

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

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/>

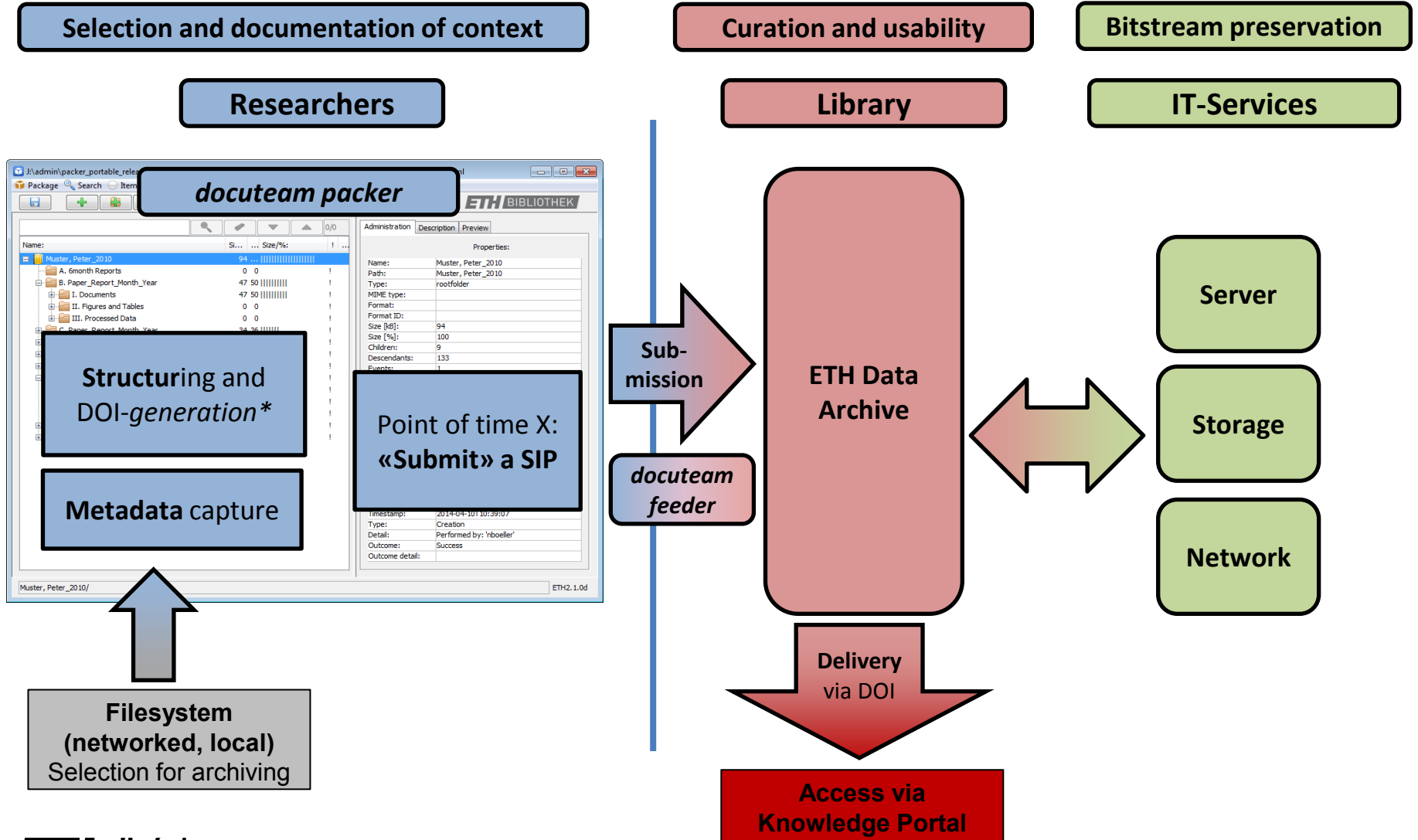
Index of /docuteam-packer

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 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»



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 Zurich's university archives...**
- **...archives' staff appraises and selects content and adds metadata**

The GUI and its elements (1)

The screenshot displays the ETH BIBLIOTHEK interface. On the left, a tree view shows a folder structure for 'Forschungsprojekt'. The right pane shows metadata for the selected 'Forschungsprojekt' folder, including properties and events.

Tree view of folders and files

Name	Größe	Größe/%
Forschungsprojekt	4'837
Messreihen	4'664	96
Messreihe A	1'554	32
Bilder_A	
0bar_001.tif		
300bar_002.tif		
MessreiheA_Auswertung.xlsx		
Messreihe B	
Bilder_B	1'545	31
0bar_003.tif	772	15
300bar_004.tif	772	15
MessreiheB_Auswertung.xlsx	9	0
MessreiheC_Auswertung.xlsx	1'554	32
Publikation	102	2
Abstract	13	0
Abstract.docx	13	0
Finale_Version	39	0
Finale_Version_des_Textes.docx	13	0
Finale_Version_des_Textes.pdf	25	0
Grafiken	23	0
Grafik_Balken.xlsx	11	0
Grafik_Kuchen.xlsx	11	0
Textversionen	26	0
Textentwurf_1.docx	13	0

Statistics per element

Technical metadata

Eigenschaften:	
Name:	Forschungsprojekt
Pfad:	Forschungsprojekt
Typ:	rootfolder
MIME-Typ:	
Format:	
Format-ID:	
Größe [kB]:	4837
Größe [%]:	100
Kinder:	4
Nachkommen:	34
Ereignisse:	1
Submit-Status:	Undefiniert

Events

Zeitpunkt:	Typ:	Ergebnis:
2013-11-14T09:11:51	Creation	Success

Event details

Zeitpunkt:	2013-11-14T09:11:51
Typ:	Creation
Detail:	Performed by: docuteam packer
Ergebnis:	Success
Ergebnisdetails:	

The GUI and its elements (2)

The screenshot shows the ETH Bibliothek metadata editor interface. On the left, a tree view displays the folder structure of a research project, including subfolders for 'Messreihen A', 'B', and 'C', and a 'Publikation' folder. On the right, the 'Vorschau' (Preview) tab is active, showing a metadata table with various fields and their values. The 'Vorschau' tab is circled in red and labeled 'Preview functions'. The metadata table is labeled 'Descriptive metadata'.

Tree view of folders and files

Name:	G...	Grösse/%:
Forschungsprojekt	4'837 ...	
Messreihen	4'664 96	
Messreihe A	1'554 32	
Bilder_A	1'545 31	
0bar_001.tif	772 15	
300bar_002.tif	772 15	
MessreiheA_Auswertung.xlsx	9 0	
Messreihe B	1'554 32	
Bilder_B	1'545 31	
0bar_003.tif	772 15	
300bar_004.tif	772 15	
MessreiheB_Auswertung.xlsx	9 0	
Messreihe C	1'554 32	
Publikation	102 2	
Abstract	13 0	
Abstract.docx	13 0	
Finale_Version	39 0	
Finale_Version_des_Textes.docx	13 0	
Finale_Version_des_Textes.pdf	25 0	
Grafiken	23 0	
Grafik_Balken.xlsx	11 0	
Grafik_Kuchen.xlsx	11 0	
Textversionen	26 0	
Textentwurf_1.docx	13 0	

Descriptive metadata

Bezeichnung:	Inhalt:
* X Arbeitsbereich	Standard-DOI
O+ Zusätzlicher Titel	ABC-Projekt in Zusammenarbeit mit Uni ...
*+ Verfasser/Urheber	Peter Muster
*+ Institution	ETH Zürich
* X DOI	10.5905/ethz-DOI-4582
*+ Datum/Jahr	2012-2013
*+ Verantwortliche Person	Lehrstuhl Prof. Hans Müller
O+ Supervisor	
O+ Partnerinstitution	Universität Zürich
O+ Veröffentlichungen	[hier z.B. Publikationstitel aufführen]
O+ DOI der Publikation	
O+ Schlüsselwörter	
O+ Inhaltsbeschreibung	
* Zugriffsrechte	Open Access
O Status	
O Aufbewahrungsfristen	Dauerhaft
O+ Bemerkungen	

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

The screenshot shows the ETH BIBLIOTHEK software interface. On the left, a file explorer displays a hierarchical structure for a 'Forschungsprojekt' (Research Project). The structure includes folders for 'Messreihen' (Measurement Series) and 'Publikation' (Publication), with various files like 'Obar_001.tif', '300bar_002.tif', 'Abstract.docx', and 'Finale_Version.pdf' listed with their sizes and dates.

On the right, a metadata form is visible, titled 'Beschreibung' (Description) and 'Vorschau' (Preview). The form contains fields for 'Titel' (Title), 'Stufe' (Level), and a table for 'Bezeichnung' (Designation) and 'Inhalt' (Content). The 'Bezeichnung' table has columns for 'Arbeitsbereich' (Working Area), 'Zusätzlicher Titel' (Additional Title), 'Verfasser/Urheber' (Author/Creator), 'Jahr' (Year), 'Verantwortliche Person' (Responsible Person), 'Institution', 'Publikation', 'Schlüsselwörter' (Keywords), 'Beschreibung', and 'Rechte' (Rights). The 'Inhalt' column contains values like 'Standard-DOI', 'ABC-Projekt in Zusammenarbeit mit Uni ...', 'Peter Muster', 'ETH Zürich', '10.5905/ethz-DOI-4582', '2012-2013', 'Lehrstuhl Prof. Hans Müller', 'Universität Zürich', and '[hier z.B. Publikationstitel aufführen]'. The 'Rechte' field is set to 'Open Access'.

At the bottom of the interface, there is a 'Metadatum einfügen:' (Insert Metadata) section with a dropdown menu and a '+' button.

Why the effort?

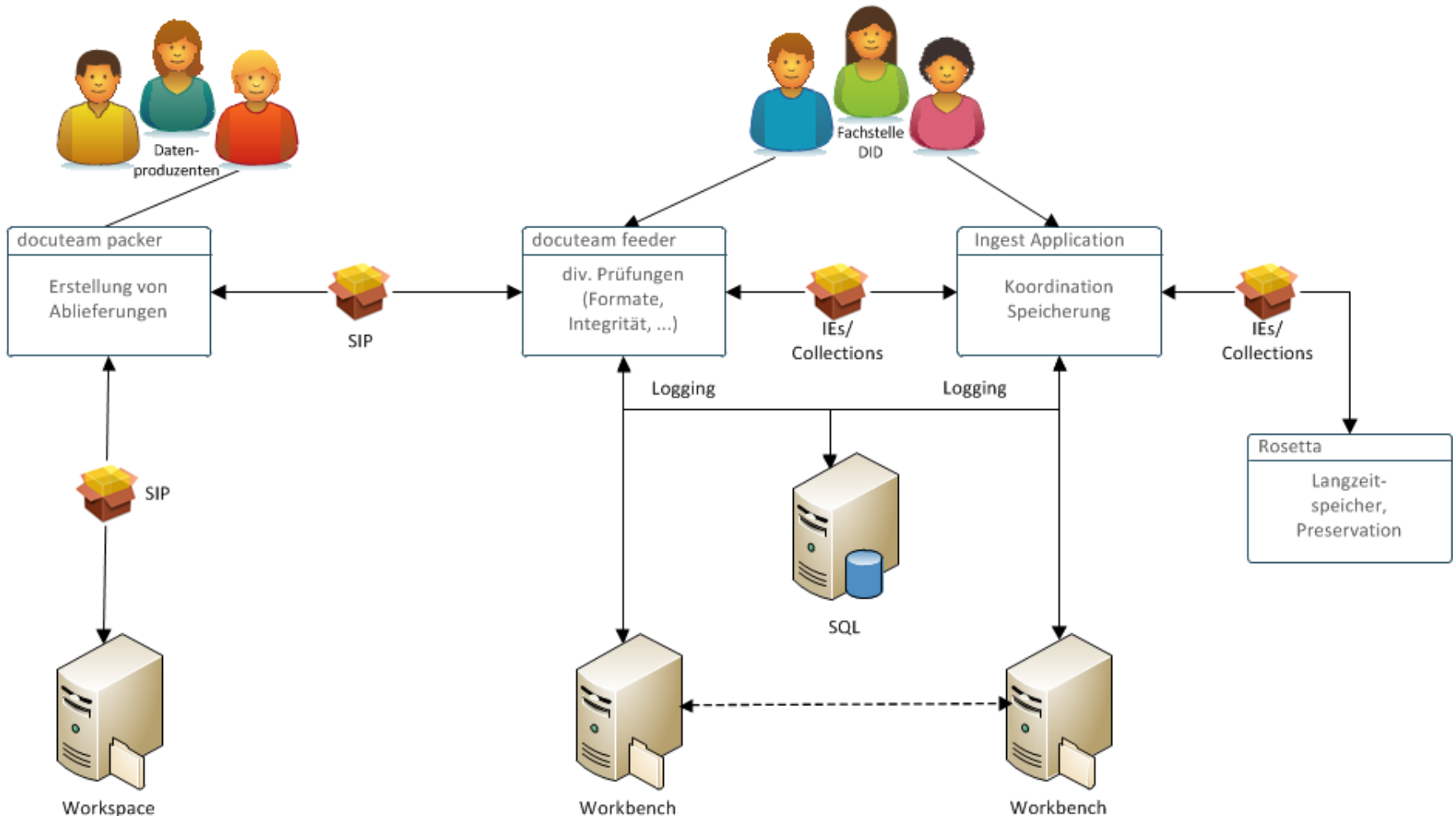
- «Local data management light»:
Data are **structured and described** locally within the group
- **Group retains full control, but important work is already done** to facilitate long term preservation
- **Metadata can be configured** – within reasonable limits
- **Structure and metadata in METS-XML** 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

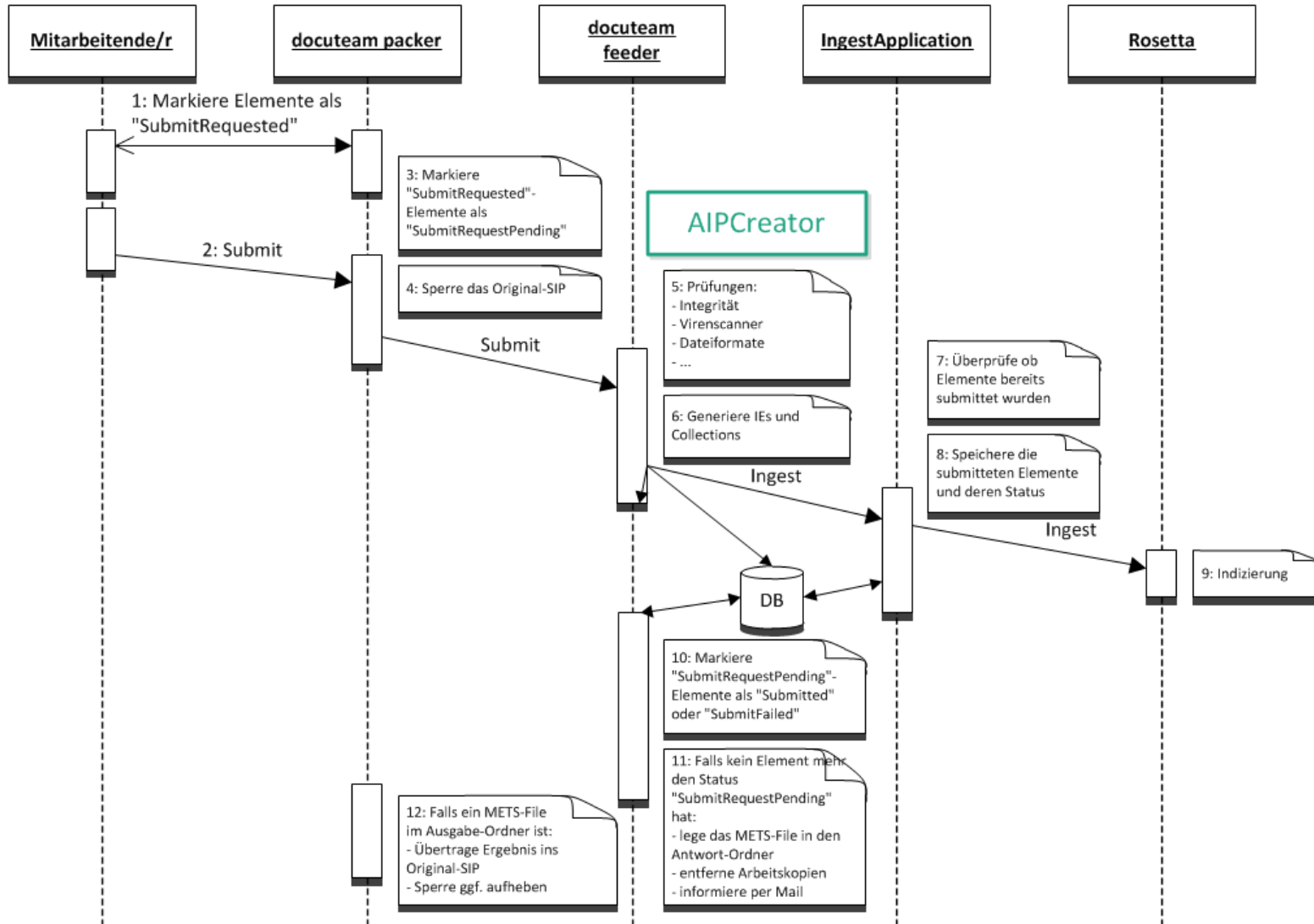
Martin Bärlocher
Library IT Services
ETH-Bibliothek
Rämistrasse 101
8092 Zurich
044 632 04 05

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

