Experiences from the DLCM pilot projects and possible implications for research data management at universities (of applied sciences)

Andreas Fürholz[A] (fueh@zhaw.ch)

Julian Durrer[B] (durj@zhaw.ch)

Dr. Joël F. Pothier[C] (poth@zhaw.ch)

22.10.2020

[A] Research and Development Unit
[B] Institute of Chemistry and Biotechnology / Section of Polymer Chemistry
[C] Institute of Natural Resource Sciences / Research Group Environmental Genomics und Systems Biology

Personal view

What is that, «Research Data»?

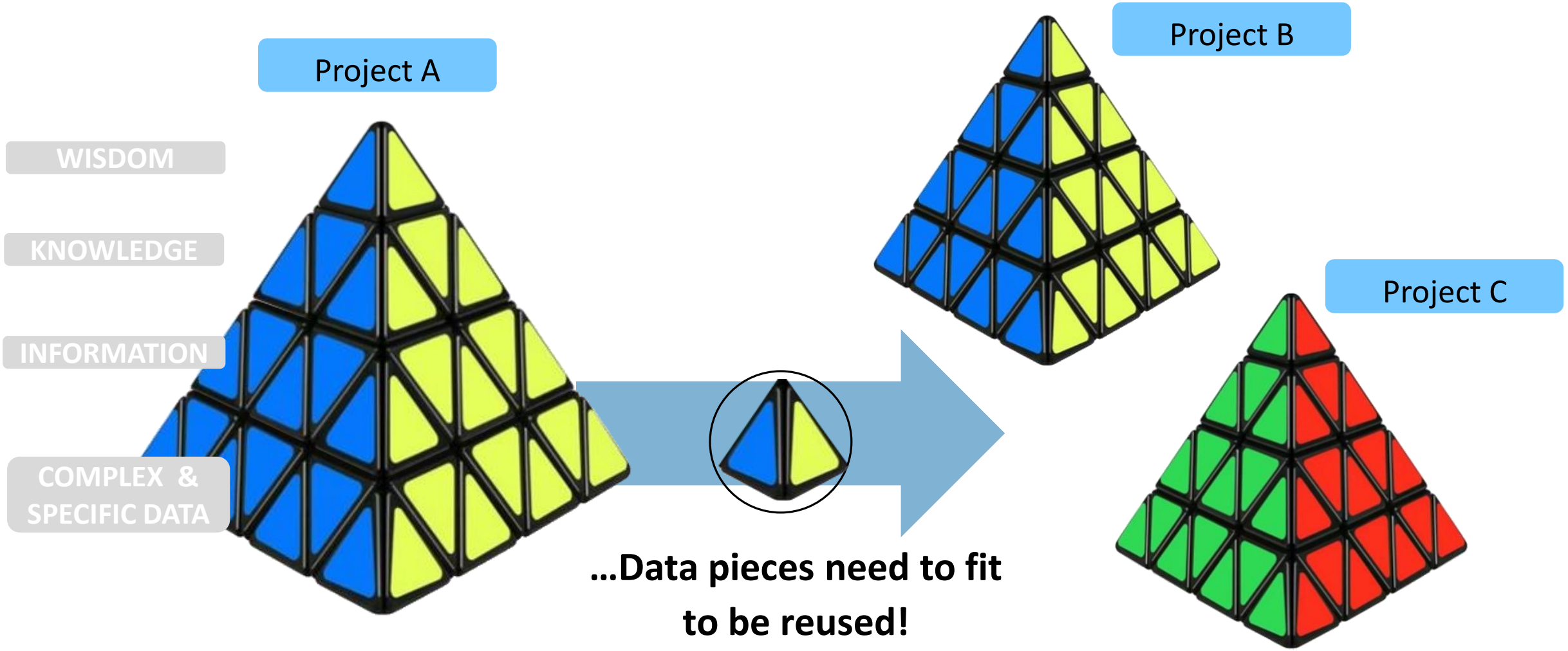
- Like an image/drawing
- Perspective
- Focus
- Boundaries
- …

…Research Data is a (very) complex product!

Painting ("Wimmelbild") about the 31st Chaos Communication Congress in Hamburg
Artist: Caro Wedekind / foxitalic retrieved from Wikipedia

Personal view

…Sharing of data is complex too!



…Data pieces need to fit to be reused!

Personal view

…More "product-thinking" for open research data!?

60g

Maggi Quick Lunch Bolognese Pasta

Maggi

3 ratings

20% discount, 2 or more

3.30

art

availability in store

…which product…

- … solves problem of a customer
- … has a defined quality
- … producer gets rewarded

- … is comprehensible, regarding…
  - / Content / Standards
  - / Quality / How (not) to use

- … it has (most likely) a…
  - / Support / Marketing
  - / Product Management

And what about research data?

Road to an Open Science / Open Data Culture

…DLCM-project has contributed in various ways

Zurich University of Applied Sciences

zh aw

Open Data Culture

ZHAW: Publication/archiving of research data → "ORD-pilots"

Co-development OLOS web portal

Use and test DLCM professional services (trainings, consulting, support)

Publication of research data and measuring of impact
(including identification of discipline specific repositories)

DLCM

Archiving

Repositories

…

Infrastructure

ELN/LIMS

openBIS

Tools

…

RDM-Services
(Training, Consulting, etc.)

DMPs

Policies

ZHAW: Active research data management → "openBIS-pilots*"

Contribution to shape solution, training and support

Connection to archiving- / publication system

openRDM .swiss

*openBIS is **an Electronic Lab Notebook** / Laboratory Information and Management System (ELN/LIMS) developed by Scientific IT services of ETHZ.
Further info: https://openbis.ch

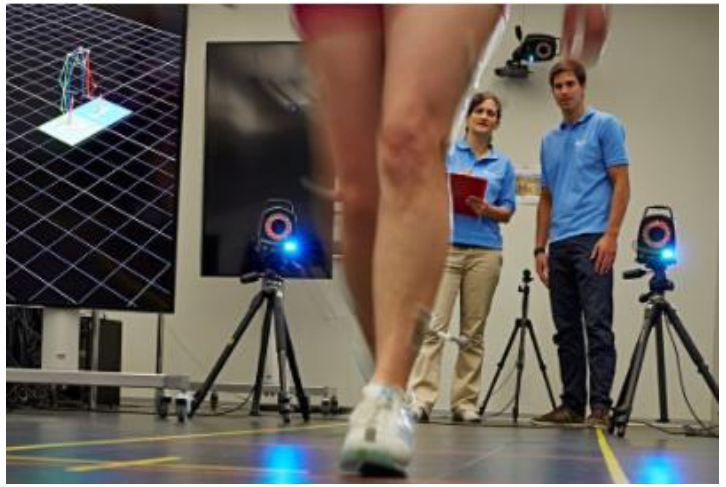
Overview ORD-pilots

		Description of Research Data						
Departement	**Funding of related project**	**Observational**	**Experimental**	**Simulation**	**Derived**	**Reference**	**Digitalisation**	**Description (published data only)**
Architecture, Design and Civil Engineering	BAK/OFC/FOC, Foundations	(I)			(x)	(x)	**x**	Digitized physical architectural models
Health Professions	SAMW/ASSM, Käthe-Zingg-Schwichtenberg-Foundation	**S** (I)						Survey
Applied **Linguistics**	various				**x**	**x**		Text-Data (XML, raw)
Life Sciences and Facility Management	EU (FP7, Nr. 613678)		**x**					Genome sequence data
Applied **Psychology**	SNSF (Nr. 132278)	**S**						Survey
	Swiss Health Observatory OBSAN	**S**						Survey
Social Work	SNSF (Nr. 169727)	**I**						Interviews
Engineering	CTI (Nr. 16851.1 PFNM-NM)		**x**	(x)	(x)			Tomography data
	SNSF (NRP 70/71, "Energy Turnaround")	**x**	**x**	**x**	**x**	(x)		Survey Code/Software Tomography data
Management and Law	SNSF (Nr. 162948)	**S**						Survey

S: Survey I: Interview (x): Data used in project but not published

Overview openBIS-pilots

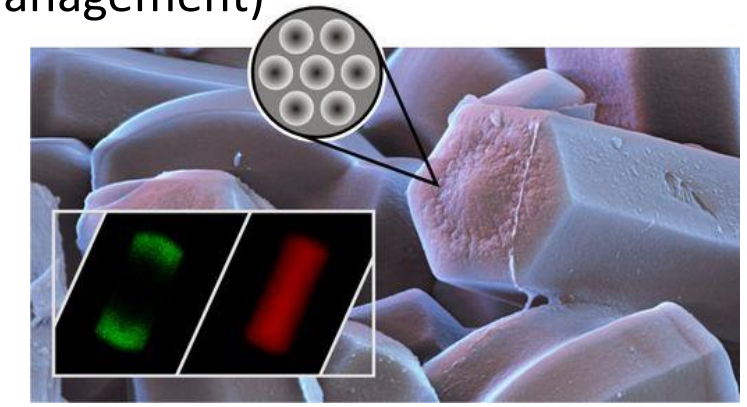
Use case at Movement Laboratory (Health Professions)



Main challenges:

- **Implementation of a sophisticated case report form** (e.g. with validation, flexibility number of data fields)
- **In situ data capturing with presence of probands** (requiring a good usability)

Use case at Polymer Chemistry Lab (Life Sciences and Facility Management)



Main challenges:

- **Integration of tools and workflows**
- **Different users (from students to PhDs)**

Side note

Why using Electronical Laboratory Notebooks?

- **Supports good scientific practice; Structuring/standardising of research workflows and documentation**
- Reach **compliance to Good Laboratory Praxis (GLP)** and other **quality systems** (access control, full audit trail, digital signatures, backup, etc.)
- **Managing data according FAIR principles**
- **(Many) advantages over paper notebooks:**

Support collaboration

Cannot get lost/destroyed

Suitable for clean/bio labs

Legible records

Stealing more unlikely

Searchable & accessible → reuse more likely

Better overview for lab managers

…

openBIS-pilot at Polymer Chemistry Lab

Julian Durrer (durj@zhaw.ch)
Institute of Chemistry and Biotechnology / Section of Polymer Chemistry

ORD-pilot “OMICS-Data”

Dr. Joël Pothier (poth@zhaw.ch)
Institute of Natural Resource Sciences / Research Group Environmental Genomics und Systems Biology

Zurich University of Applied Sciences

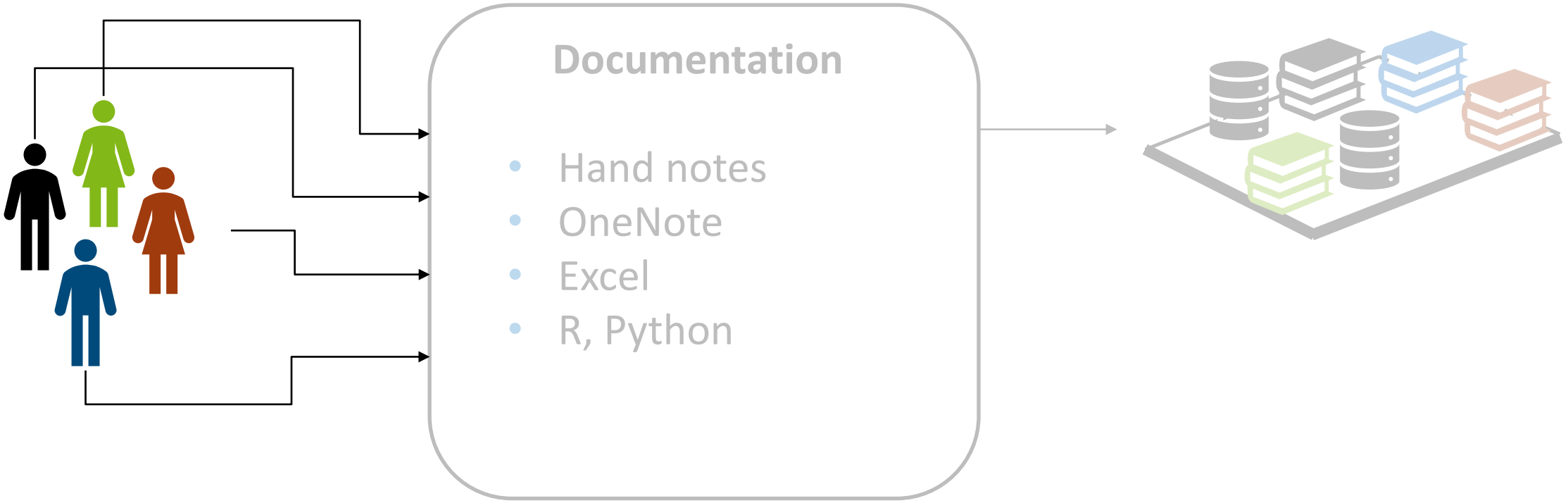
openBIS-pilot: Polymer Chemistry Lab

Institute of Chemistry and Biotechnology / Section of Polymer Chemistry

Zurich University of Applied Sciences

Section of Polymer Chemistry

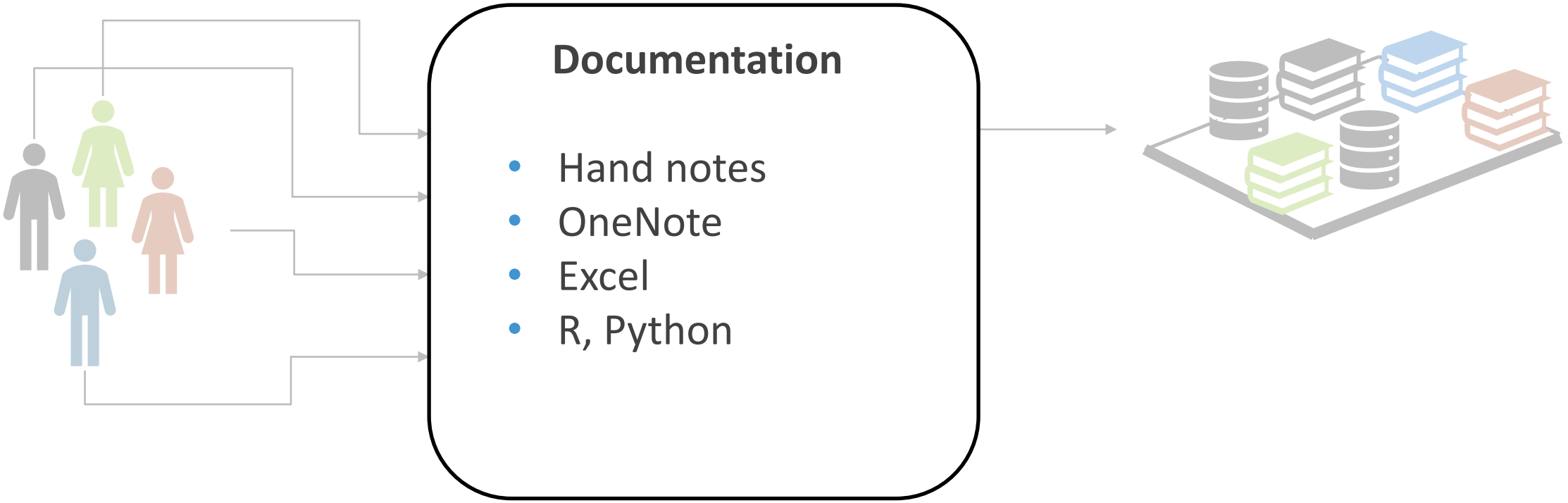
Demand for an Electronic Lab Notebook (ELN)



Zurich University of Applied Sciences

Section of Polymer Chemistry

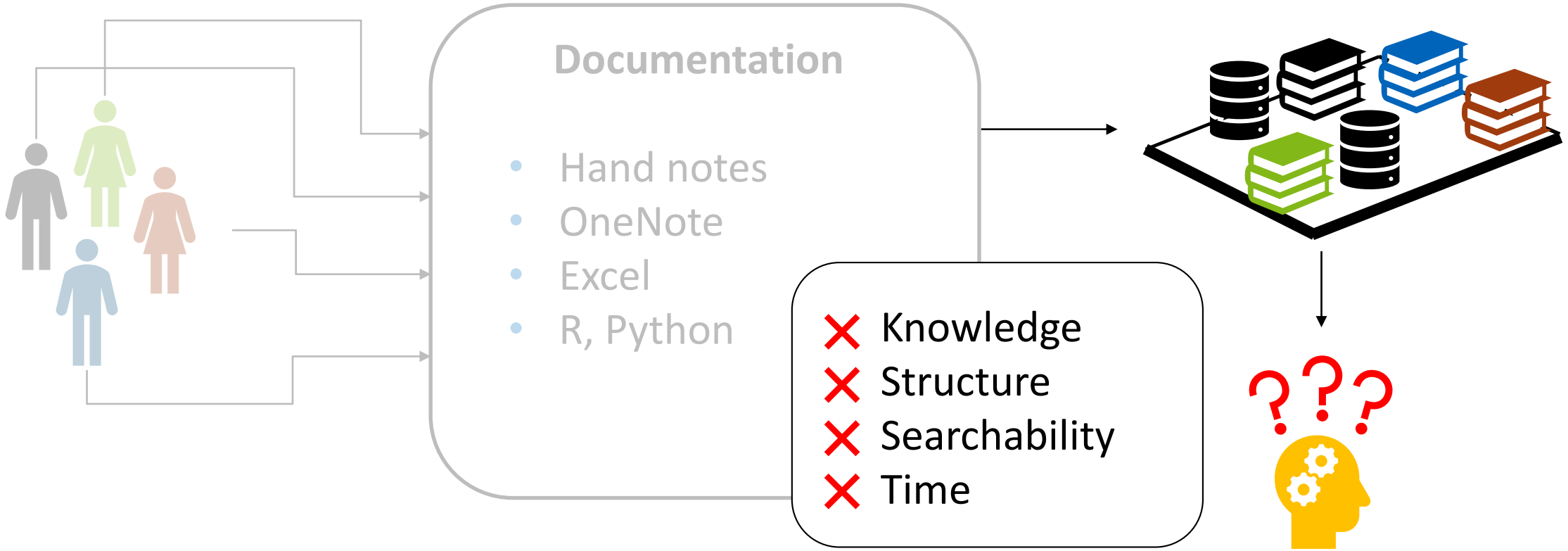
Demand for an Electronic Lab Notebook (ELN)



Zurich University of Applied Sciences

Section of Polymer Chemistry

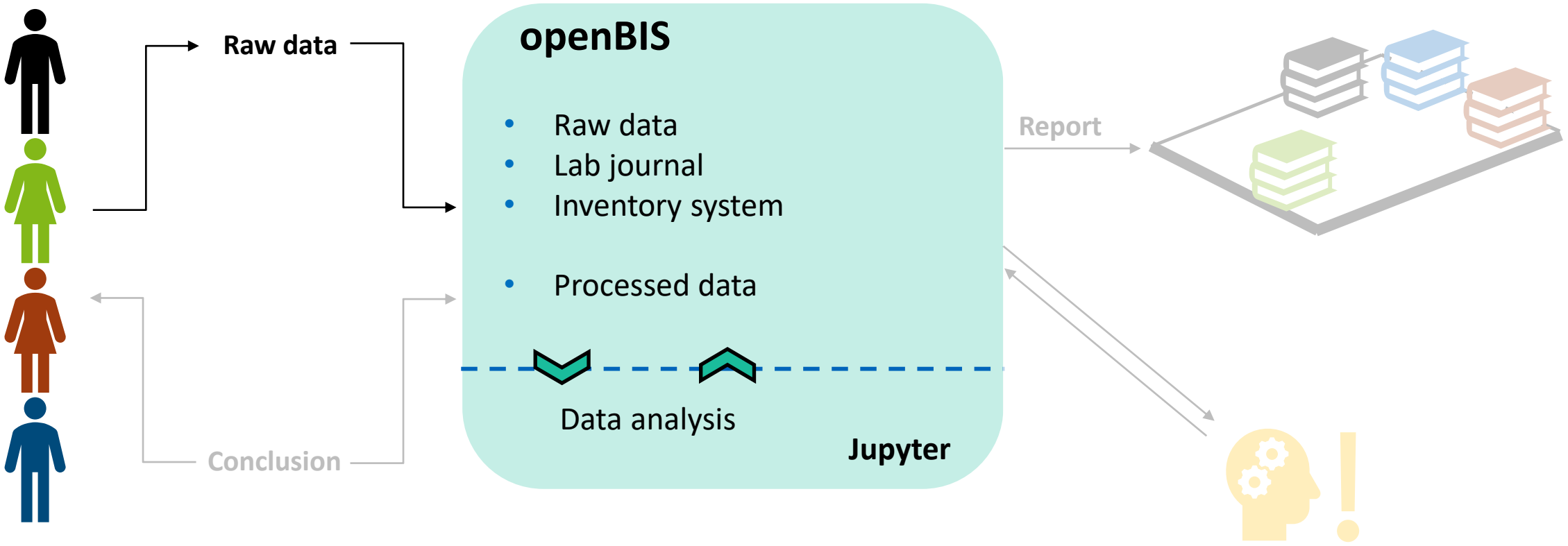
Demand for an Electronic Lab Notebook (ELN)



Zurich University of Applied Sciences

Experience with openBIS

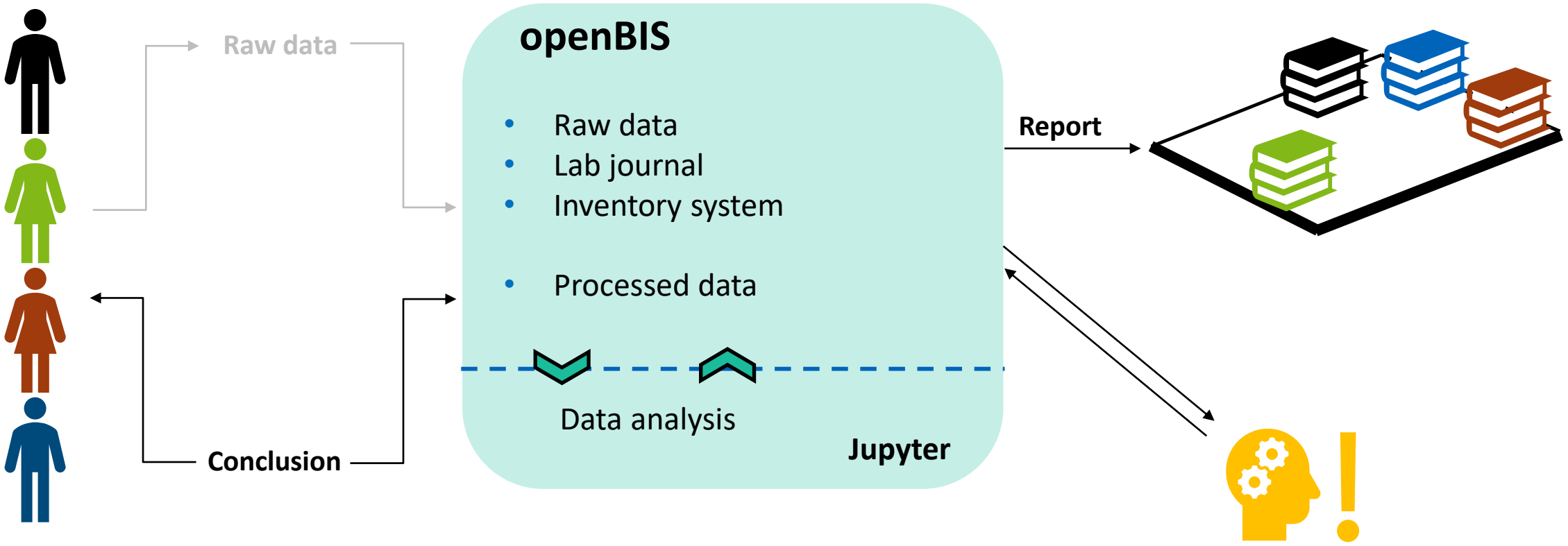
View from an academic side



Zurich University of Applied Sciences

Experience with openBIS

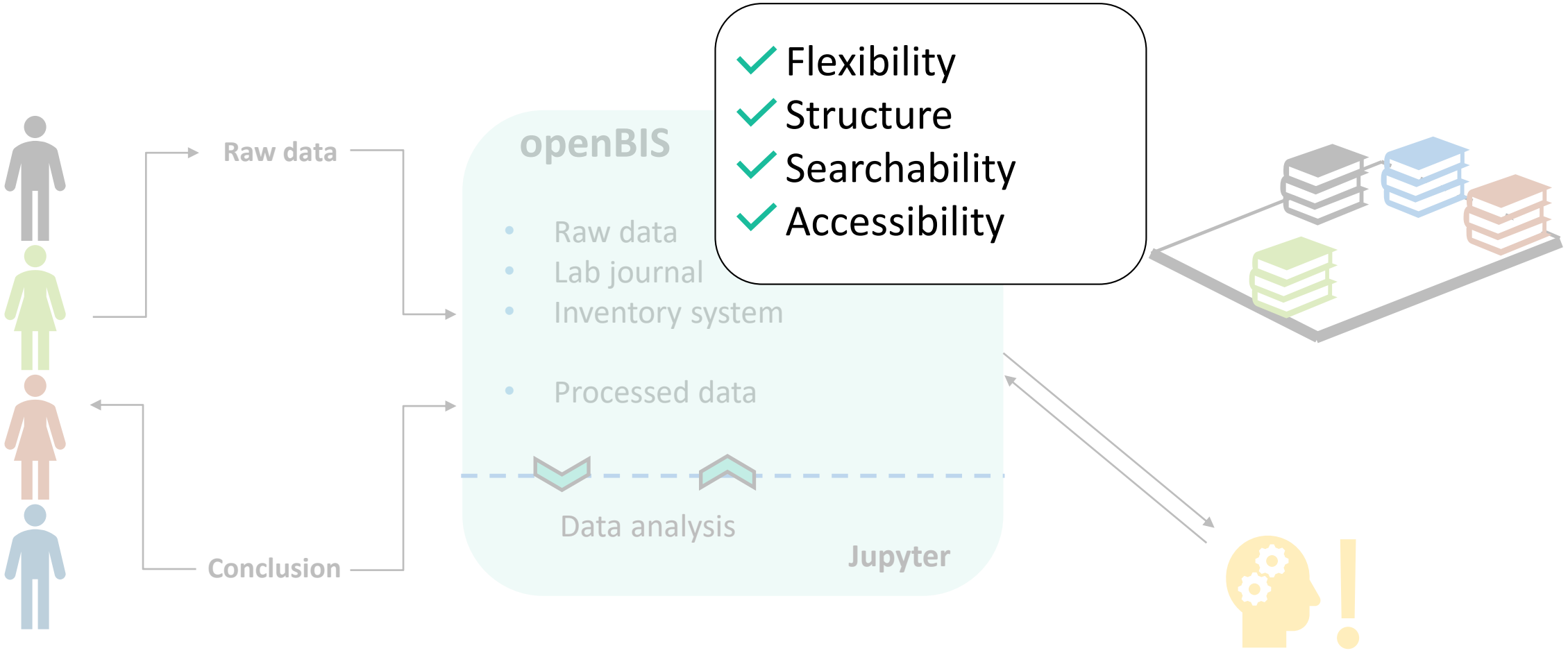
View from an academic side



Zurich University of Applied Sciences

Experience with openBIS

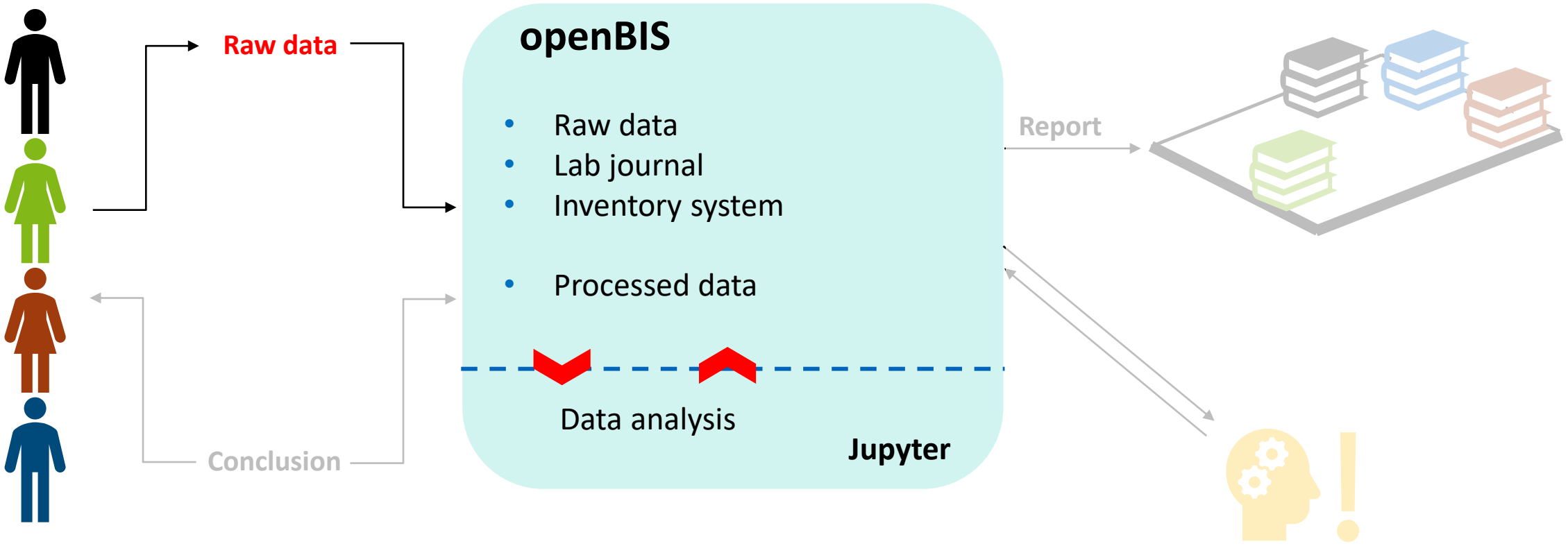
View from an academic side



Zurich University of Applied Sciences

Challenges during implementation

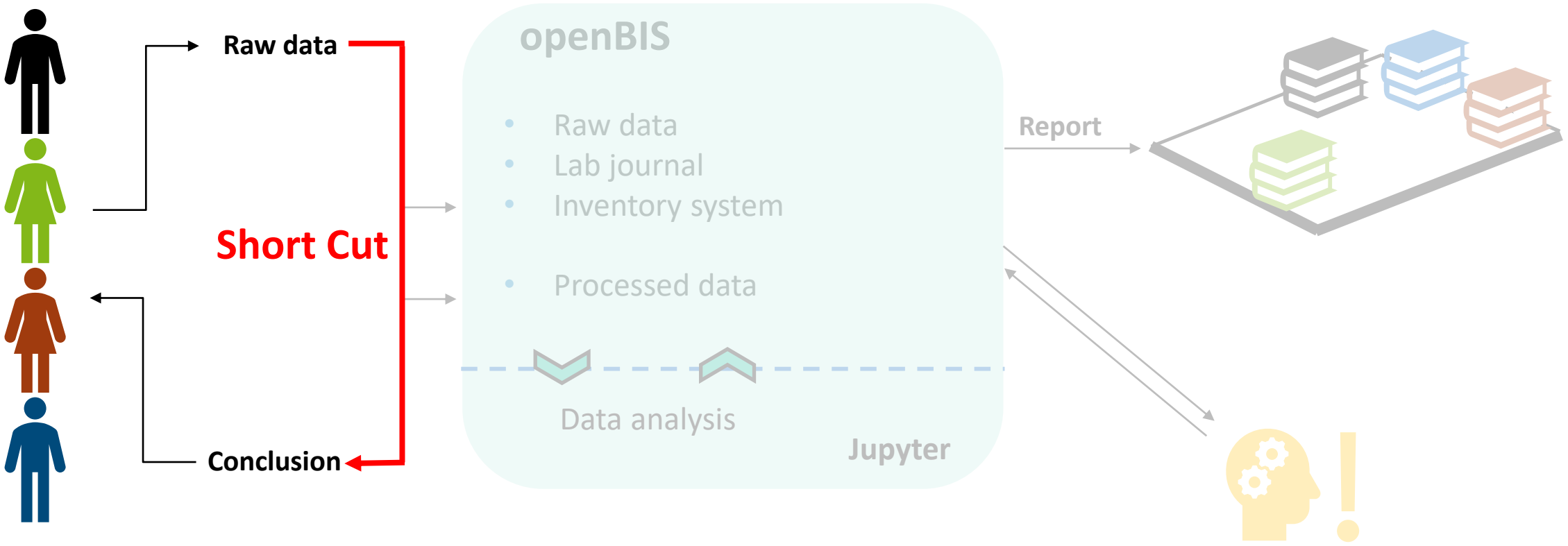
View from an academic side



Zurich University of Applied Sciences

Challenges during implementation

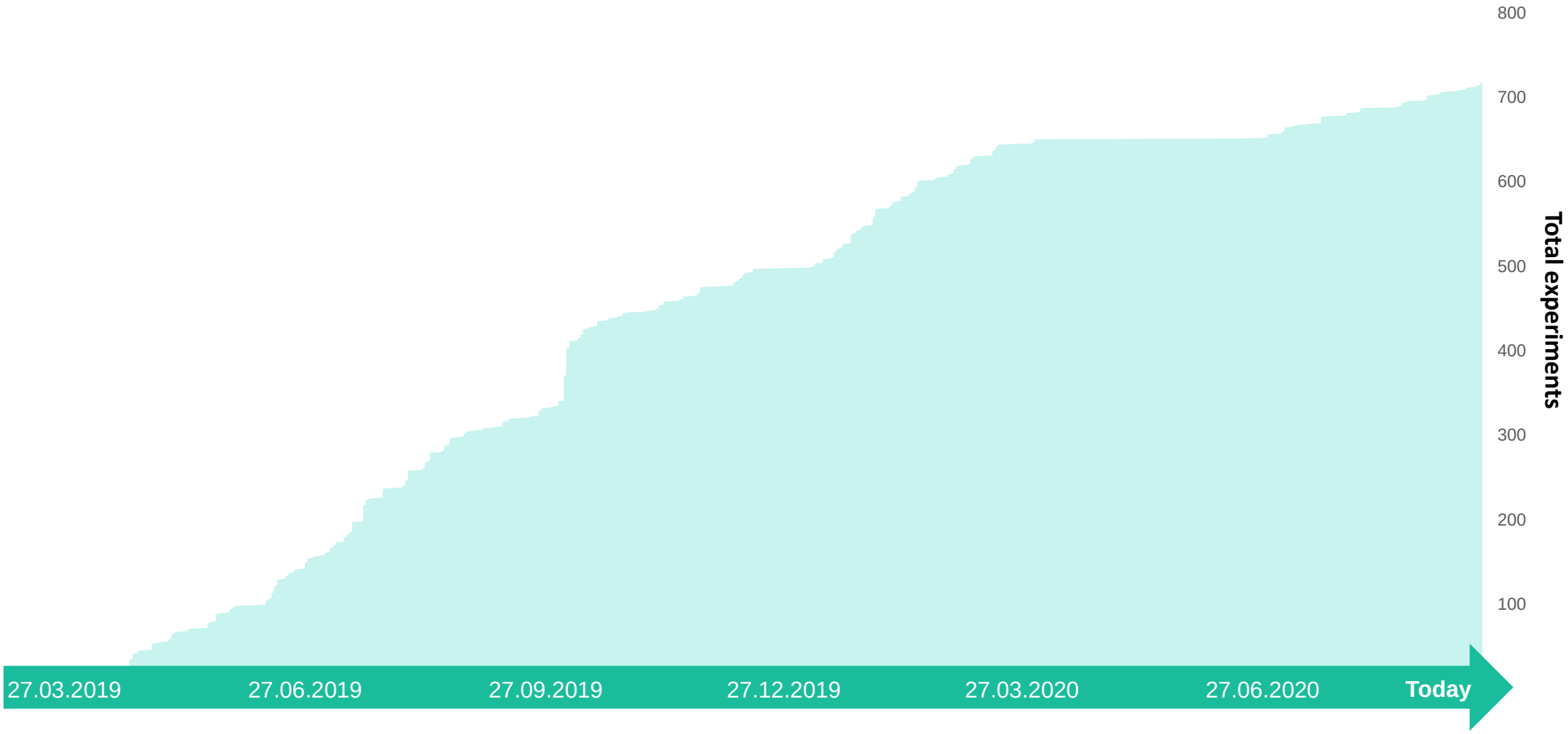
View from an academic side



Zurich University of Applied Sciences

Pilot implementation: Summary

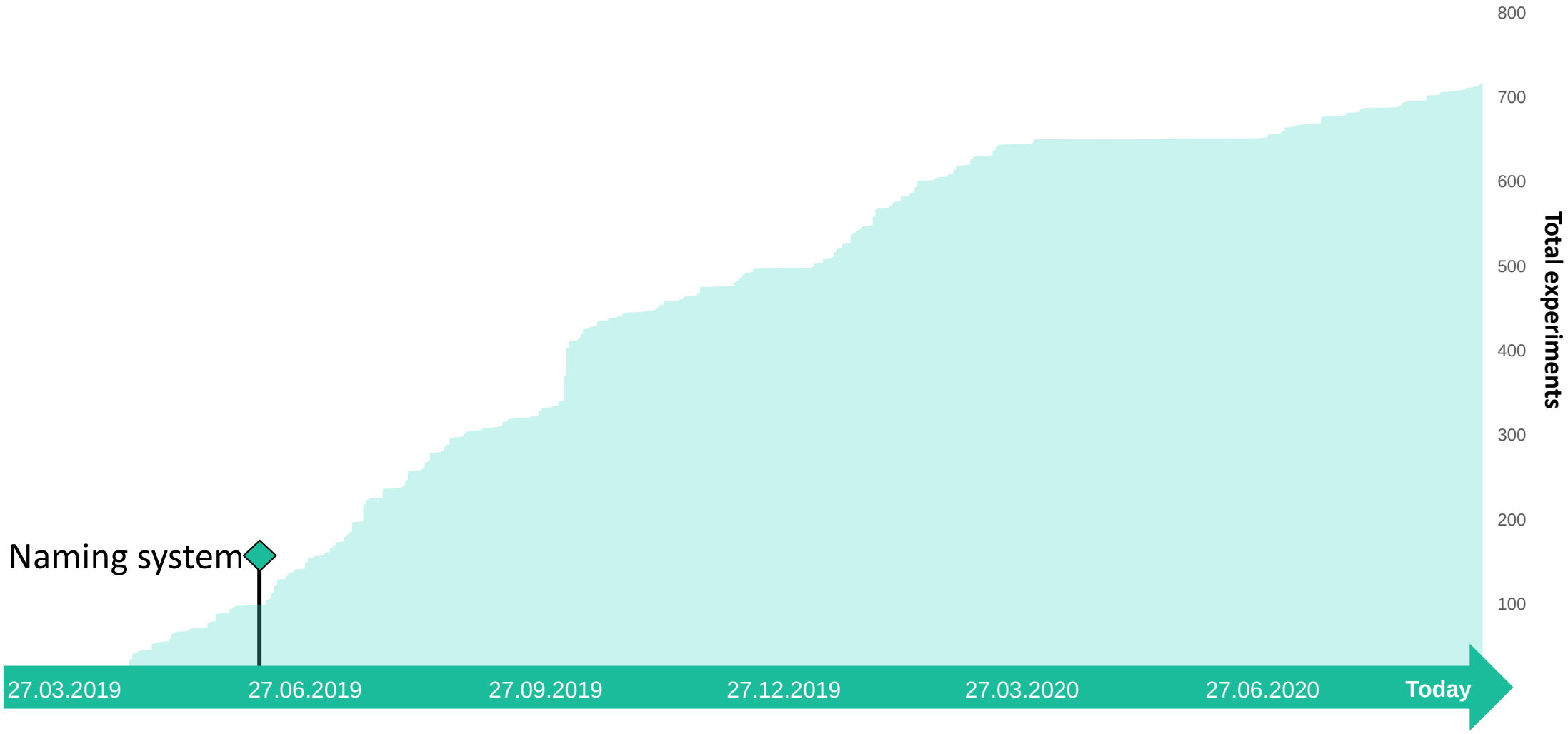
openBIS-Experience



Zurich University of Applied Sciences

Pilot implementation: Summary

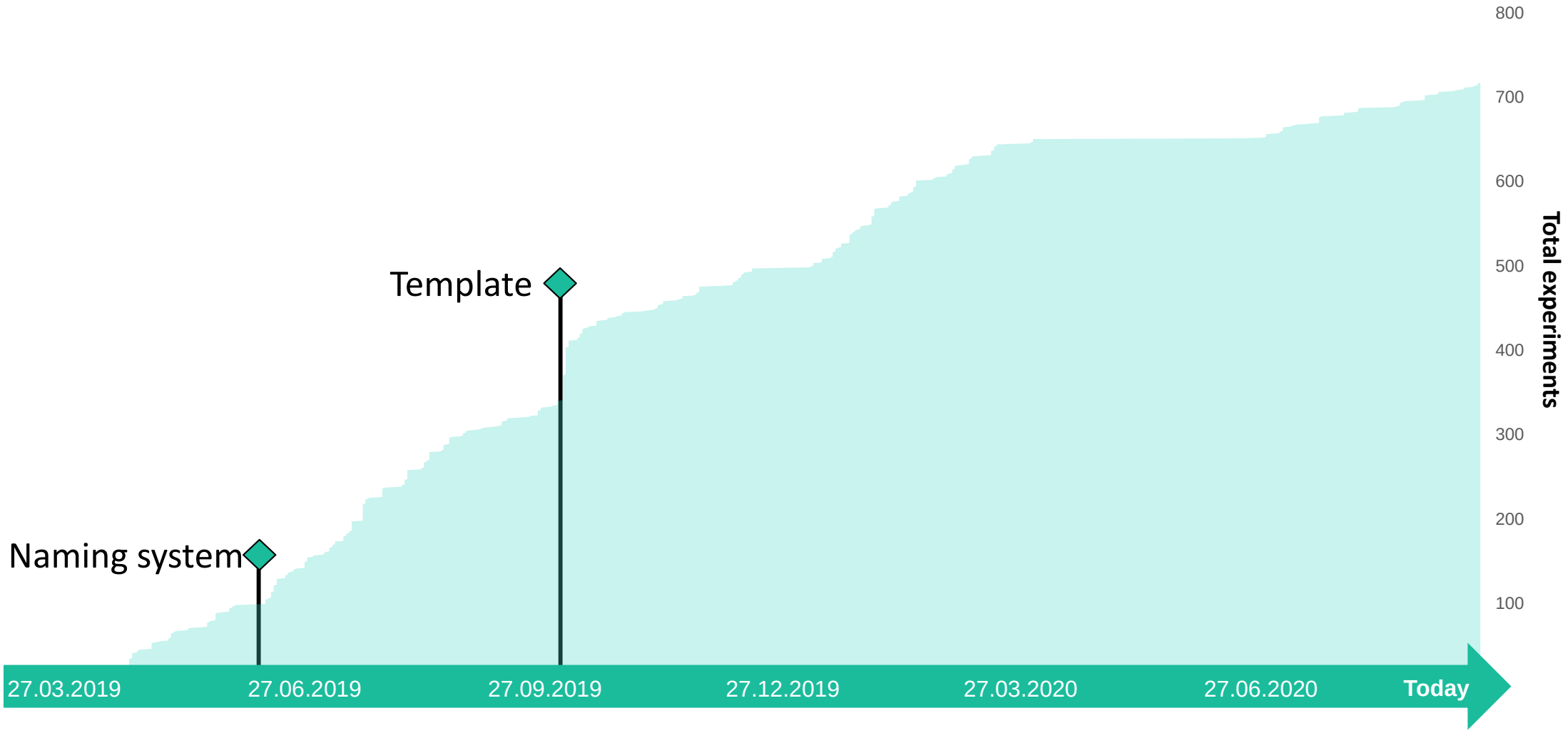
openBIS-Experience



Zurich University of Applied Sciences

Pilot implementation: Summary

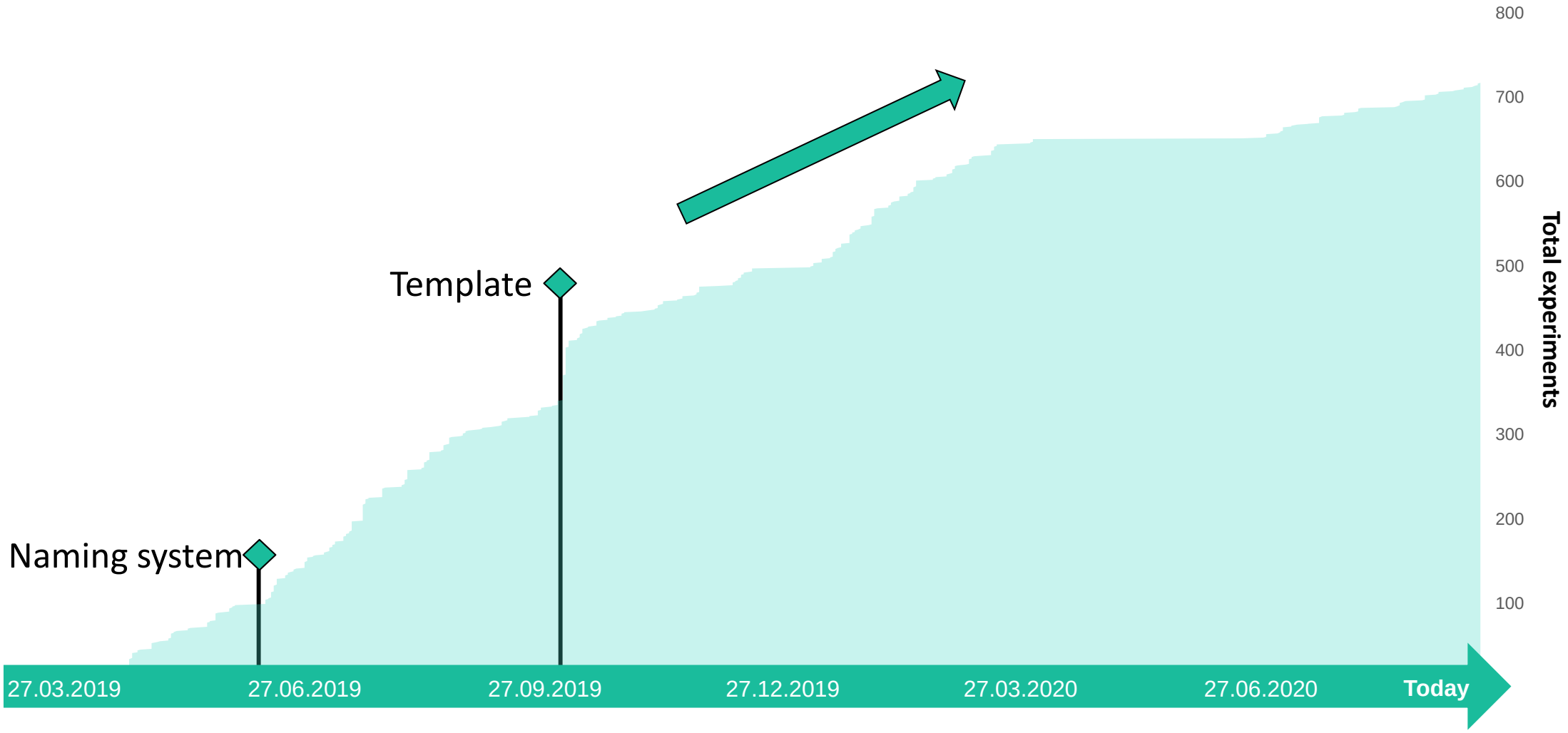
openBIS-Experience



Zurich University of Applied Sciences

Pilot implementation: Summary

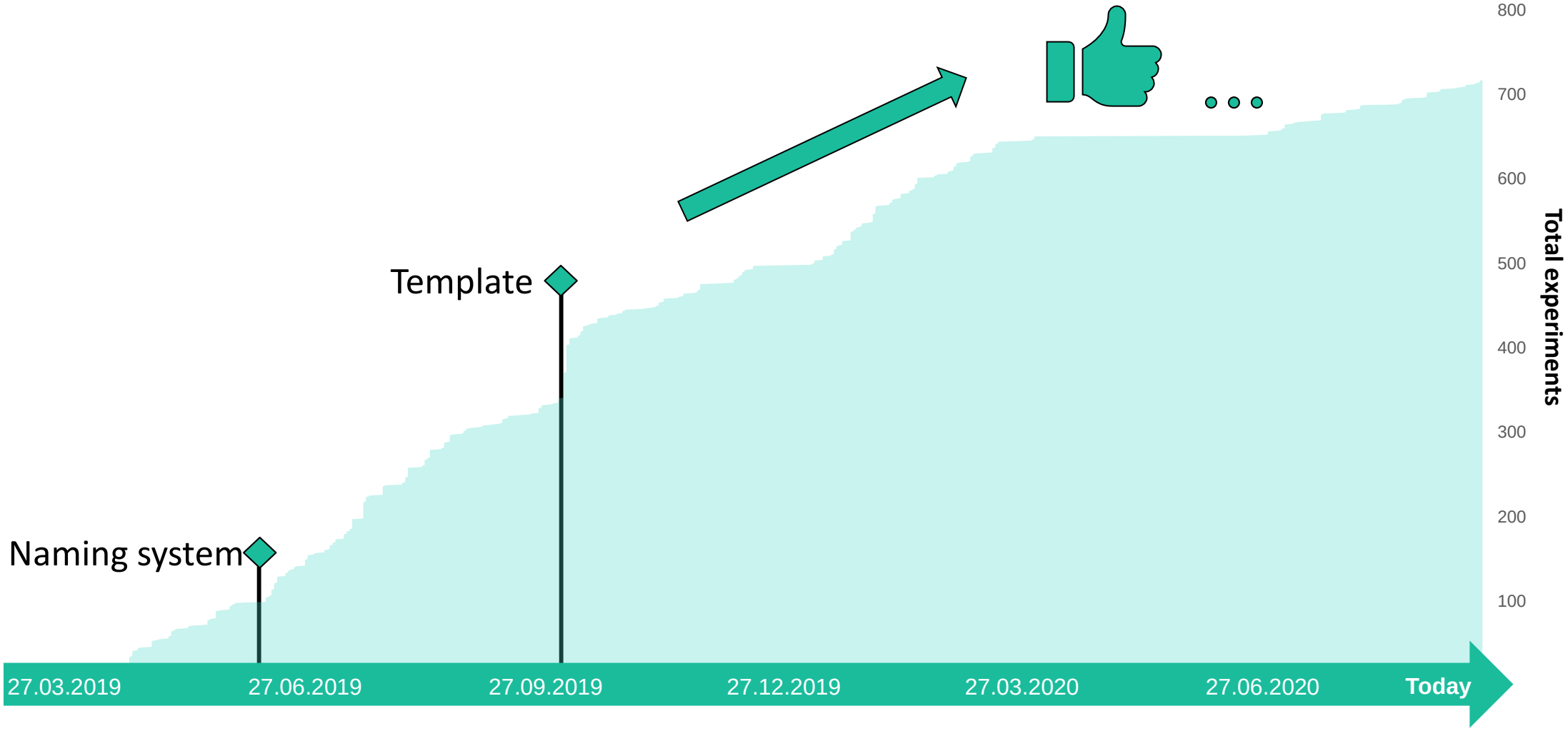
openBIS-Experience



Zurich University of Applied Sciences

Pilot implementation: Summary

openBIS-Experience



Zurich University of Applied Sciences

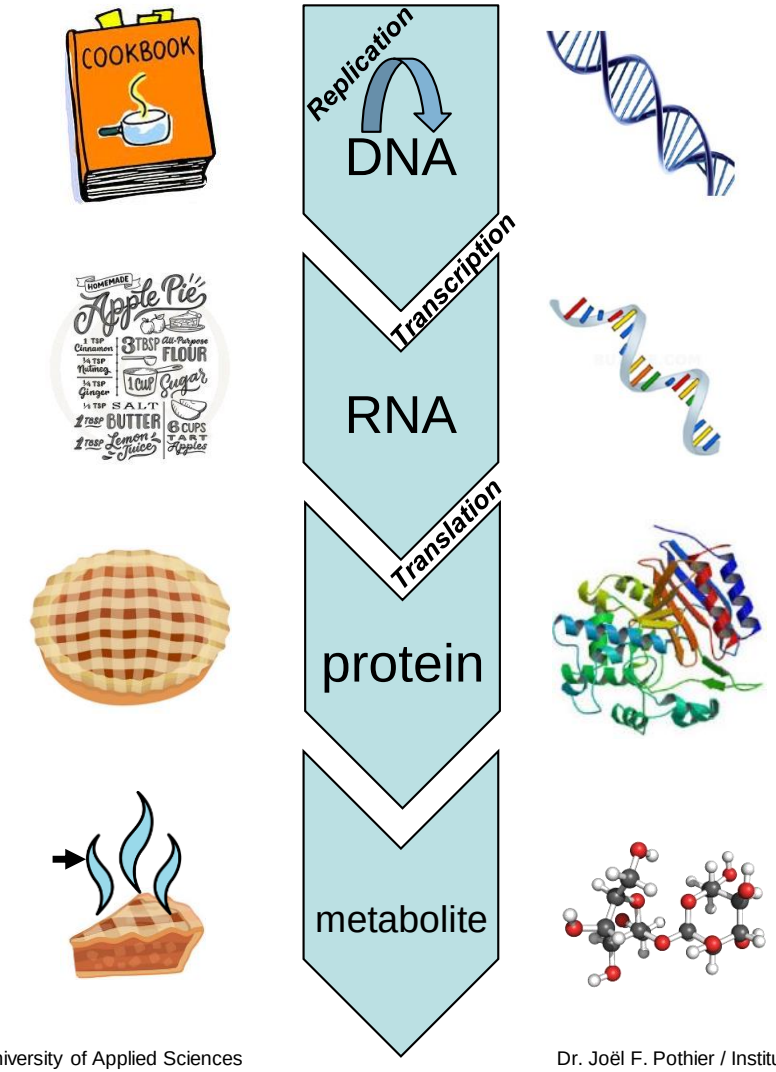
ORD-pilot “OMICS-Data”

Dr. Joël F. Pothier (poth@zhaw.ch)
Institute of Natural Resource Sciences / Research Group Environmental Genomics and Systems Biology

Zurich University of Applied Sciences

OMICS-data

What are OMICS data?



= (*meta*)Genomics

= (*meta*)Transcriptomics

= (*meta*)Proteomics

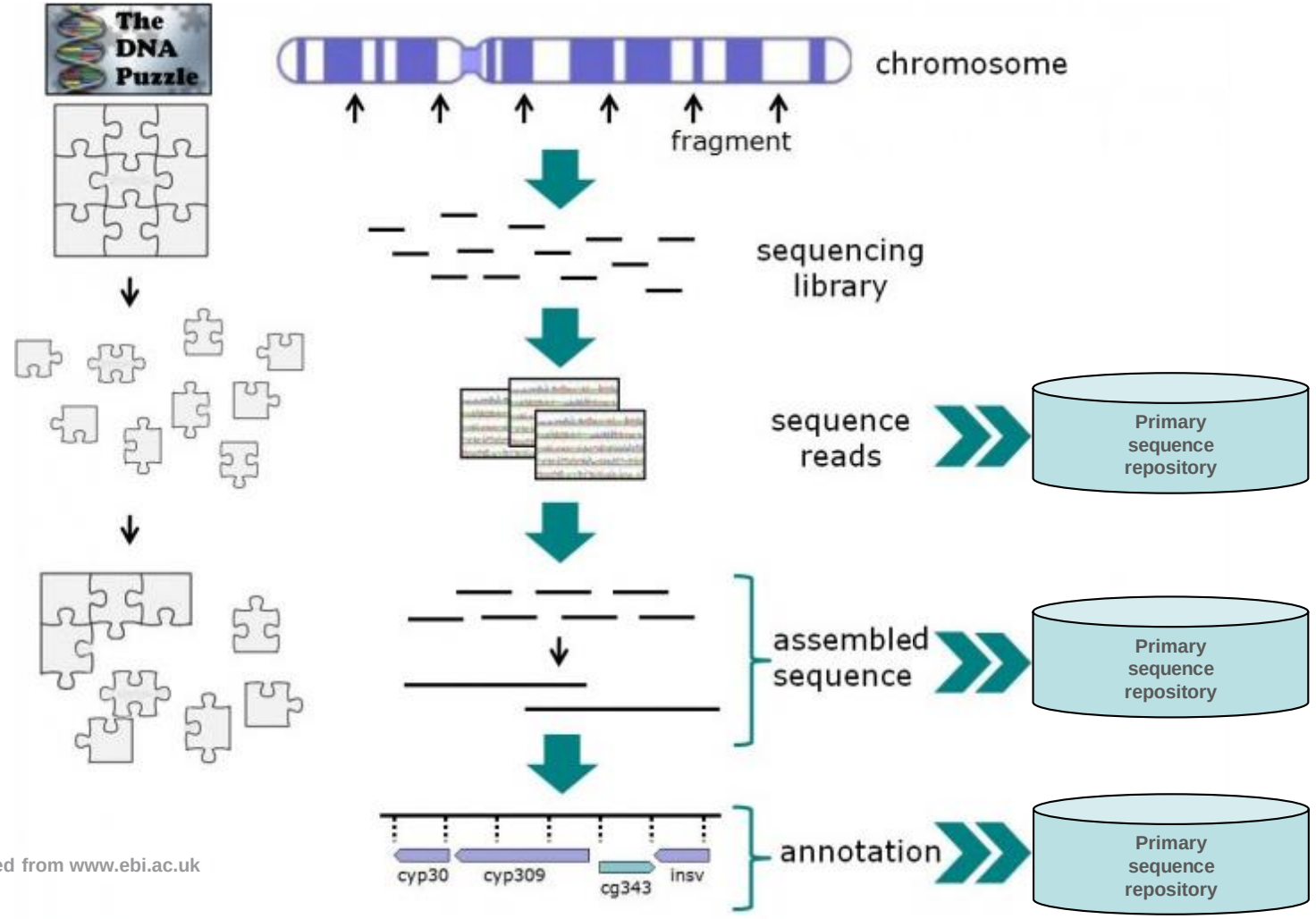
= Metabolomics

OMICS approaches

Zurich University of Applied Sciences

Nucleotide primary data

The DNA puzzle and primary sequence data

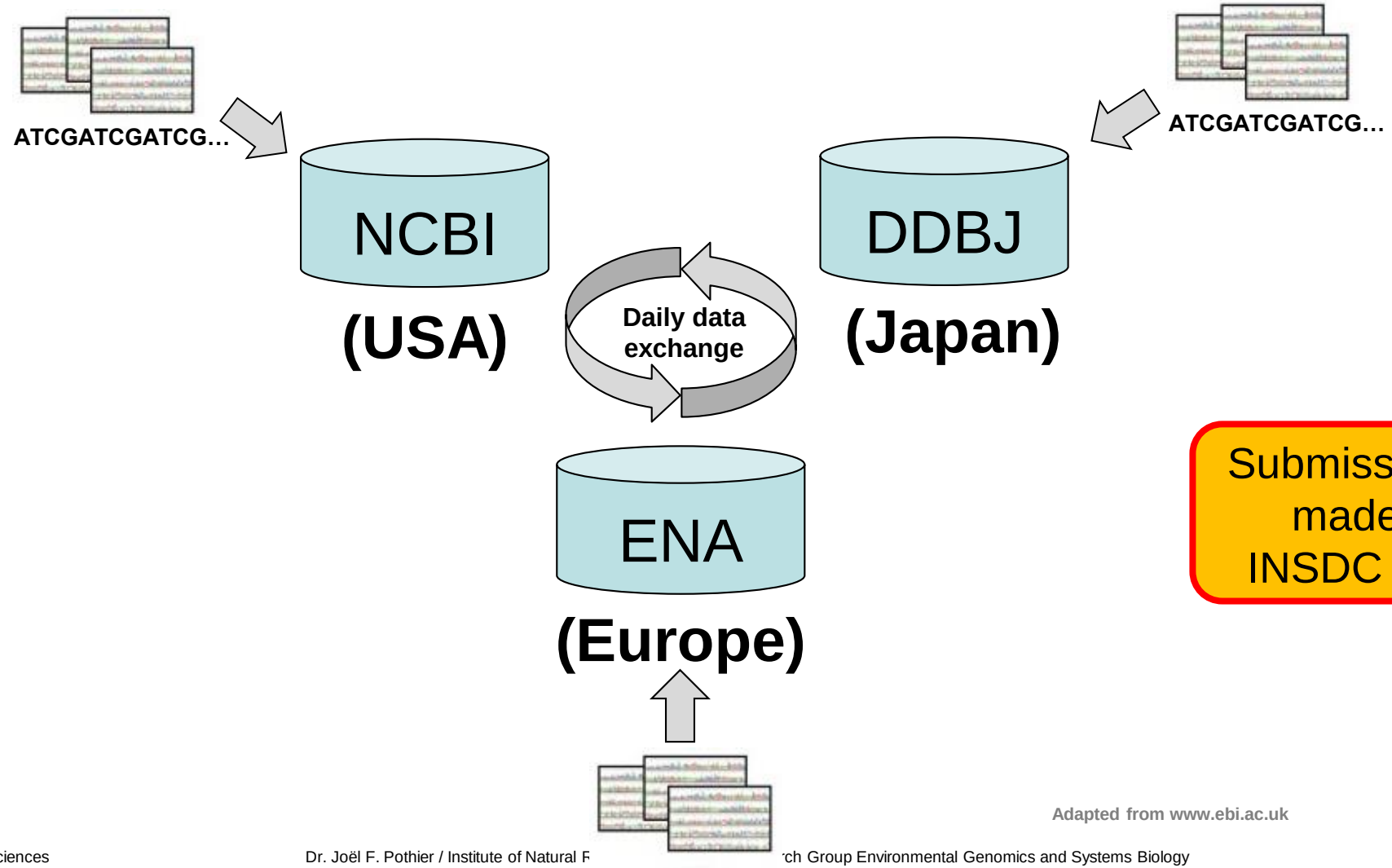


Adapted from www.ebi.ac.uk

Zurich University of Applied Sciences

INSDC

International Nucleotide Sequence Database Collaboration



Submission can be made to <u>any</u> INSDC database

Adapted from www.ebi.ac.uk

Zurich University of Applied Sciences

INSDC

International Nucleotide Sequence Database Collaboration

Data type	DDBJ (Japan)	ENA (Europe)	NCBI (USA)
Studies	BioProject	Study	BioProject
Samples	BioSample	Sample	BioSample
Next generation reads	Sequence Read Archive	**E**uropean **N**ucleotide **A**rchive	Sequence Read Archive
Capillary reads	Trace Archive		Trace Archive
Annotated sequences	DDBJ		GenBank

Adapted from www.insdc.org

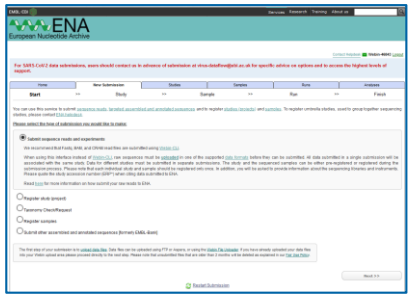
Zurich University of Applied Sciences

Data submission to ENA

Sufficient flexibility for data submission

Three routes of submissions are possible:

- **Interactive**: web forms directly filled in the browser or spreadsheets completed off-line and then uploaded
  => most accessible



- **Command line**: Webin-CLI program (.jar); validates the submission before completing them
  => maximum control

- **Programmatic**: preparation as XML documents either send to ENA (ex. with cURL) or using the Webin Submissions Portal
  => maximum control and traceability

Zurich University of Applied Sciences

Data submission to ENA

Sufficient flexibility for data submission

BUT depending on the data more than one route may be required:

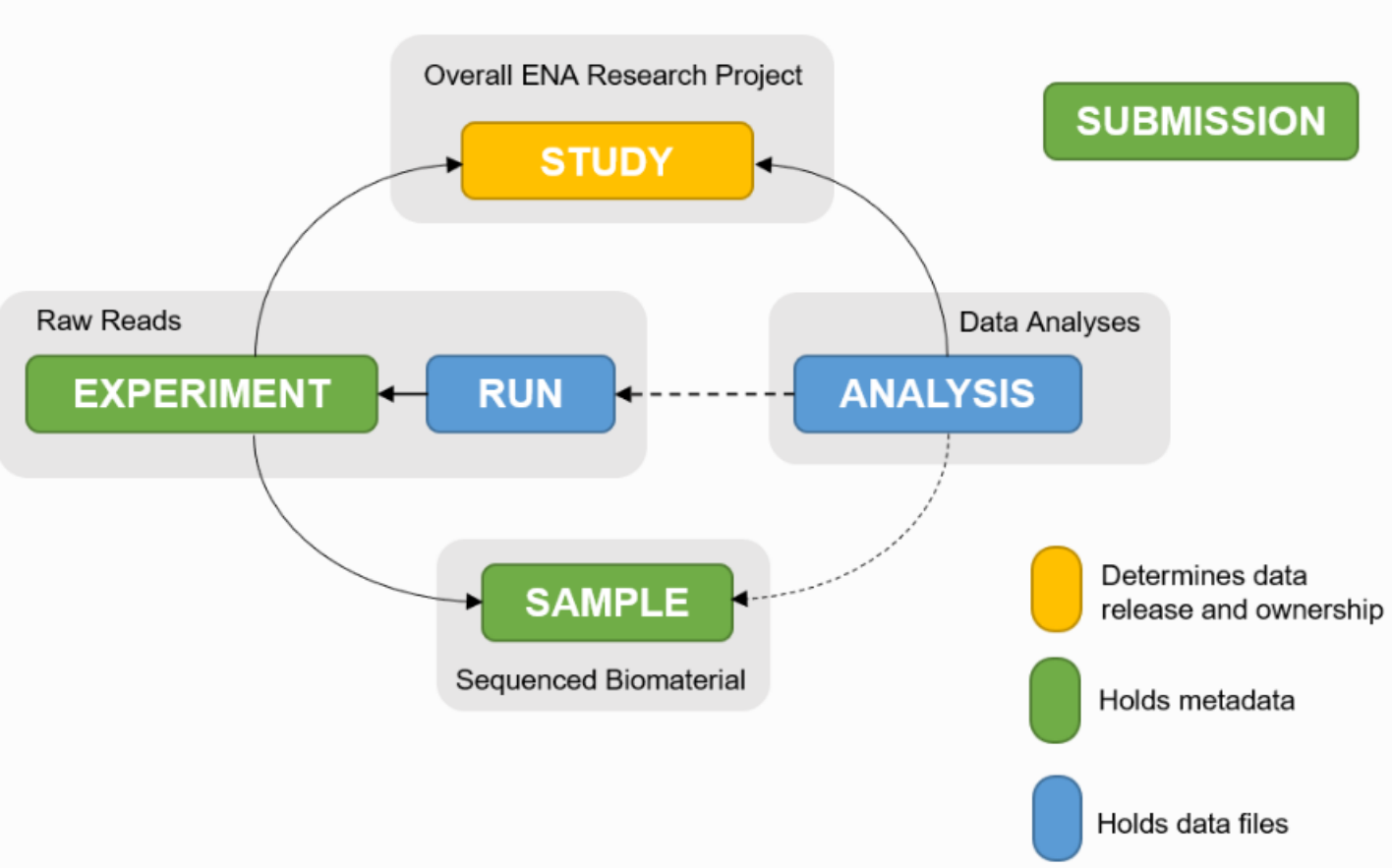
Data type	Interactive	Command line	Programmatic
Study	Y	N	Y
Sample	Y	N	Y
Read data	Y	Y	Y
Genome assembly	N	Y	N
Transcriptome assembly	N	Y	N
Template sequence	Y	Y	Y
Other analyses	N	N	Y

Adapted from www.ebi.ac.uk

Zurich University of Applied Sciences

Data submission to ENA

ENA metadata model and relationships between object types



Adapted from www.ebi.ac.uk

Zurich University of Applied Sciences

Remaining challenges with OMICS data

… and the added value of the OLOS-archiving solution

- Only the **original submitter** can modify a submission
- **Checklists** do not always work for all type of samples
- **Metadata standardization** has improved but some flexibility is still required
- **Metadata formatting** is highly relevant but not always obvious at first
- **Information on software** version used but absence of:
  - information on the settings used/changed
  - the version of the dependencies
  - possibility to provide informative log files, benchmarks or relevant scripts
  - etc.

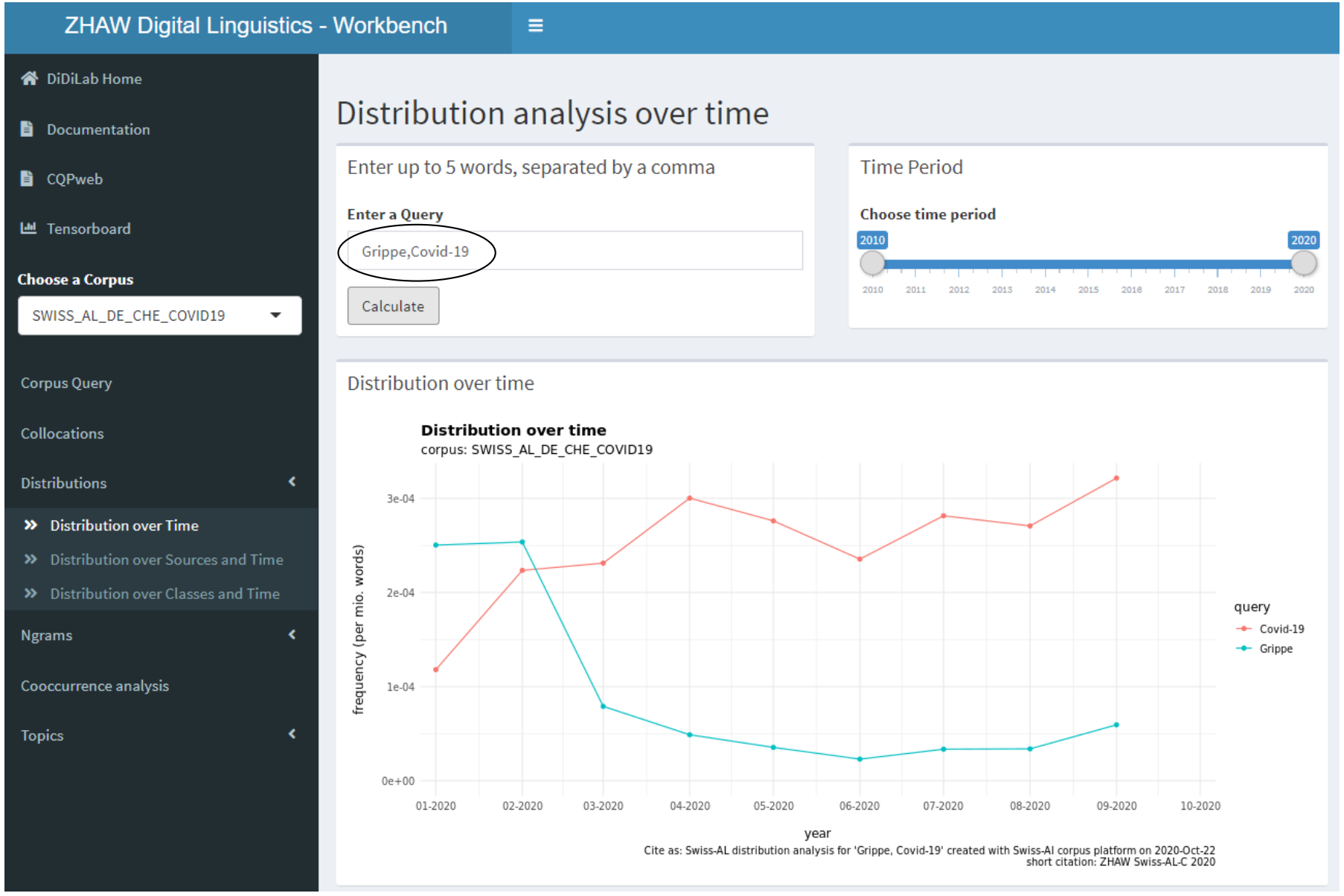
ORD-pilots

Repositories

	Description of Research Data							
Departement	**Observational**	**Experimental**	**Simulation**	**Derived**	**Reference**	**Digitalisation**	**Description (published data only)**	**Chosen Repositories**
Architecture, Design and Civil Engineering	(I)			(x)	(x)	**x**	Digitized physical architectural models	DaSCH
Health Professions	**S** (I)						Survey	Harvard Dataverse
Applied **Linguistics**				**x**	**x**		Text-Data (XML, raw)	ZHAW Swiss-AL
Life Sciences and Facility Management		**x**					Genome sequence data	ENA database, GenBank (NCBI)
Applied **Psychology**	**S**						Survey	FORSbase
	S						Survey	
Social Work	**I**						Interviews	FORSbase
Engineering		**x**	(x)	(x)			Tomography data	zenodo
	x	**x**	**x**	**x**	(x)		Survey Code/Software Tomography data	Zenodo, Mendeley Data
Management and Law	**S**						Survey	FORSbase

ORD-pilots

*ZHAW Digital Linguistics – Workbench** (https://swiss-al.linguistik.zhaw.ch)



* under development

Key findings from the pilots

- **Discipline-specific repositories should be preferred** over generic repositories

  → Offer a **better discovering** (e.g. data pre-view, visualisations)

  → **Closer to community**

- It needs **more of …data management, …standards, …processes, …support …to make data publication a success!**

Outlook (…and what you will find in the paper)

- Used **evaluation criteria for discipline specific repositories**
- Data **processing workflows** (e.g. anonymization, interview transcription)
- **Considerations for the reuse of data** in research and education
- **Impact of data publication** (e.g. number of views, downloads, requests)

Implications from the pilots

Implications at 3 levels:

- for researchers
- for institutions
- national level

Implications from the pilots

for researchers («messages from researchers to researchers»)

- **Take yourself time for a best practice in Research Data Management (RDM)**

  → Plan RDM, including costs (partly funded)

> The SNSF is aware that it takes time and money to ensure adequate data management. Therefore it allows applicants to request funds for data upload (but not download), data preparation and validation (data stewardship). The SNSF may allocate up to CHF 10,000 for these activities.

Source: SNSF

- **Engage in your community to set or use standards**

  → Possible starting points: colleagues, conferences, project consortia, or:

DCC

Because good research needs good data

Digital Curation Centre
www.dcc.ac.uk

RDA RESEARCH DATA ALLIANCE

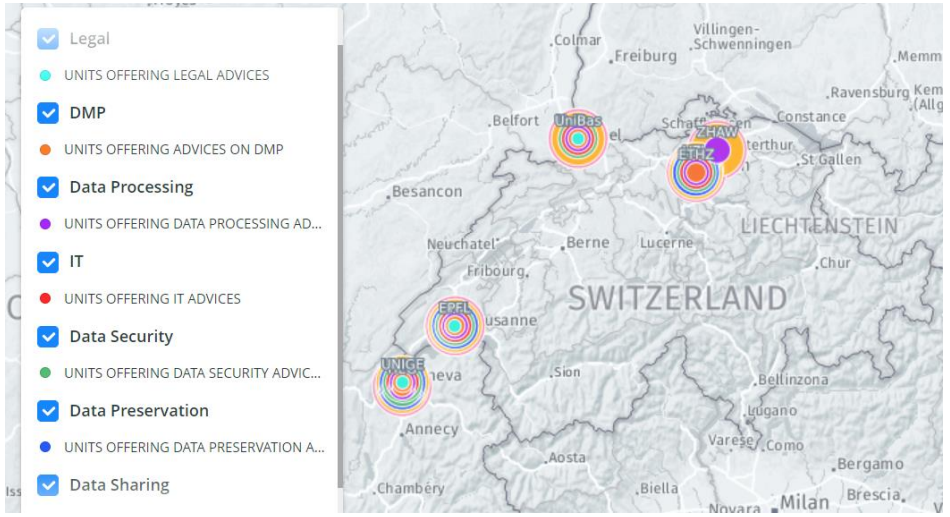
www.rd-alliance.org

Implications from the pilots

for researchers («messages from researchers to researchers»)

- **Use open source** or **open format tools for (Active) Research Data Management (they raise the value of data!)**
  - → such as: Electronical Laboratory Notebooks (ELN), scripting languages and version controlling software

- **Check out support, training and consulting at your institution or national networks (e.g. dlcm.ch)**



Source: DLCM

Implications from the pilots

for institutions

Zurich University of Applied Sciences

No advices here, at ZHAW we do the following:

- Open science / open research data **policy** (implemented in the general ZHAW R&D-policy)
- **New cross-organizational unit* «ZHAW Services Research Data»** as **central contact point** and for **local support**
- Implementing a **data stewardship model**, where specialists and data scientists give **hands-on support** throughout the **entire data life cycle** (complementary to existing services)
- **Build up or connect** to **tools, trainings, communities** or **services** (example: help with data anonymization, choosing discipline specific repository)
- **Make use of national support and infrastructure wherever possible**

* pooling of resources and competencies from library, ICT and Research and Development & more (e.g. legal services)

Implications from the pilots

on national level

- **Development of national infrastructure**
  (as done within P5-projects from swissuniversities)

- **Support of communities and bottom-up initiatives**
  (to foster discipline specific standards; e.g. Domain Data Protocols DDP)

- **National Coordination Desk / Swiss Research Data Alliance**

Thank you for listening!
Questions?