ETH zürich

ETH Library

Docuteam packer

Viewer and editor for file structures and metadata

Other Conference Item

Author(s):
Töwe, Matthias; Bärlocher, Martin

Publication date:
2014

Permanent link:
https://doi.org/10.3929/ethz-a-010347496

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

ETH BIBLIOTHEK

OPEN RESEARCH DATA, EPFL, 28 October 2014, M. Töwe, M. Bärlocher

docuteam packer: viewer and editor for file structures and metadata

Overview

- **Aim of the Workshop**
- ***docuteam packer***
  - Purposes and limitations
  - Use cases
  - Configuration
- **Demo**
- **Testing**

  (Link to download: http://download.library.ethz.ch/docuteam-packer/)

Download

- http://download.library.ethz.ch/docuteam-packer/

Index of /docuteam-packer

Name	Last modified	Size	Description
Parent Directory		-	
Manual_docuteam_packer_Workshop.pdf	24-Oct-2014 11:54	1.5M	Manual in German
Manual_docuteam_packer_en.pdf	27-Oct-2014 15:40	1.3M	Manual in English
ooo.zip	10-Dec-2013 14:49	333M	Optional Open Office, for file preview only
packer_portable.zip	24-Oct-2014 14:15	38M	Application

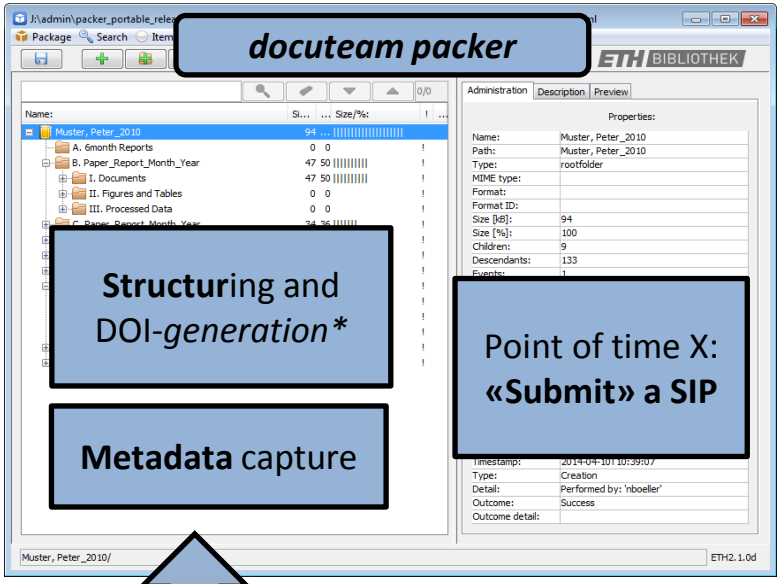
Aims of the workshop

- **Aims of using *docuteam packer*** are clear
- Participants have a rough understanding of the tool's **strengths and limitations**
- **Technical prerequisites** for using the tool are known

Workflow «Small Data»

Selection and documentation of context | **Curation and usability** | **Bitstream preservation**

Researchers | **Library** | **IT-Services**



Filesystem (networked, local)
Selection for archiving

Sub-mission

docuteam feeder

ETH Data Archive

Server

Storage

Network

Delivery via DOI

Access via Knowledge Portal

What is *docuteam packer*?

For users

- **Viewer and editor for local preparation of archival packages for transfer to ETH Data Archive**
  - Create and edit folder structure, as it should be reflected in ETH Data Archive
  - Enter and edit metadata
  - DOI-creation (Digital Object Identifier; to be registered by ETH Data Archive)
  - Assign access rights and retention periods to be enforced by ETH Data Archive

In the background

- Create a **Submission Information Package (SIP)** or Archival Information Package (AIP) **of metadata + structure (METS-format,** Metadata Encoding and Transmission Standard) **and data**

What *docuteam packer* is not!

- **No comprehensive data management solution**
- **No records management solution**
- **No collaboration platform**
- **No data repository**
- **No long term archive** - but a tool to **prepare** for and submit to archive
- **No solution for *local* rights management**
- **Not tied to use with Rosetta** as the only long term archive

➢ ***Consider alternative approaches*** *where these are more appropriate*

➢ ***Be careful with using the tool without submitting to an archive***

Example Use Cases

Research groups

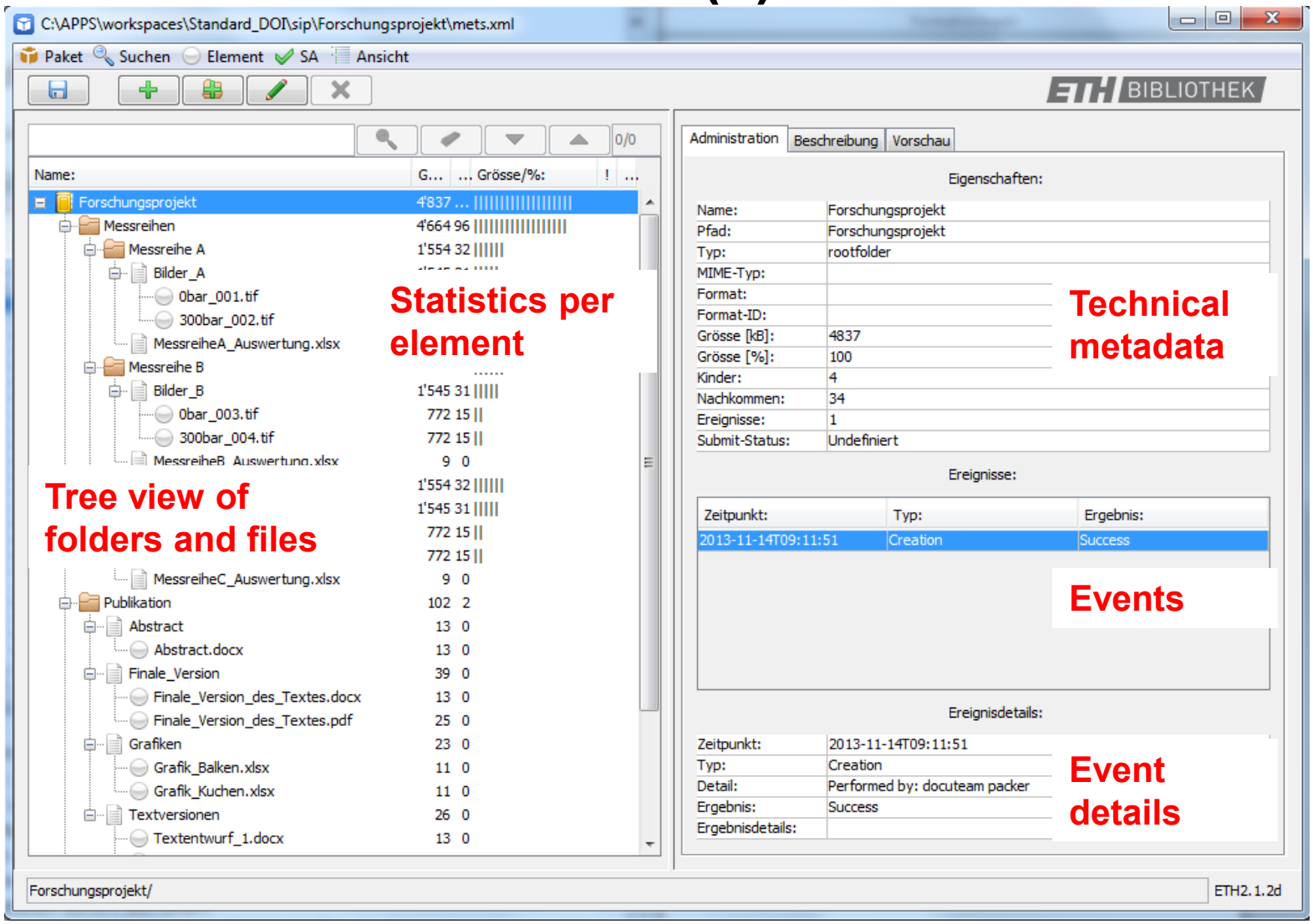
- **Data belonging to a manuscript** are collected, submitted to the long term archive and made accessible via DOI for reviewers and readers
- **Research group has a structured filing** without metadata; it should be edited and submitted into the long term archive
- **PhD students** of a group **are presented with a filing structure** they should follow when managing their data

…

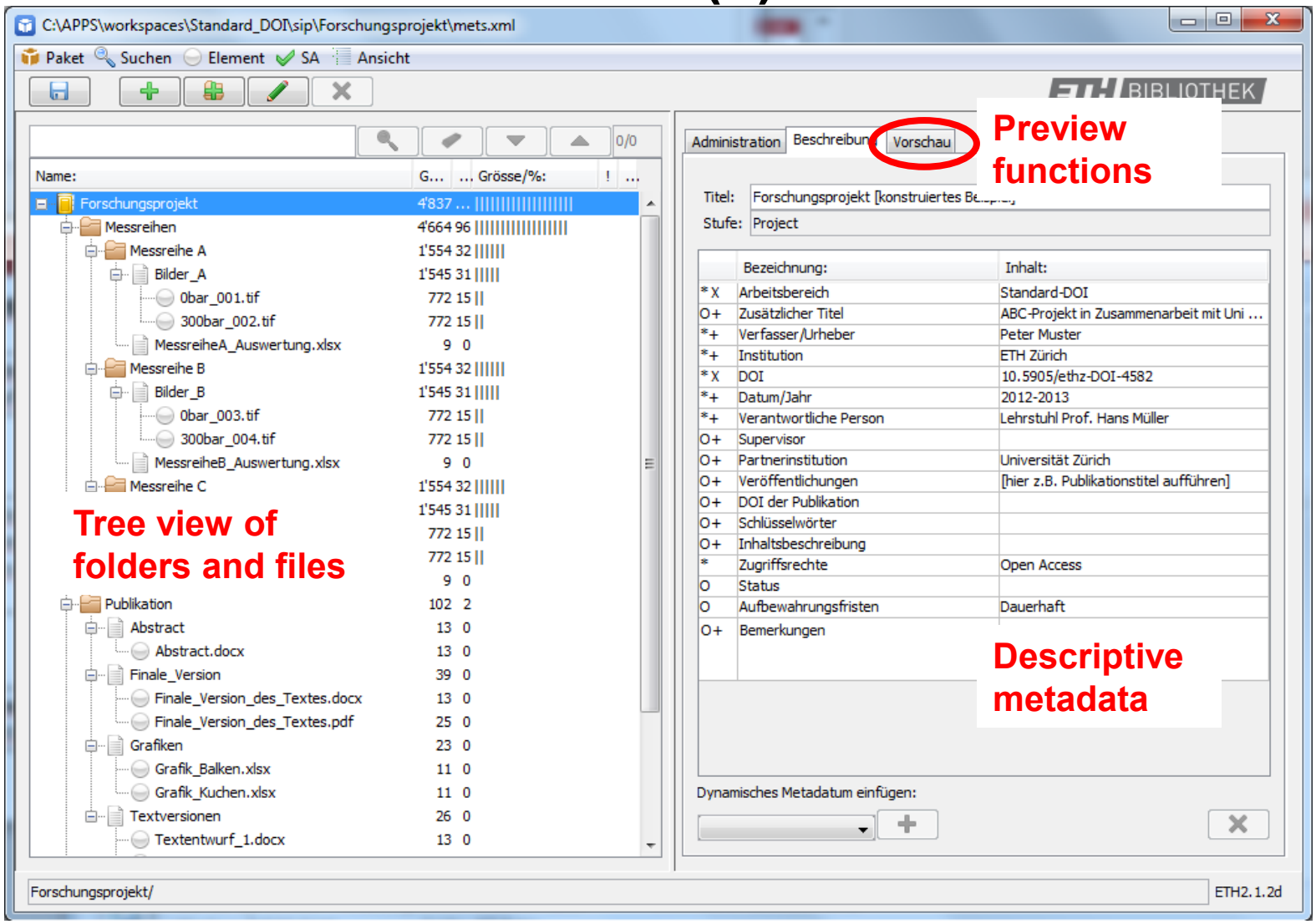
Administrative staff within ETH

- **Delivers structured data to ETH Zurichs university archives**…
- …**archives' staff appraises and selects** content and adds metadata

The GUI and its elements (1)



The GUI and its elements (2)

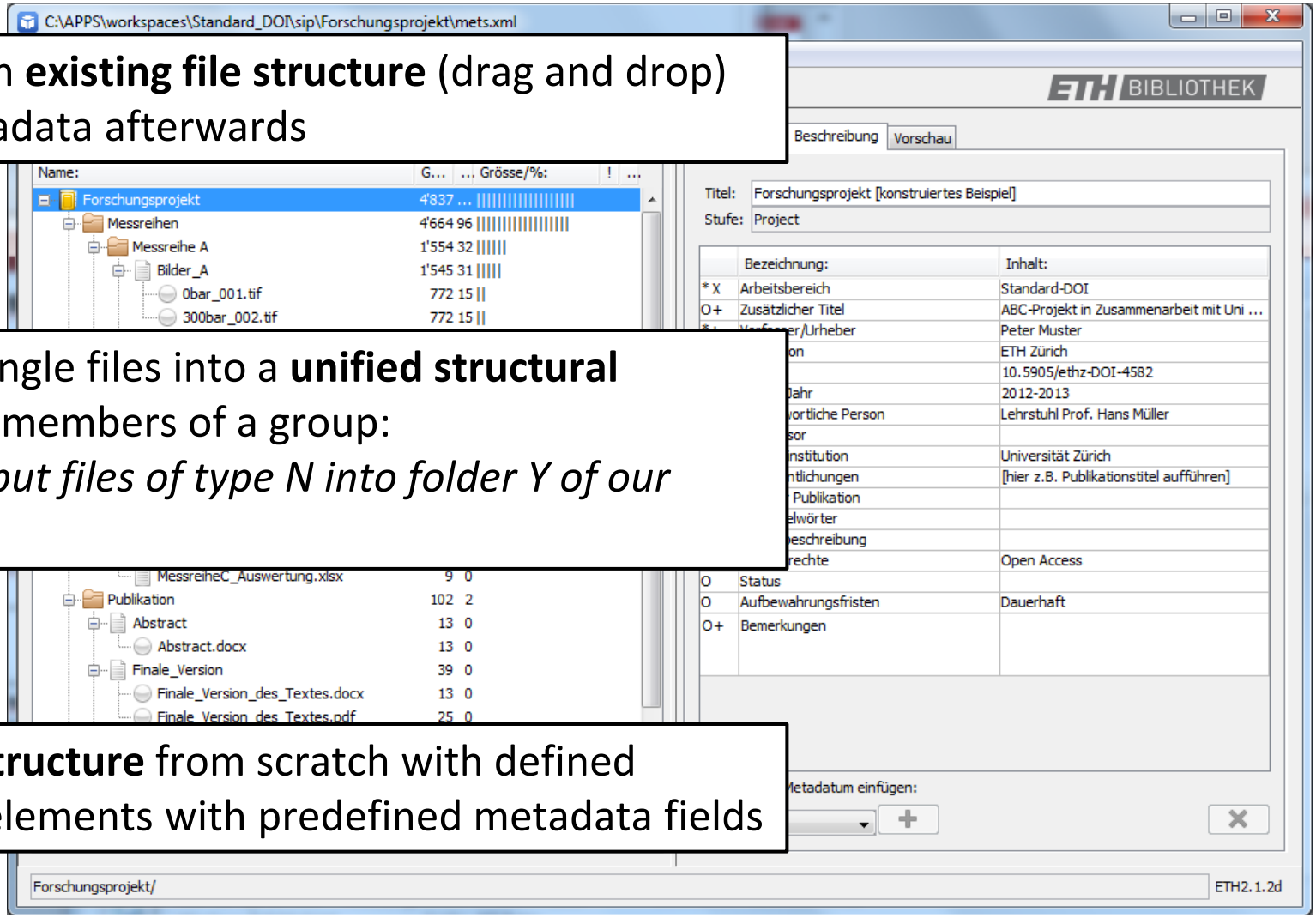


Practical use

1. Import an **existing file structure** (drag and drop) and add metadata afterwards

2. Import single files into a **unified structural template** for members of a group:
«We always put files of type N into folder Y of our structure»

3. **Build a structure** from scratch with defined hierarchical elements with predefined metadata fields



Why the effort?

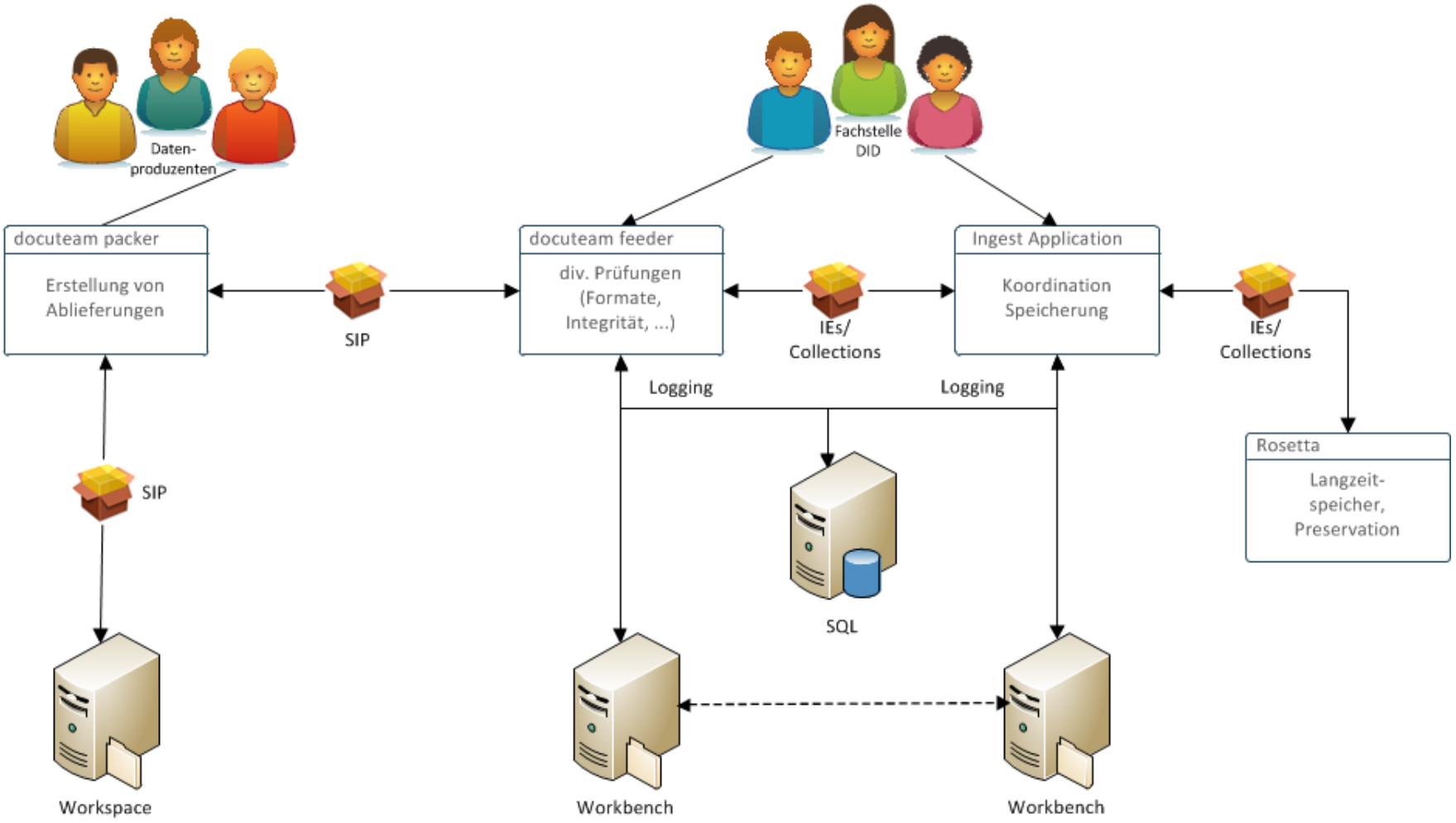
- «Local data management light»:
  Data are **structured and described** locally within the group
- **Group retains full control, but important work is already done** to facilitate long term preservation
- **Metadata can be configured** – within reasonable limits
- **Structure and metadata in METS-XML** can be submitted to ETH Data Archive automatically (via *docuteam feeder* as Submission Application)
- DOI are generated and can be used in citations (registration follows later in ETH Data Archive)
- **Selection of retention periods and access rights** to be enforced in ETH Data Archive

Issues to observe

- **Configuration** is flexible, but must remain consistent with MD schema
- **If flexibility** of configuration **is exploited**, effort for maintenance strongly increases and **the approach will not scale well**
- **Early discussion with the Digital Curation Office** is important!
- No installation, but it must be possible to **run *docuteam packer* locally**
- **Users can get themselves into trouble** by manipulations on the file system
- **If data remains on local storage for years**, problems with respect to long term preservation can occur once data is submitted to long term archive

Complete process into the archive

Gesamtprozess



Questions?

www.library.ethz.ch/Digital-Curation

data-archive@library.ethz.ch

Dr. Matthias Töwe
Head Digital Curation
ETH-Bibliothek
Rämistrasse 101
8092 Zurich
044 632 60 32
matthias.toewe@library.ethz.ch

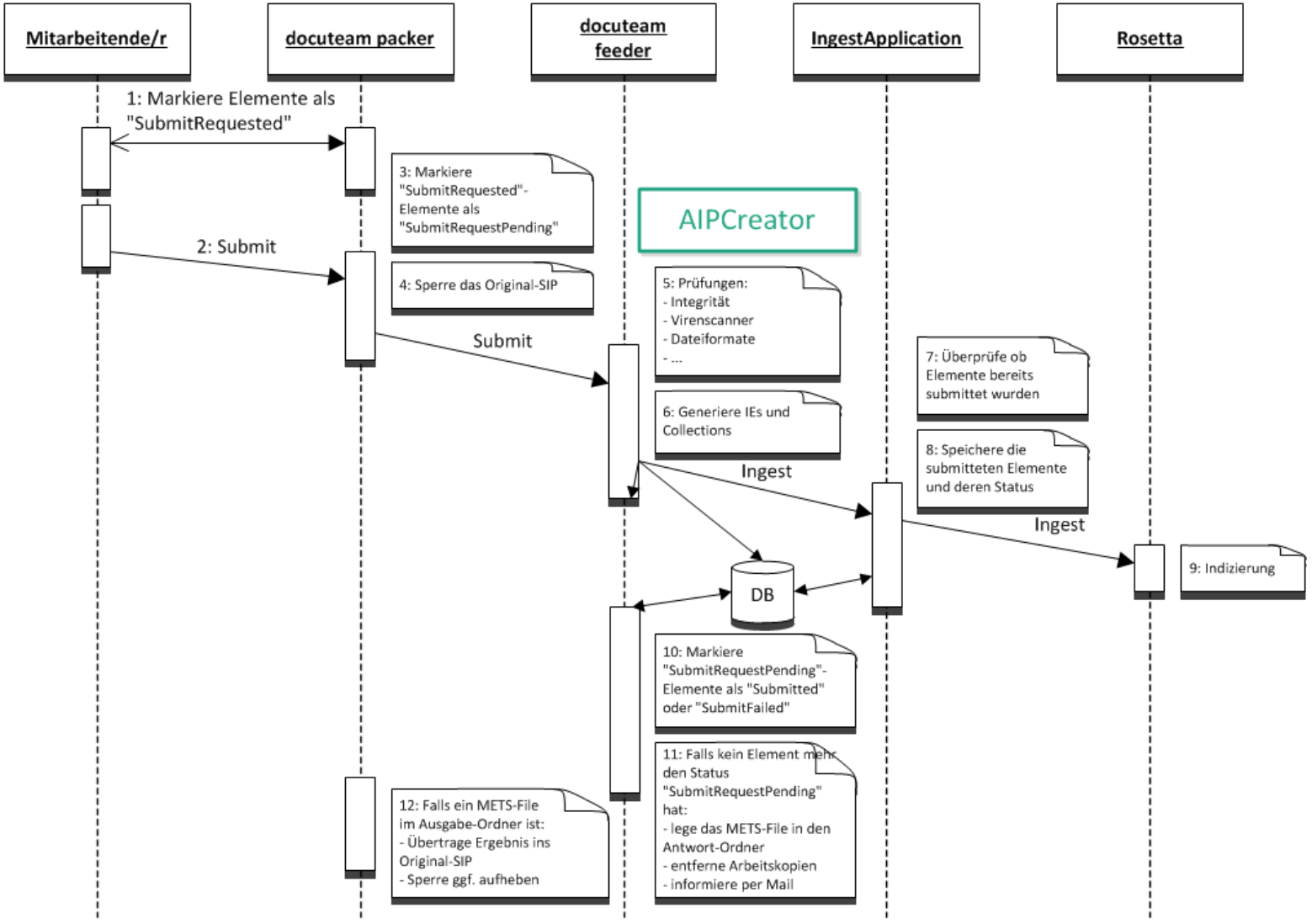
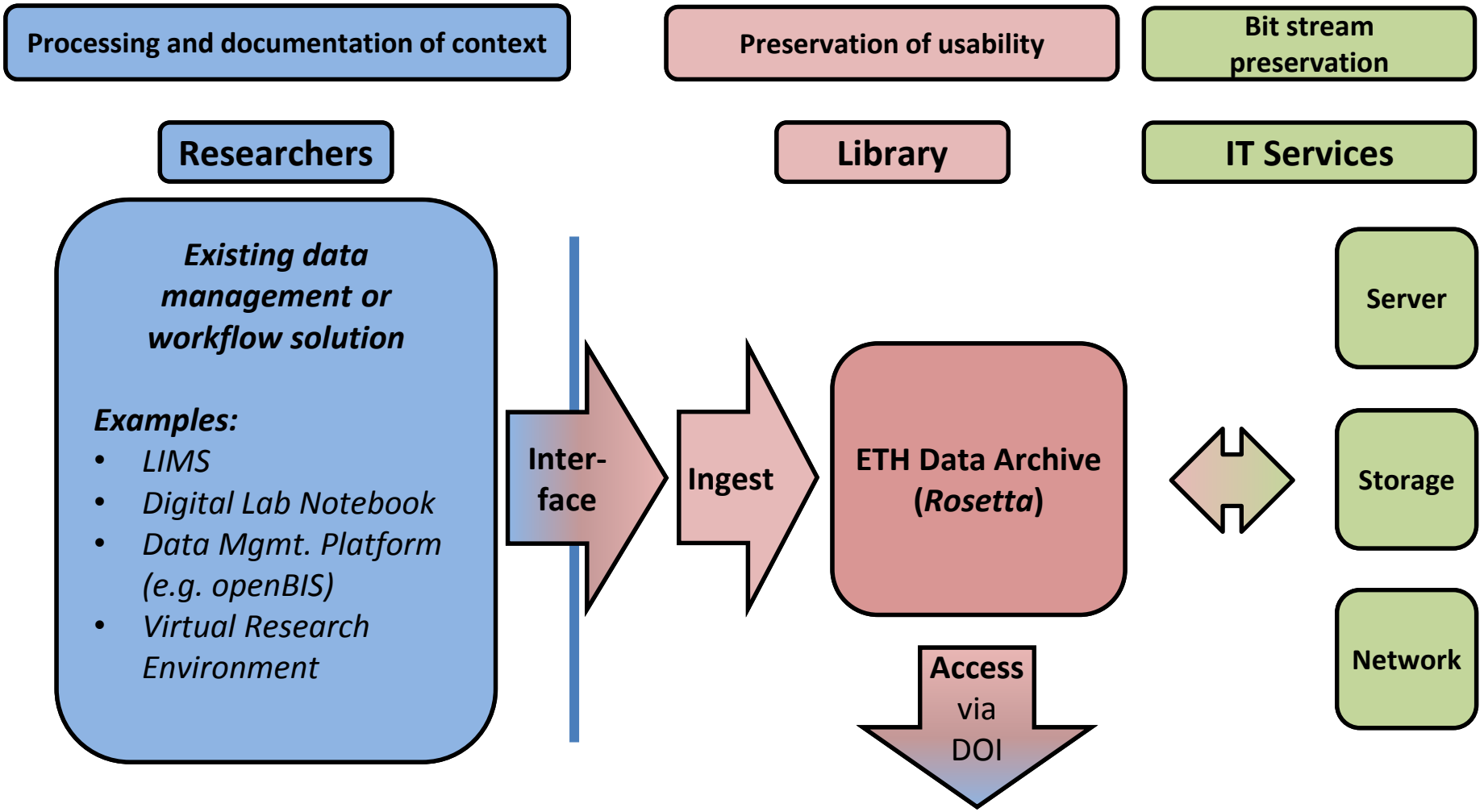
Martin Bärlocher
Library IT Services
ETH-Bibliothek
Rämistrasse 101
8092 Zurich
044 632 04 05
martin.baerlocher@library.ethz.ch

Use Case Research Data – «Small Data»

- **Distinct from «Big Data»**
- **Structured data in discrete files**; produced everywhere – even in projects which actually deal with «Big Data» as their research topic
- **Interface between data management ←→ Long term preservation**
- **Facilitate** compliance with accountability and **verifiability**
- **Ensure citability** of data → DOI-registration
- Support producer's own **re-use**, access by colleagues or *Open Data*
  → From ***Restricted Access* to *Open Access***
- **Retention period at least 10 years**

➢ *We expect increasing **requirements** by funders and universities **for data management***

Submission Process



(Close to) ideal approach for Research Data