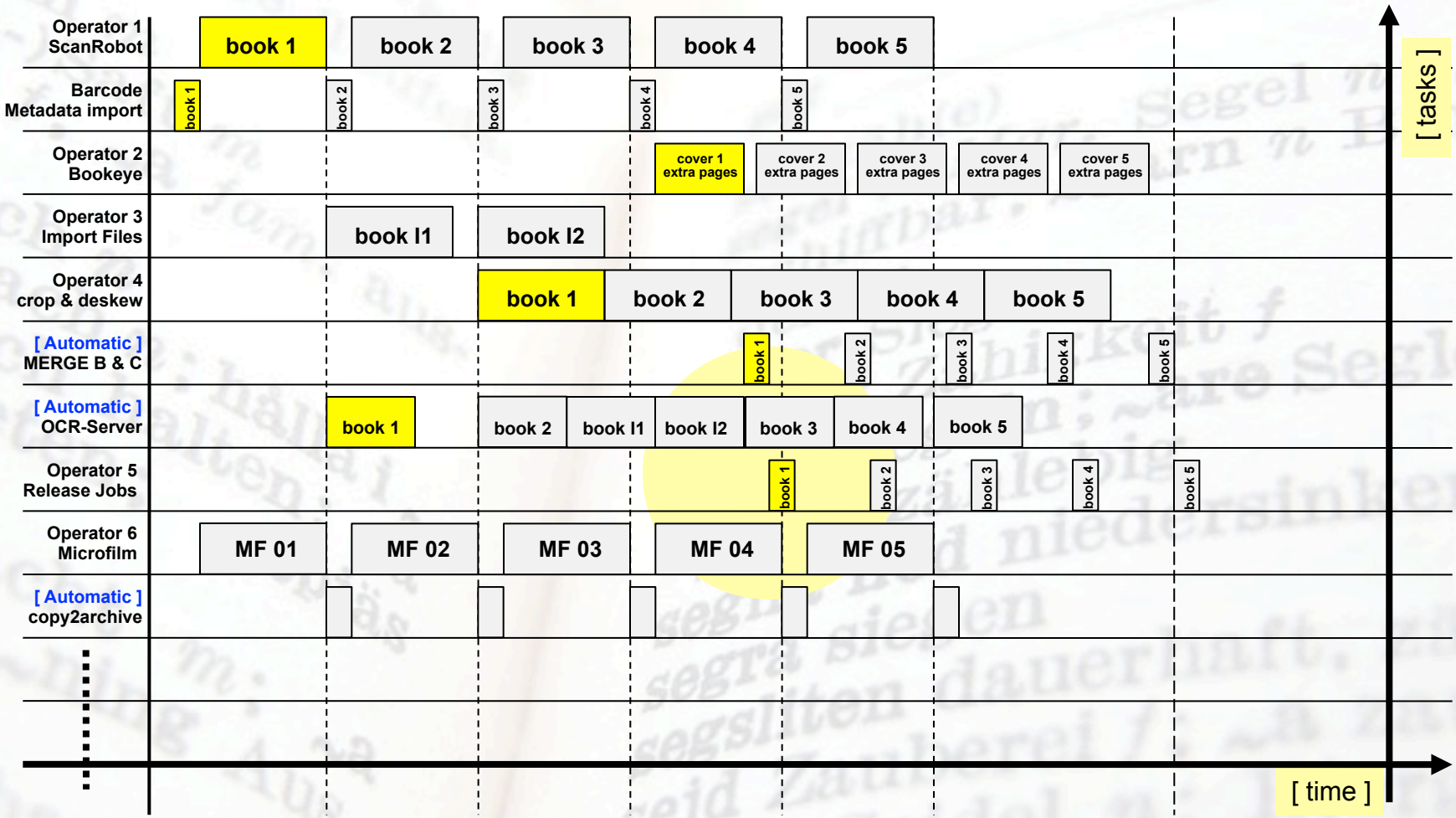
ScanFlow™: Jobs in the timetable



ScanFlow™

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

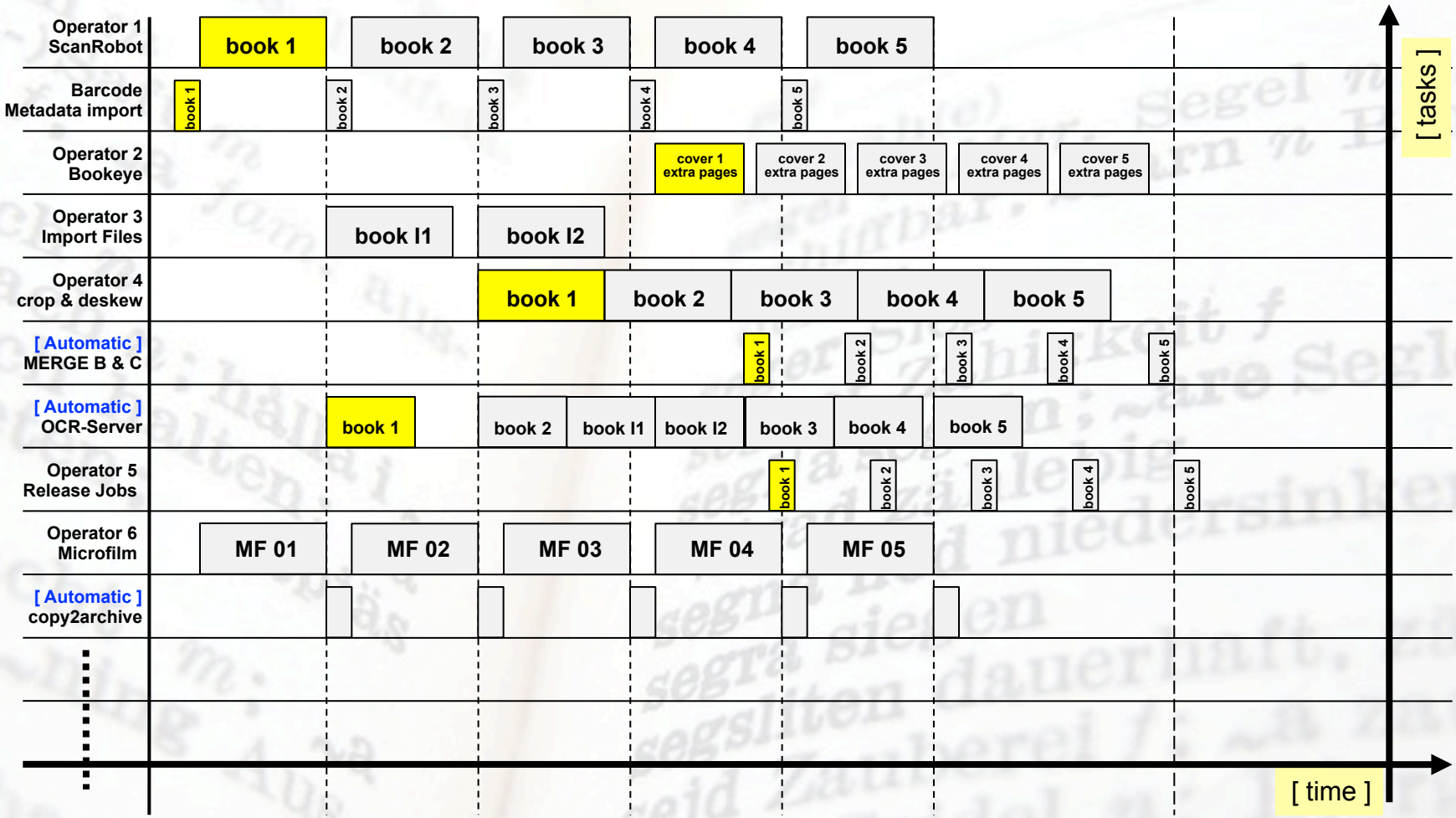
ScanFlow™: Jobs in the timetable



ScanFlow™

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

ScanFlow™: Jobs in the timetable



ScanFlow™

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

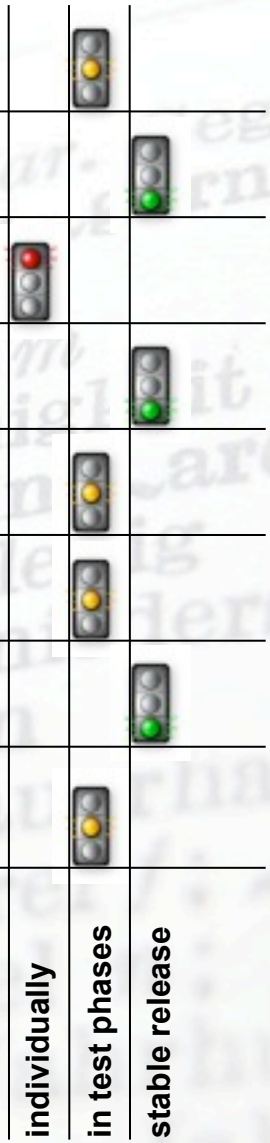
ScanFlow™: Examples

- renaming example
  - impossible to automize all the tasks
  - good preparation for manual tasks
- deskew and croping issue
  - prevent wrong results,
    by “live preview” of what will be after the i.e. croping
  - to avoid information lost

TREVENTUS

ScanFlow™: Modules

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



TREVENTUS

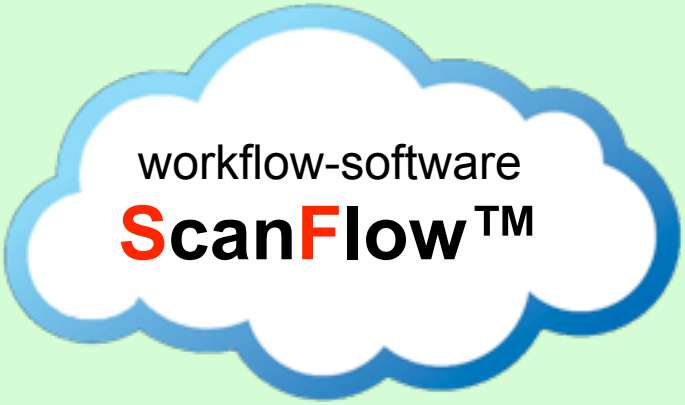
ScanFlow™: in test phases & development

- 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

TREVENTUS

Conclusio: What means “workflow”?

The workflow-software:



- organizing & coordinating
- monitoring & overview

→ to handle the resources of a scan-center

TREVENTUS

www.treventus.com