



# Digital Curation at ETH-Bibliothek: From a Survey to a Customer Oriented Service for Researchers

LIBER Annual Conference 2014, July 3

Dr. Arlette Piguet, ETH-Bibliothek, ETH Zurich

# General Conditions at ETH Zurich

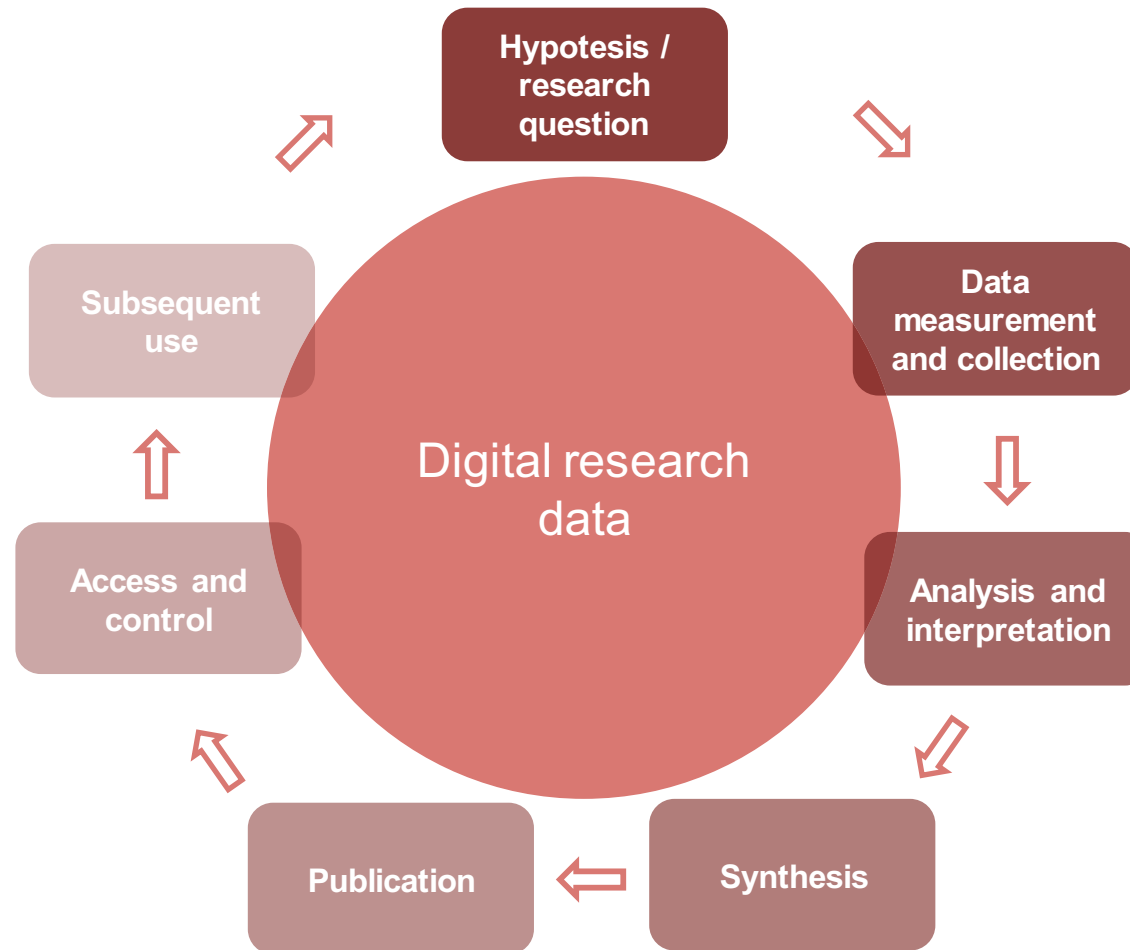
## ETH Zurich

- Swiss Federal Institute of Technology Zurich
- Founded in 1855
- University for technology and natural sciences
- More than 18'000 students from 80 countries, 2'600 among them are doctoral candidates, 500 professors
- Central library: ETH-Bibliothek

## Research data

- The responsibility for data archiving lies with the researcher (ETH Guidelines for Research Integrity and Good Scientific Practice)
- That means: archiving of data is formally clear; in reality archiving is desperate, unclear, confusing

# Idealized Life Cycle of Research Data



# Objectives of the Project «Data Curation» at ETH Zurich

## Provide services for the researchers

- Fulfill the requirements of the research funding organizations / good scientific practice
  - Long term archiving of research data in electronic format, make data accessible, guarantee reproducibility
- Guarantee citability
  - DOI registration
- Facilitate reuse
  - For own purposes, for colleagues or globally
- Support the required documentation and description
- No competition with subject specific solutions
- Complementary internal solutions for subjects without such data archives or for data which should or may not be put into these archives
- Consultancy and support for using external or internal services, with selection of formats suitable for long term archiving and with quality control
- Central archiving: minimize single solutions



**Longterm archiving of further electronic data at ETH Zurich:** Administrative records from ETH Archives and Library materials (born digital theses, digitization masters)

# Why Institutional / Local Approaches?

## Institutional approaches

- Do researchers need «external» support?
  - In reality «no». But infrastructure services can relieve them for the benefit of their core business
- Are subject specific repositories and services not enough?
  - Quite a broad range of services are not covered

## Local approaches

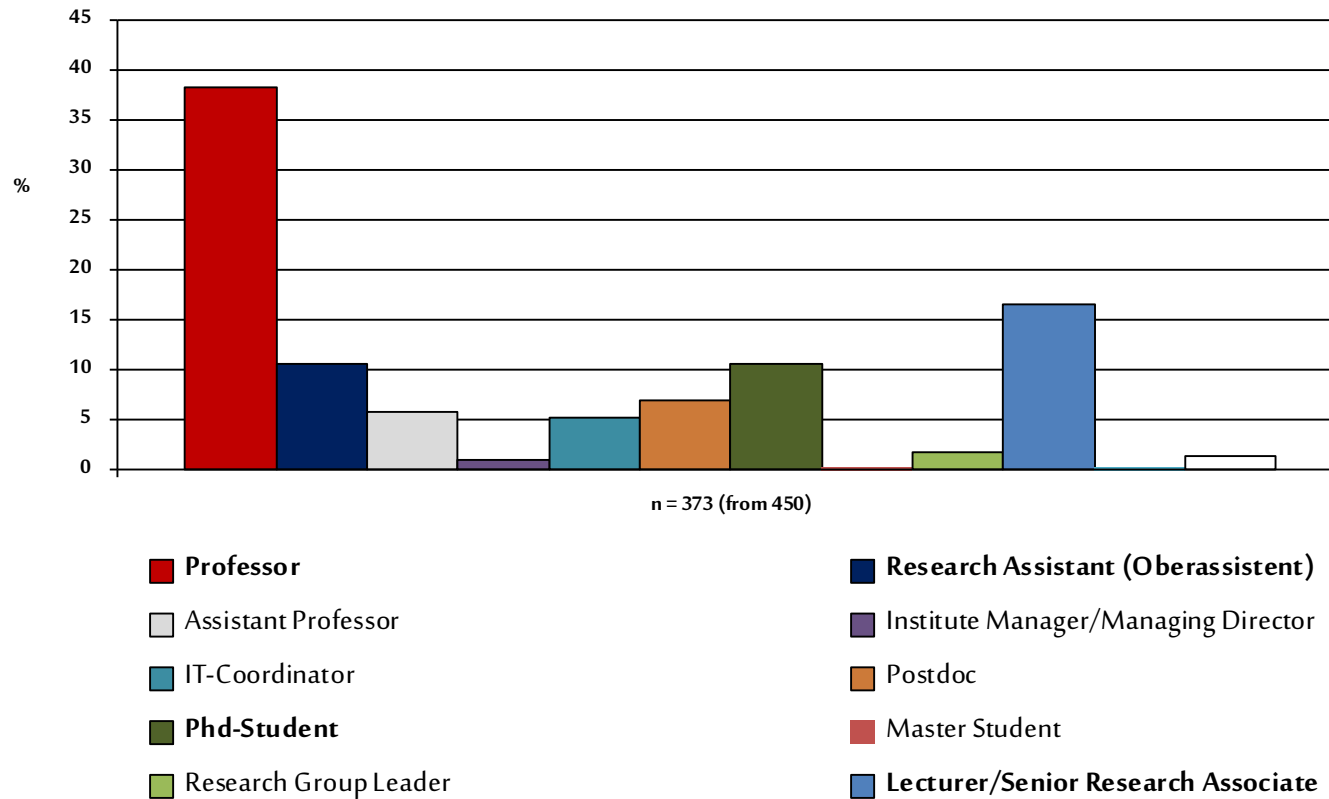
- Differentiation from the international subject specific services
  - The scientists decide about their preferred solution
- A national infrastructure is not available
  - Acceptance increased, but for this local contact persons become even more important

# Online Survey: «Handling Research Data»

## Selected Results

Survey and diagrams: S. Scheid

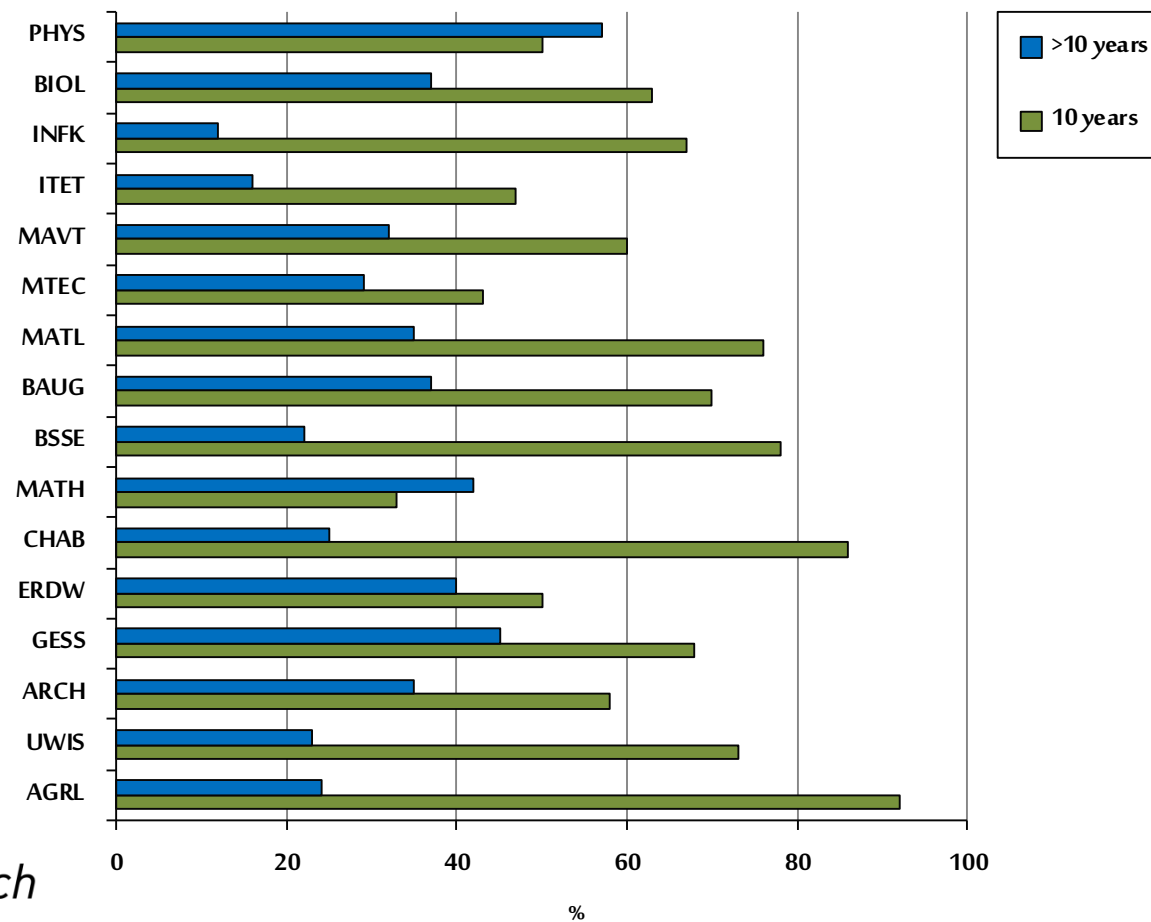
# Participants



D-AGRL	n = 25
D-UWIS	n = 33
D-ARCH	n = 26
D-GESS	n = 21
D-ERDW	n = 20
D-CHAB	n = 36
D-MATH	n = 13
D-BSSE	n = 9
D-BAUG	n = 26
D-MATL	n = 17
D-MTEC	n = 21
D-MAVT	n = 25
D-ITET	n = 20
D-INFK	n = 24
D-BIOL	n = 27
D-PHYS	n = 30

# Period of Storing Data

Which period of time do you or your research group have in mind for storing data?

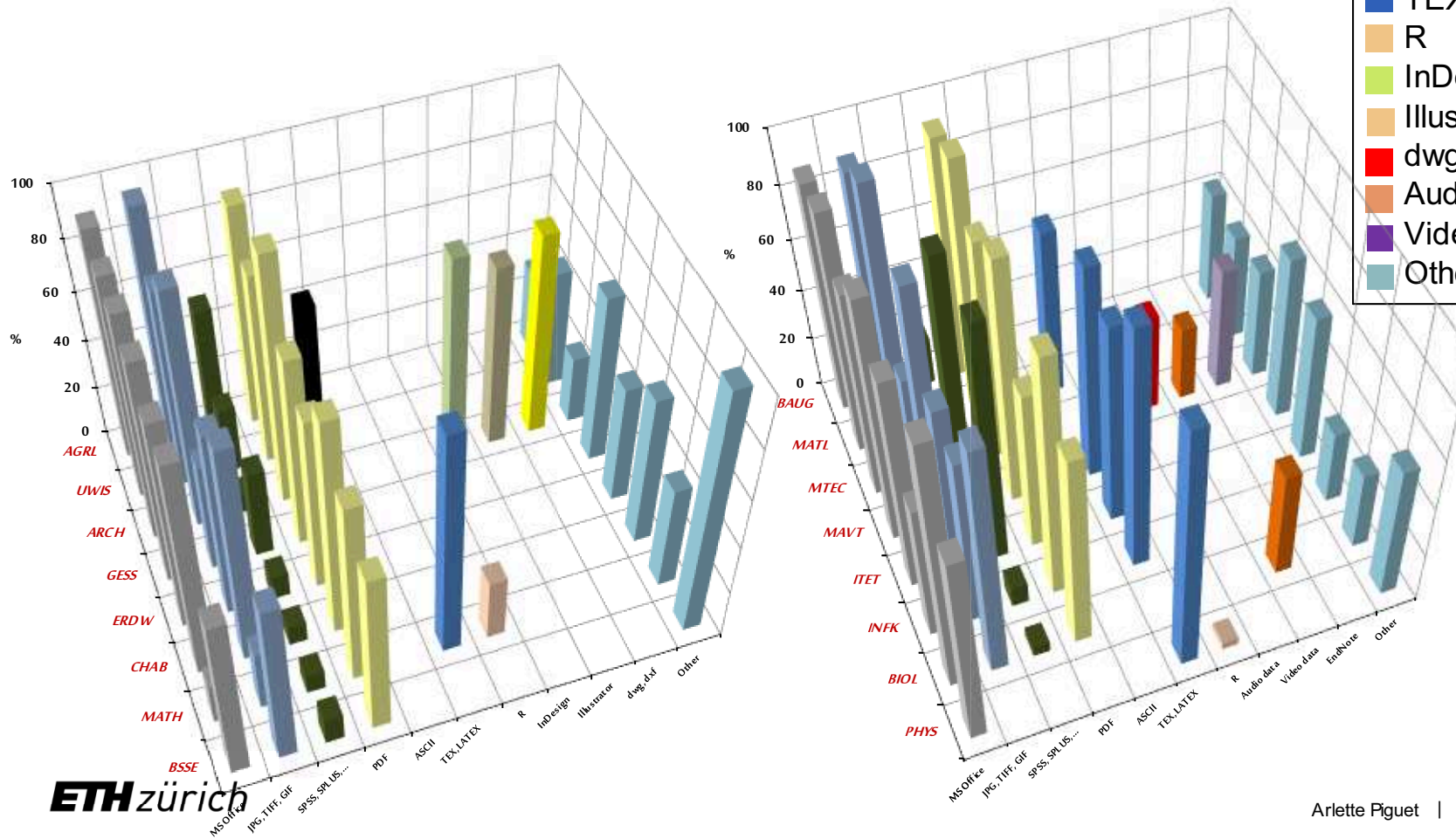




# Formats in Use

Which formats do you and your research group work with?

- MS Office
- JPG, TIFF, GIF
- SPSS, SPLUS
- PDF
- ASCII
- TEX, LATEX
- R
- InDesign
- Illustrator
- dwg, dxf
- Audio data
- Video data
- Other



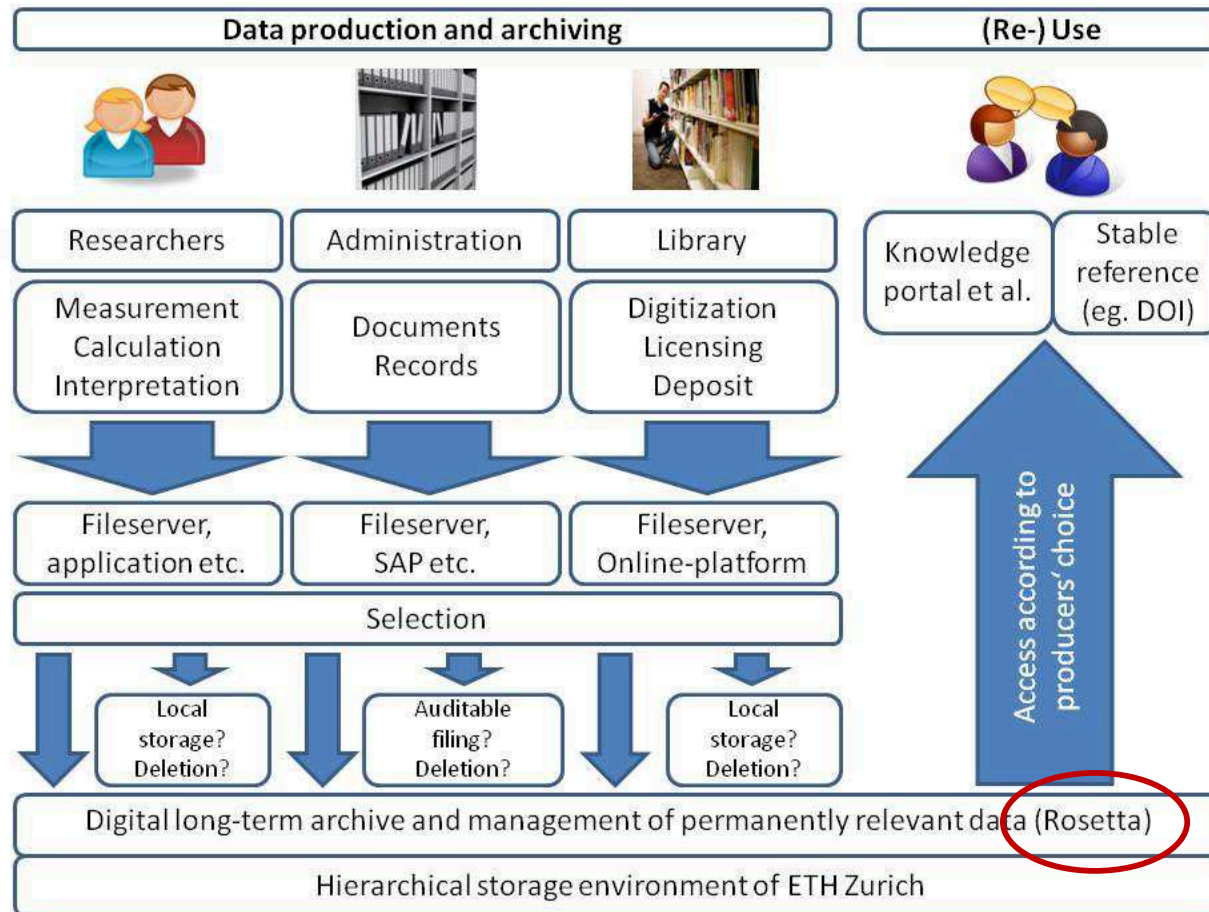
# Points to Consider for a Future «Data Curation Service»

## Researchers...




- want to keep full control at least of who accesses their data - even though they support principally Open Data
- give not generally access to their data to third parties
- show very different levels of understanding and awareness with the matter
- are interested in support for preservation and quality control (checklists, support concerning metadata...)
- want to restructure, select and document data before ingest
- need to keep certain data for limited periods only (e.g. 10 to 12 years)
- see archiving needs often related to data and materials used for publications and want to persistently reference them
- make mixed statements concerning a data policy valid for the whole university
- want no additional workload without additional benefit

# Work Done so Far

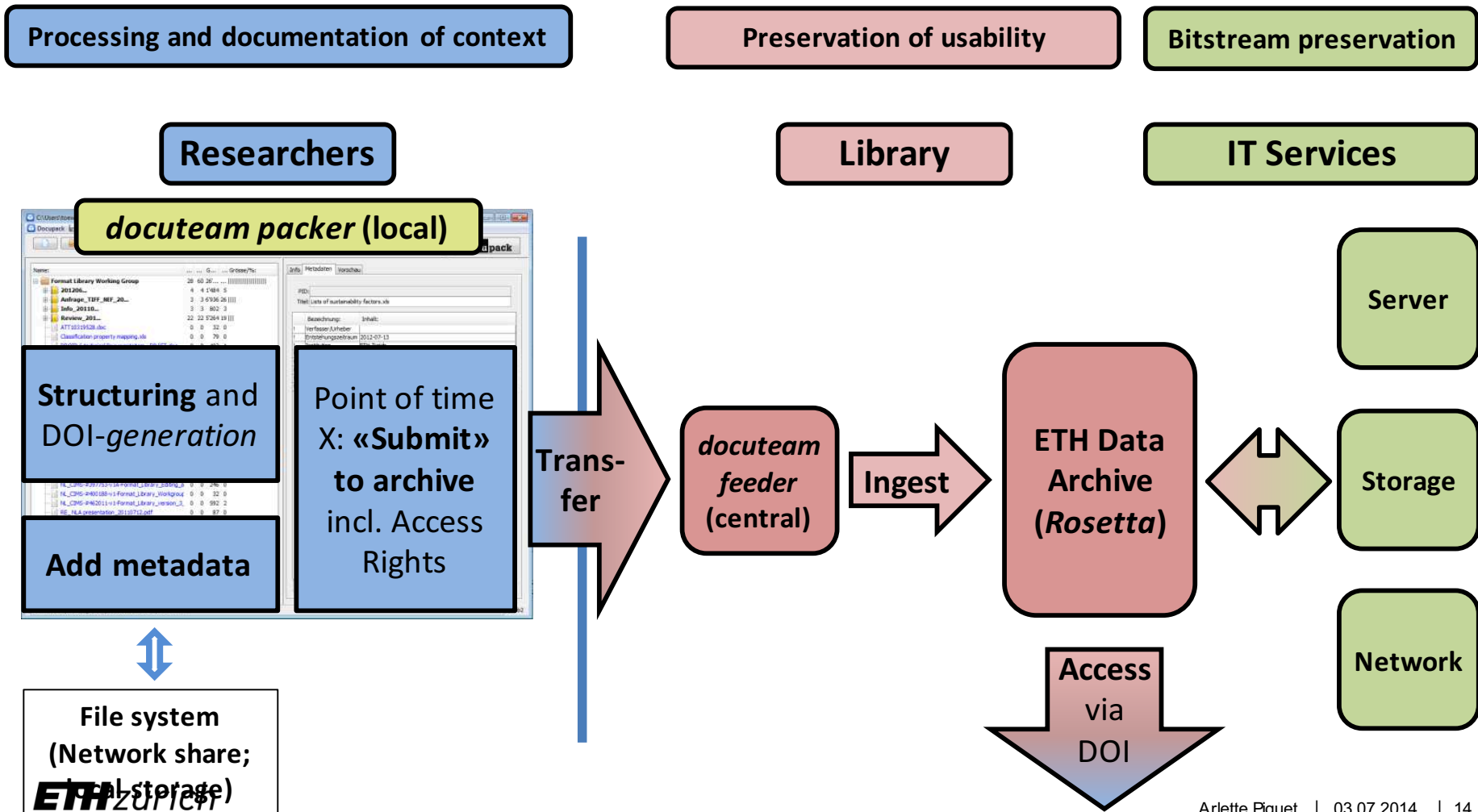
# Rosetta as Common Basis



# One Solution for All Use Cases? (I)

- **For «big data»** (= large quantities of rather unstructured data, within or outside of applications)
  - The need has to be checked, solutions still unclear
- **For structured data from running applications** 
  - E.g. LIMS (Laboratory Information Management System), digital labourjournals
  - Big potential for automation via flexible interfaces
  - Pilot project to be initiated
- **For rather small structured data** 
  - Local data management
  - Manual preparation in a viewer and editor for structuring data and adding metadata (docuteam packer)
  - Data and metadata are stored as long as needed on the local server of each research group
- **For single data sets** 
  - Manual upload

# Workflow: «Small Data»



## One Solution for All Use Cases? (II)

- **For data of the institutional repository**
  - Long term archiving of the data of ETH E-Collection
- **For data of the ETH Archive**
  - Long term archiving of structured data submitted by the record-supplying body



# Discovery and Access

## Research data

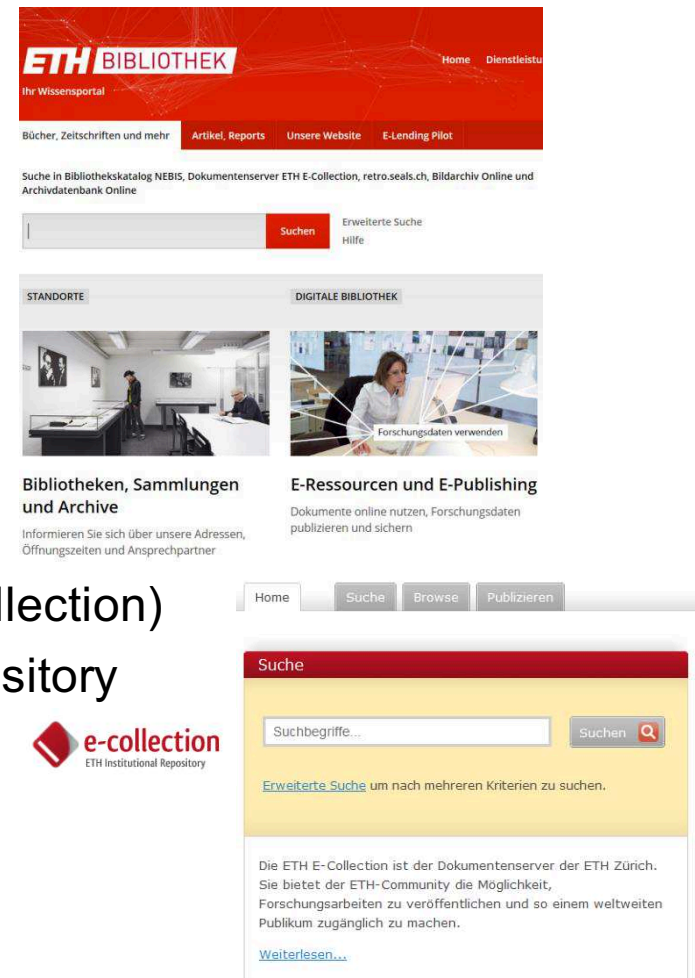
- Metadata will be searchable in Primo
- Links to access data sets in Rosetta

## Library content

- Metadata searchable in Primo (and ETH E-Collection)
- Links to access full texts in the institutional repository
- Rosetta contains the «dark archive»

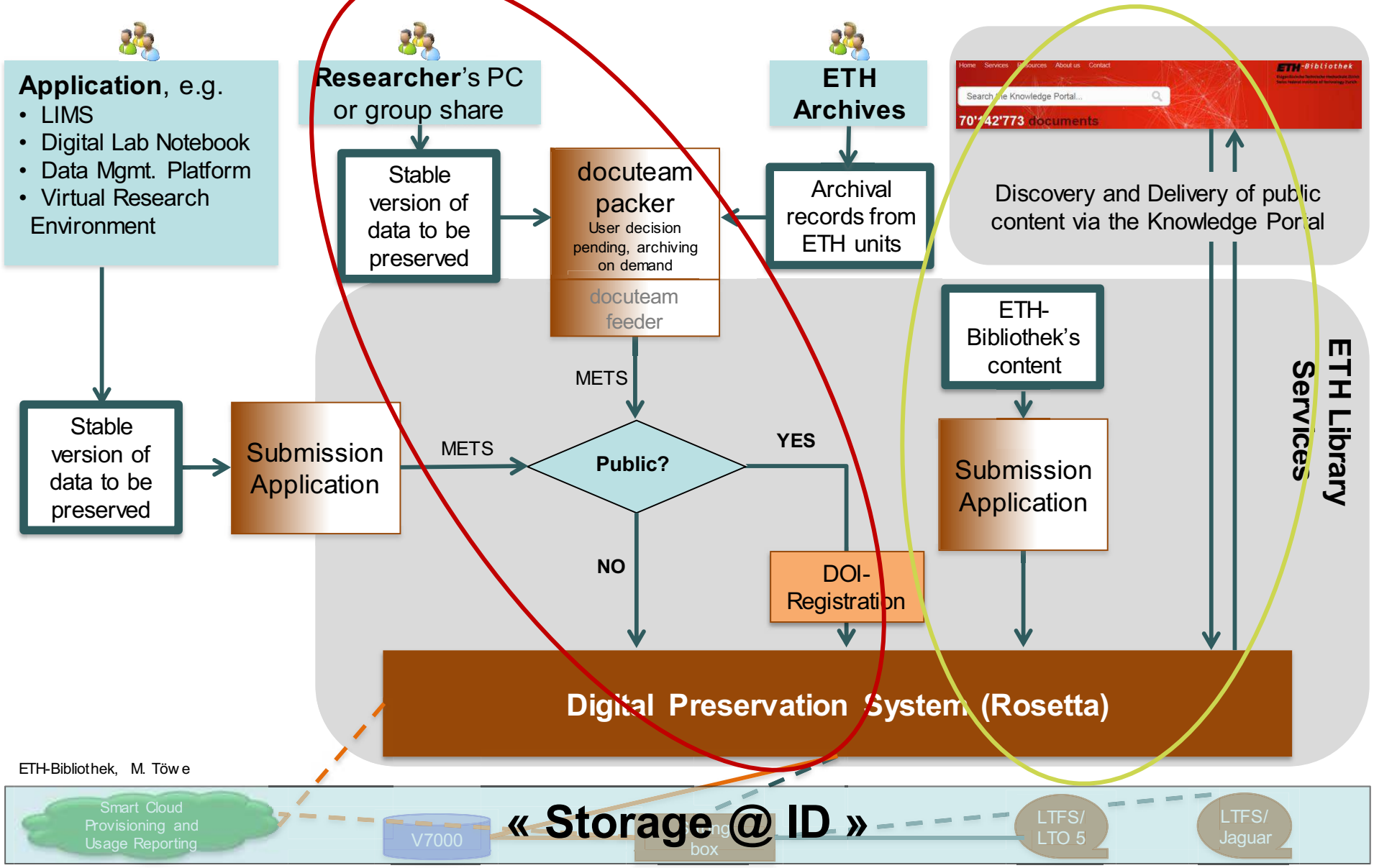
## ETH Archive

- Metadata searchable in Primo
- Links to access the records in the archival database
- Objectes are stored in Rosetta only («light online archive»)





# ETH Library Services



ETH-Bibliothek, M. Töwe

M. Töwe / M. Bärlocher

# Services and Conclusions So Far

# Digital Curation Office at ETH Zurich

- Point of contact for technical and conceptual issues concerning long-term electronic archiving...
- ...and offers researchers help and support in managing and publishing their data
- Provides an infrastructure (ETH Data Archive) for the medium and long-term storage of digital information such as research data, documents or images
- Checklists and FAQs
- ...

<https://www.library.ethz.ch/en/ms/Digital-Curation-at-ETH-Zurich>

The screenshot displays the ETH Bibliothek website interface. At the top, the logo 'ETH BIBLIOTHEK' is visible alongside navigation links for 'Home', 'Services', 'Resources', 'About us', 'Contact', and 'Deutsch'. A search bar is present with a 'Search' button and links to 'Advanced Search' and 'Help'. Below the search bar, the main content area features a section titled 'Digital Curation at ETH Zurich'. This section includes a sub-header 'Digital Curation at ETH Zurich' and a list of links: 'Why digital curation?', 'Research data', 'Digital collections', 'Our services', and 'Checklists and FAQs'. A 'Contact' section is also visible, providing the address 'Digital Curation Office', phone number '+41 44 832 60 32', and email 'data-archive@library.ethz.ch'. The background of the page features a photograph of two researchers in a laboratory setting, with binary code overlaid on the image.

## Conclusions / Final Remarks

- Researchers need additional services to fulfill the requirements for good scientific practice
  - Long term archiving services including support for data management (technical infrastructure ↔ consultancy)
- Consultancy as a central task already at the time of research projects applications?
- Potential for an «ETH E-Collection plus» for adding supplementary material?
- How can the necessary personnel be trained?
  - Research related, library and archival competence, IT-affinity...
- Costs need to be covered long term
- Culture of cooperation between the different stakeholders has to be developed
- National and international information exchange with institutions which face similar challenges are important

Thank you for your attention

I thank Dr. Matthias Töwe and Dr. Wolfram Neubauer for the contribution to this presentation.

[piguet@library.ethz.ch](mailto:piguet@library.ethz.ch)