



SLUB

Wir führen Wissen.

Workflow Automation with Rosetta – Experience and Challenges

Henrike Berthold and Gerald Hübsch

Saxon State and University Library Dresden (SLUB)

June 2017

Workflow Automation with Rosetta

Agenda

Motivation

Workflows at SLUB

Workflow Automation

Quality Assurance

Challenges

Motivation

Mass digitization and long-term preservation services

Participation in mass-digitization projects

Long-term preservation service for other institutions

State Digitization Program for Science and Culture coordinated by the SLUB:

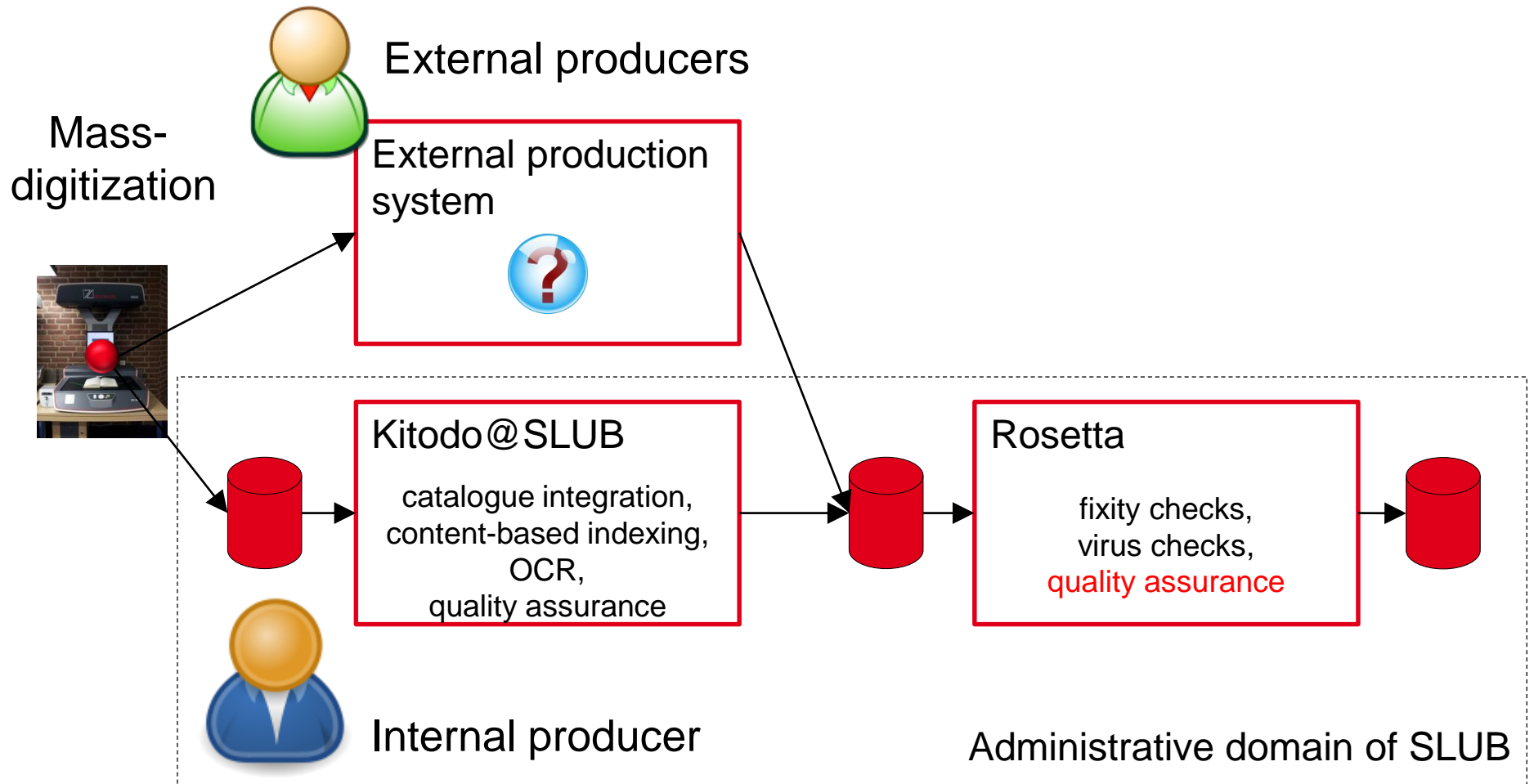
1. Licensing of digital media (consortium of Saxon university libraries)
2. Digitization of valuable materials and cultural artifacts in Saxony
- 3. Digital long-term archiving at SLUBArchiv**
 - 1.4 Mio. Scans in 2016, continues at least until end of 2018

Long-term preservation services for two other institutions

- 0.5 Mio. Scans per institution in 2017

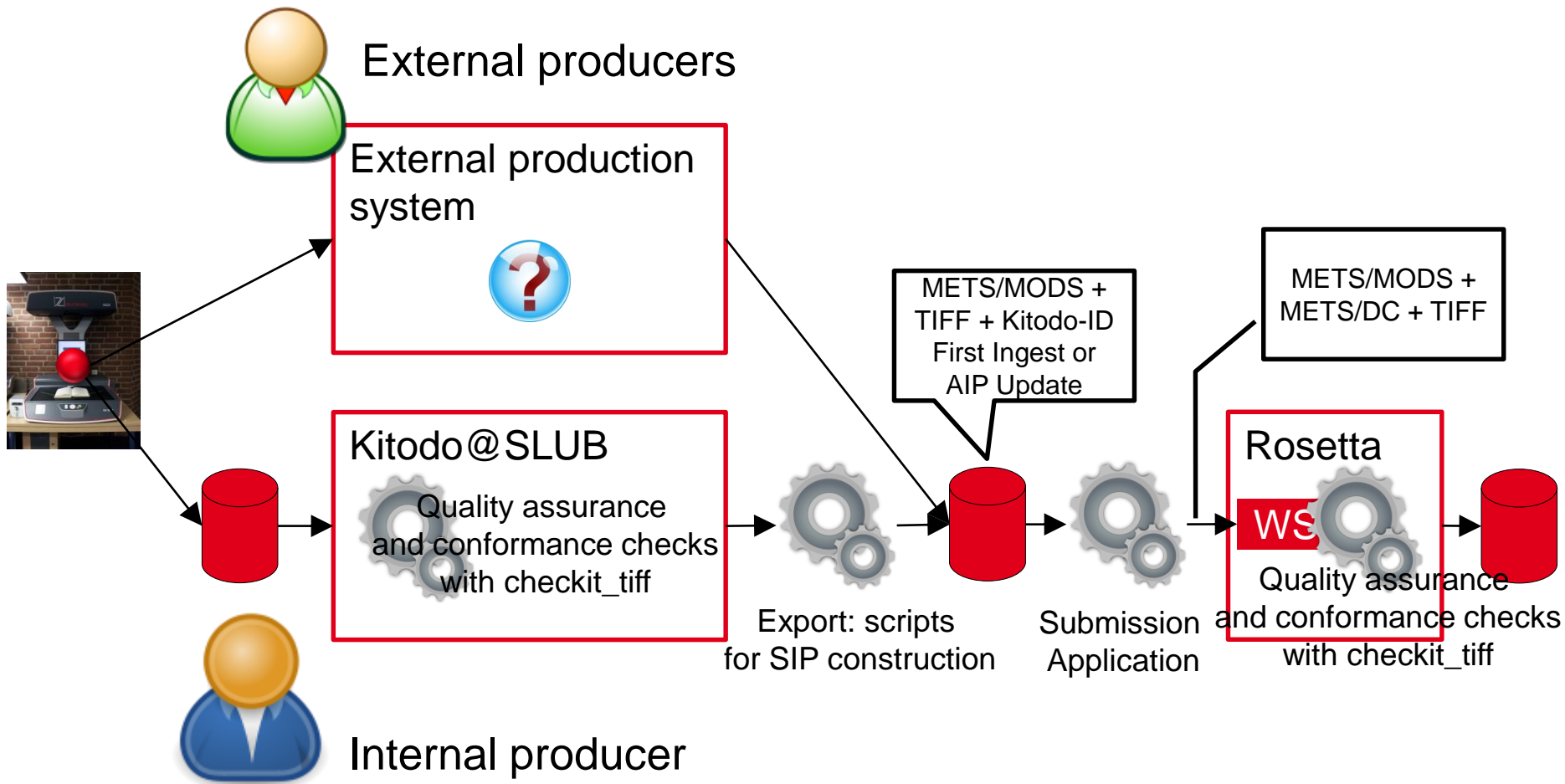
Rosetta Integration in Mass-digitization Scenarios

Simplified Workflows at SLUB



Rosetta Integration in Mass-digitization Scenarios

Workflow Automation



Rosetta Integration in Mass-digitization Scenarios

Quality Assurance with checkit_tiff

checkit_tiff [1] is *the* conformance checker for TIFF files used by all SLUB TIFF workflows

Integration in Kitodo is insufficient when serving external producers

Integration of checkit_tiff in Rosetta is mandatory, but *very* difficult to implement as a custom extension:

- Different Plugin types for format identification, metadata extraction and format validation are missing
- SLUB uses Metadata Extractor Plugin to integrate checkit_tiff
- Missing documentation of DNX significant properties
- Only format-specific, but not tenant-specific, configuration of conformance rule (e.g. “allow ICC type ‘Lino’ only for one tenant” can’t be implemented)

[1] https://github.com/SLUB-digitalpreservation/checkit_tiff

Rosetta Integration in Mass-digitization Scenarios

Quality Assurance

AIP Updates are necessary in ~ 8% of all ingests for corrections

Implemented by SLUB's submission application

An external ID (e.g. Kitodo process number) is used to find all SIPs with the same ID that are already in Rosetta (e.g. in the status 'PERMANENT' or 'DECLINED')

Work on the technical specification of a operation for Rosetta's WS API is in progress and will hopefully be available in Rosetta 5.3



Rosetta Integration in Mass-digitization Scenarios

Challenges ahead for SLUB that require ExLibris support

Improve trust in the digital archive!

- Cyber-threads: Integrity and authenticity of all Rosetta update-package must be guaranteed
- Dead reference identification as soon as performance issues are resolved

Operate Rosetta stable and reliable

- Monitoring options for central components
- Solr replication

Perform custom format validation with dedicated format validation Plugin type

- mkv/ffv1 (mediaconch), tiff (checkit_tiff), PDF (PDF Pilot SDK)

Simplify bulk data handling for TA

- Example: 3036 repaired SIPs in TA (~ 2.5TB) must be manually reingested



SLUB

Wir führen Wissen.

Workflow Automation with Rosetta – Experience and Challenges

Henrike Berthold and Gerald Hübsch

Saxon State and University Library Dresden (SLUB)

June 2017