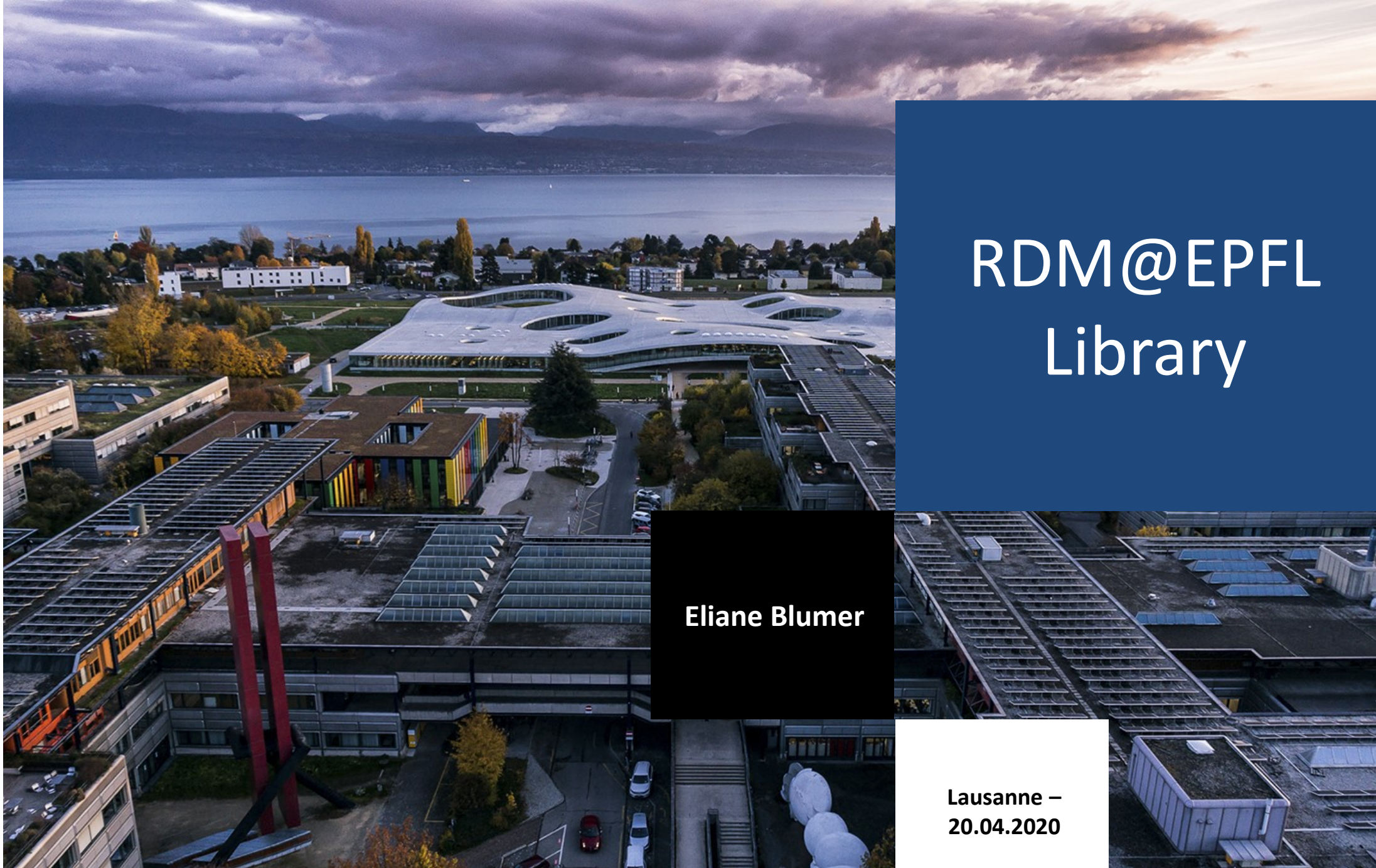


EPFL



RDM@EPFL Library

Eliane Blumer

**Lausanne –
20.04.2020**

Ecole
polytechnique
fédérale
de Lausanne

Acknowledgements

- To AILIS and CERN for having me here
- To Sitthida Samath, Alain Borel, Francesco Varrato, Antoine Masson, Alessandra Bianchi, Guilaine Baud-Vittoz, Lorenza Salvatori, Isabelle Kratz

What will I cover

History and team

Outlook

A typical day

Questions

2012-13

First reflection and two recruitments with RDM-related activities

2014

Launch of ORD Pilot in H2020

First recruitment only RDM

First contacts with EPFL Research Office

Open Data Conference

2015

Official launch of the RDM services (Research Office, VPSI)

First training offer

2016

Swiss P-2 project Data life-cycle management project (DLCM, Phase 1)

Consolidation of EPFL RDM service

- Started early
- Internal and external drivers and enablers
- Dedicated staff

2017

Martin Vetterli becomes president of EPFL, institutional importance of Open Science

Launch of new website
researchdata.epfl.ch

SNSF starts to ask for a DMPs

2018

H2020 Open Data is opened for all disciplines

First audit in a lab

Two new dedicated recruitments for RDM

2019

New EPFL website and RDM pages, NCCR RDM strategy, new training formats

EPFL Data Champions Community

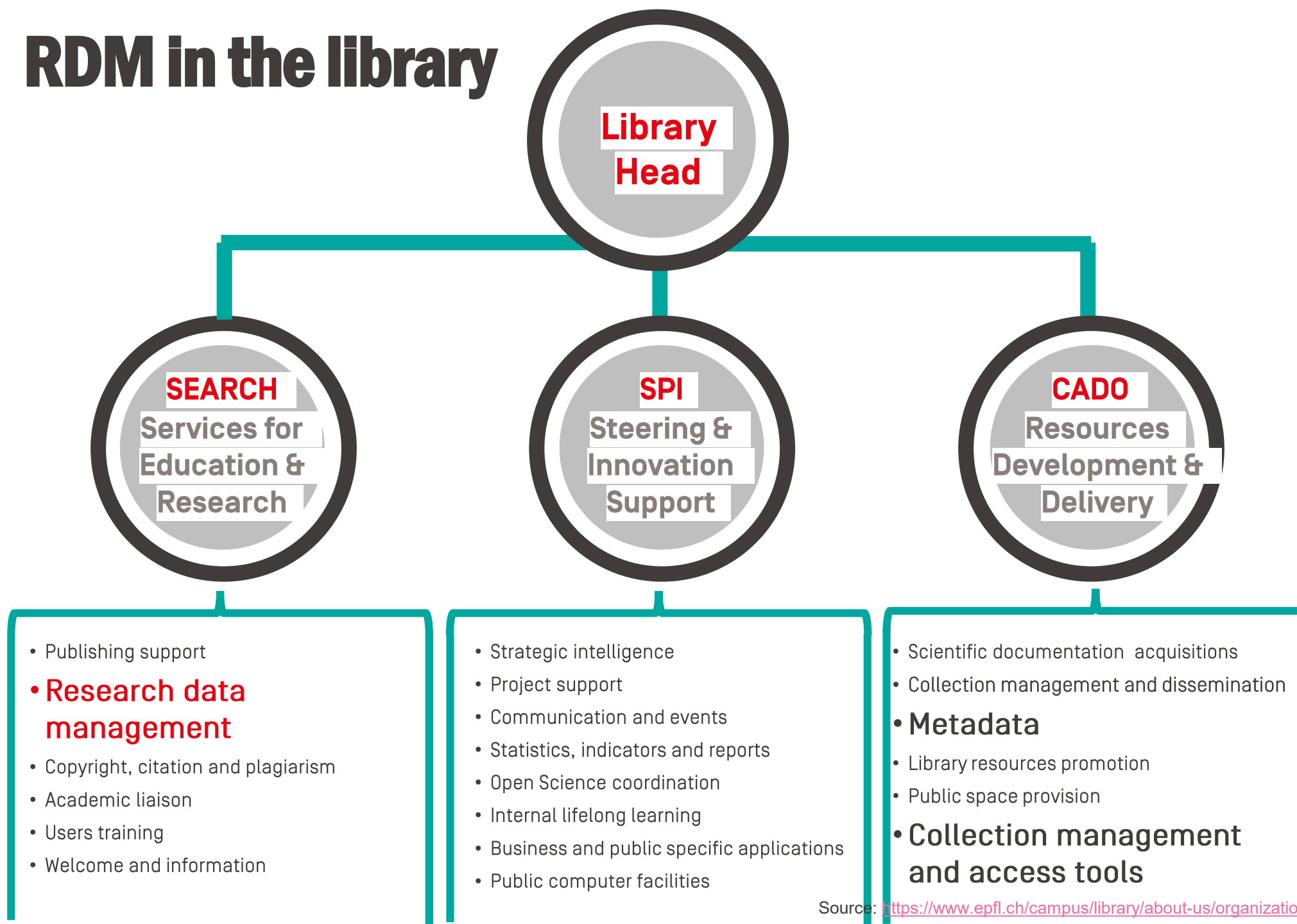
Data repository call for tender

2020

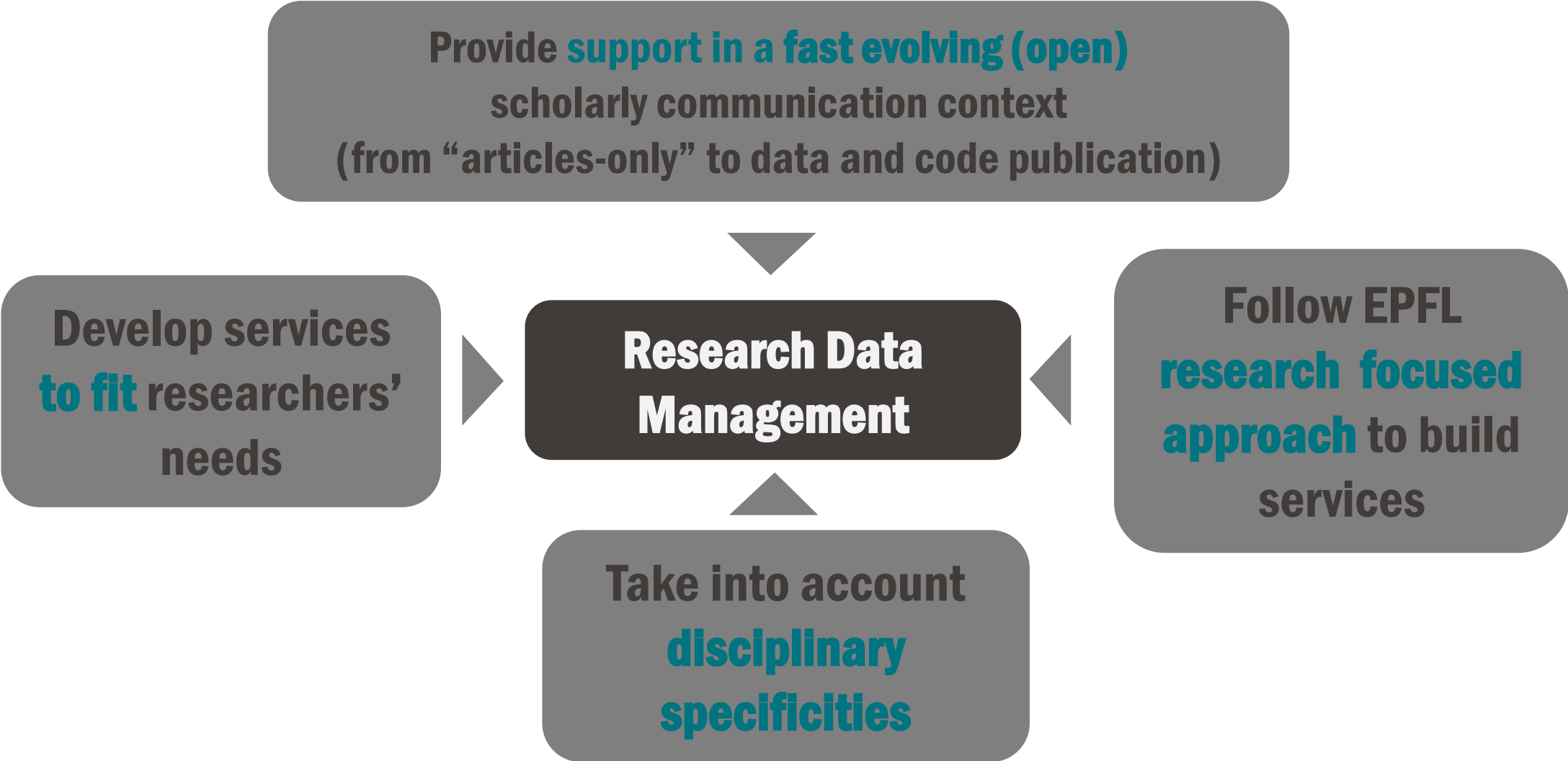
Pilot with data archive

- Continue to improve and adapt
- Communication and collaboration
- Infrastructure as well as HR

RDM in the library



RDM values



The team

Alain Borel

RDM technical support expert (0,3 FTE)
Expertise: technical systems expert, APIs, technical development, maDMP, informatics, digital archiving
Background: PhD, Former researcher in chemistry, Librarian

Eliane Blumer

RDM team coordination, user support and training (0,5 FTE)
Expertise: research data management training, digital archiving
Background: Information specialist + information systems specialist

Sitthida Samath

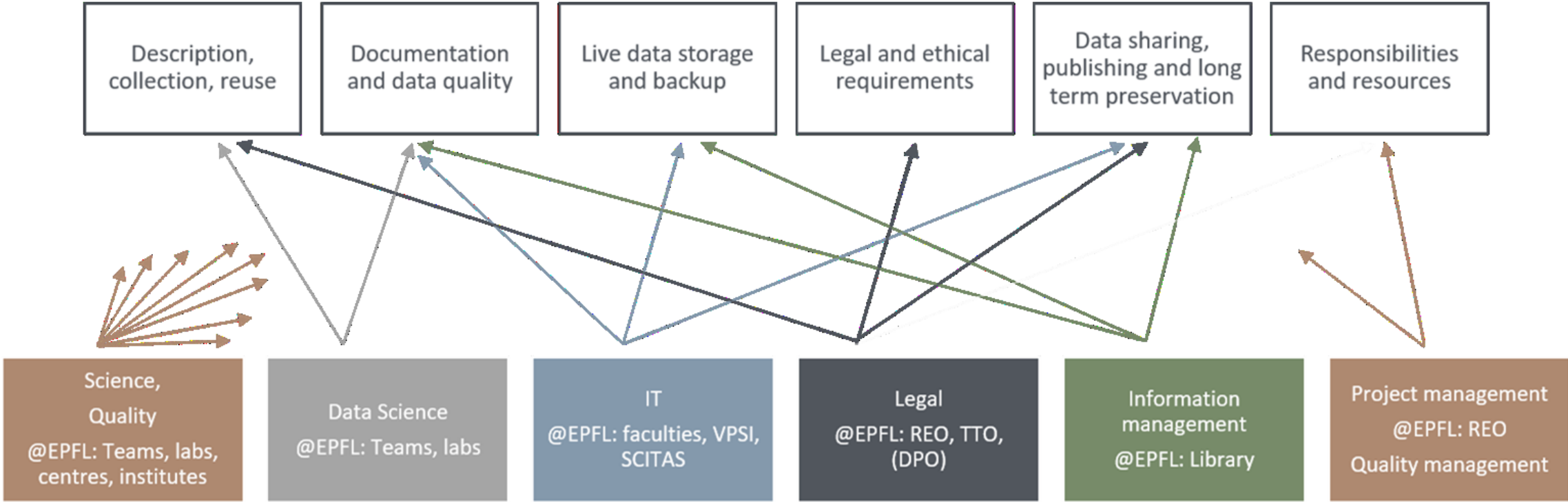
RDM metadata expert (0,3 FTE)
Expertise: metadata, ELN/LIMS, maDMP, statistics, mathematics, linguistics, informatics, training
Background: Information system specialist, computational linguistics

Francesco Varrato

RDM user support and training (0,9 FTE)
Expertise: training, community management, service improvement, statistics, physics, mathematics, user experience
Background: PhD, Former Researcher in Computational Physics

Antoine Masson

RDM technical support expert (0,7 FTE)
Expertise: ELN/LIMS, technical development, coding, open source, engineering, physics, mathematics, training
Background: PhD, Former Researcher in Physics



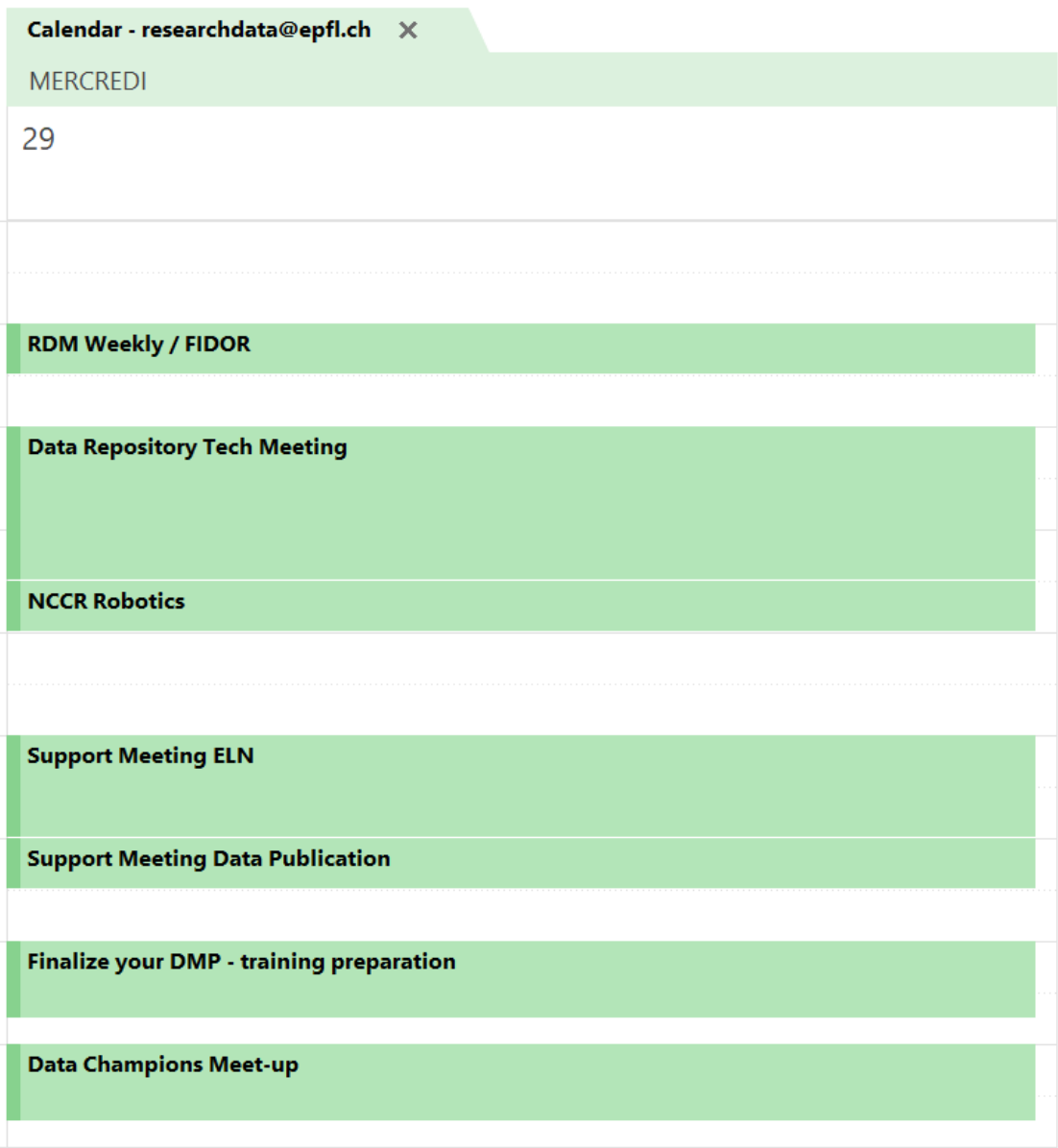
+ EPFL Data Champions

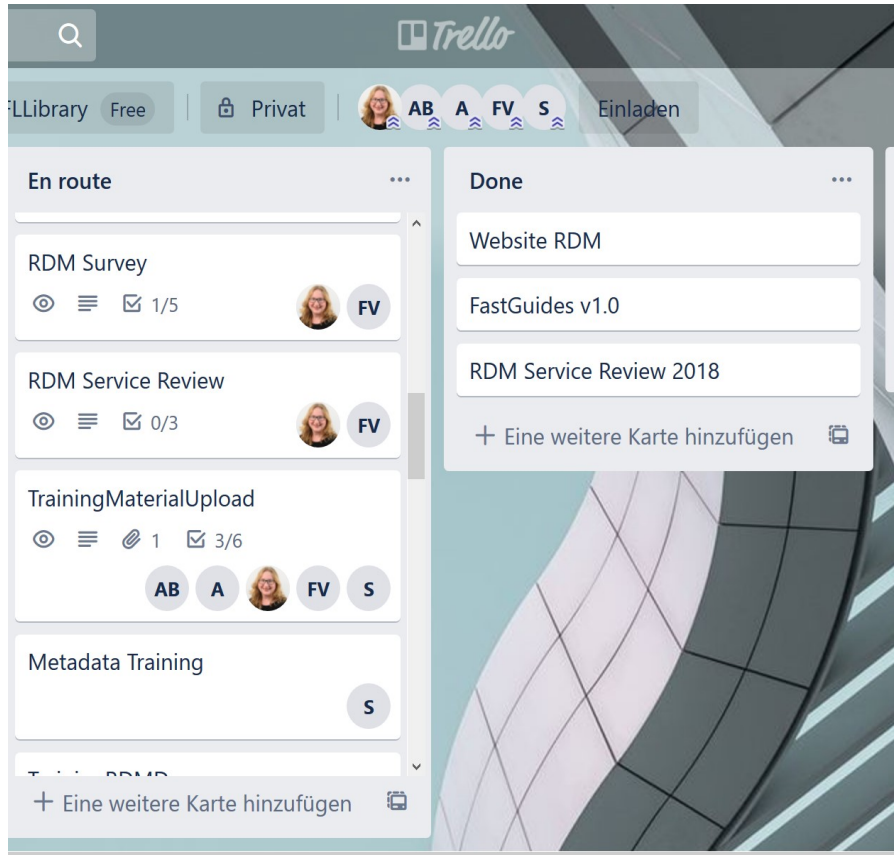
go.epfl.ch/datachampions

Credit: SFP Training, F. Varrato, S. Samath, EPFL Library

A typical day

◀ ▶ 29 avril 2020





- Trello or on real window at RLC
 - Important topics to discuss
 - Mails to answer on researchdata@epfl.ch
- FIDOR = Formation interne des données de recherche
 - Feedback from conferences
 - Topics from the field
- [Dedicated RDM chapters on library website](#)

10

Data Repository Tech Meeting

- Weekly meeting by the tech group of the EPFL data archive project
- Discussions about submission agreement, application profile, project planning

Home · Catalog · **Ingestion** · Retrieve · Active preservation · Documentation site · Advanced options ▾

- ✓ Ingestion sanitizer results ([Show](#))
- ✓ Preprocess and evolution results ([Show](#))
- ✓ Explorer results ([Show](#))
- ✗ Ingestion check results ([Hide](#))

It shows explanations only for results with errors.

- ✗ Ingestion checks applied at **object level** (1 check with errors)
 - ✗ Check: 1007 File name begins with the object name (1 object with errors) - This check verifies that each contained file in the object begins with the object name.
 - ✓ Check: 1003 Not empty: **Successful** - This check verifies that the object is not empty.
 - ✓ Check: 1010 Virus Scan: **Successful** - This check will scan an object for viruses. If a virus is found, the ingest job will stop
- ✓ Ingestion checks applied at **object folder level**
- ✗ Ingestion checks applied to **individual files into folders of objects** (1 check with errors)
 - ✗ Check: 3004 Maximum characters number (2 files with errors) - This check verifies that the filename number of characters does not exceed the specified maximum.
 - ✓ Check: 3003 Invalid filename characters: **Successful** - This check verifies that the filename does not contain invalid characters for the filesystem.

NCCR Robotics

- Meeting with the NCCR data manager
- RDM Strategy for the NCCR Robotics has been finalized
 - a **single roadmap** for all the NCCR members for publication of the data by an
 - **Embedded** data librarian during 5 months by
 - Analysis of needs via **survey**
- Training to be organized for implementation of this strategy

robotics⁺ Swiss National
Centre of Competence
in Research

FNRS
SWISS NATIONAL SCIENCE FOUNDATION

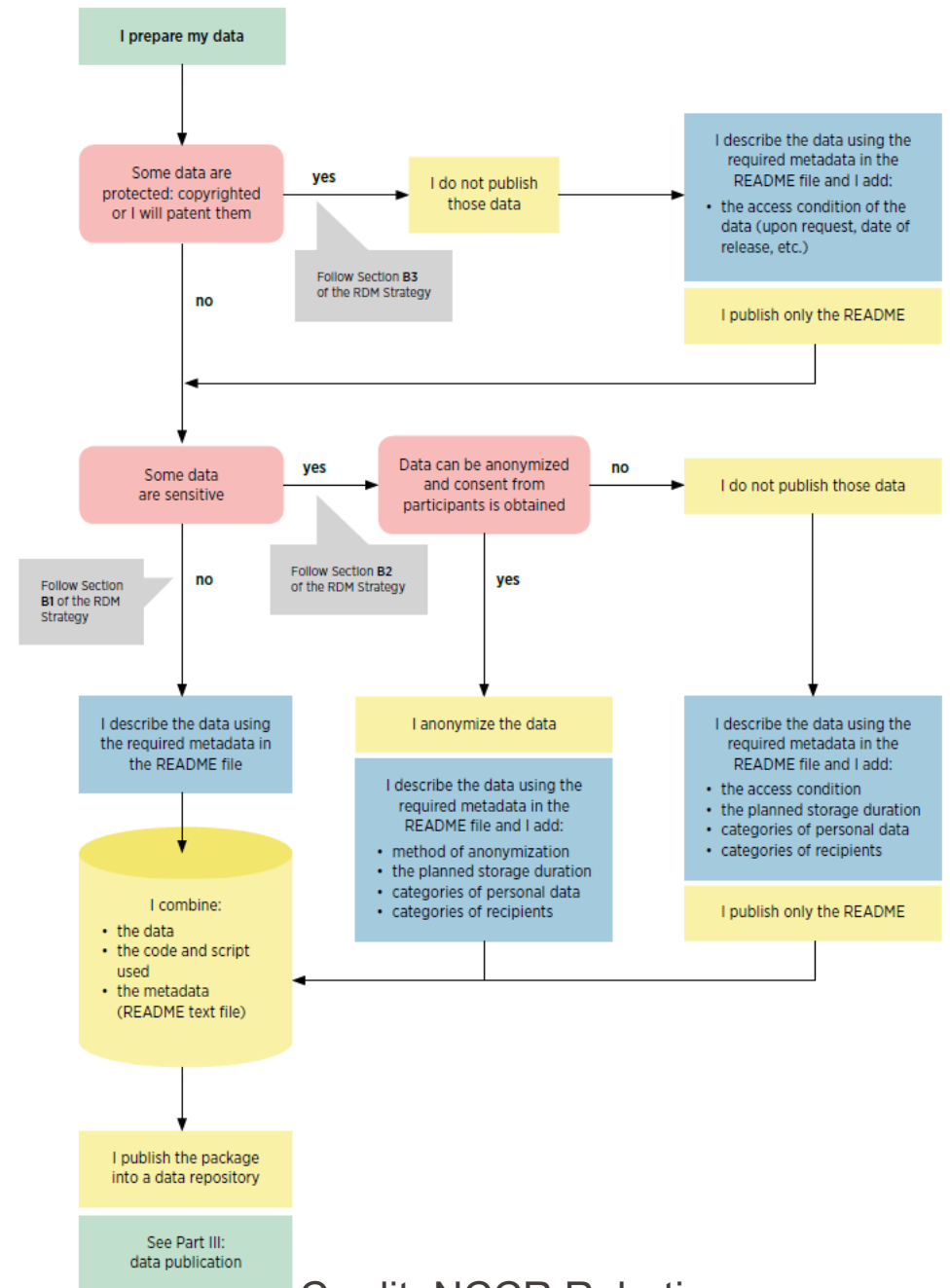
RESEARCH DATA MANAGEMENT STRATEGY



Credit: NCCR Robotics

NCCR Robotics Strategy

- extra clauses for **sensitive and for protected data**
- **Zenodo** is recommended with the creation of a community for the NCCR it's also allowed to use another repository
- **DOI** will be added on the **NCCR website**
- a **specific email address** has been created for handling data publication



13

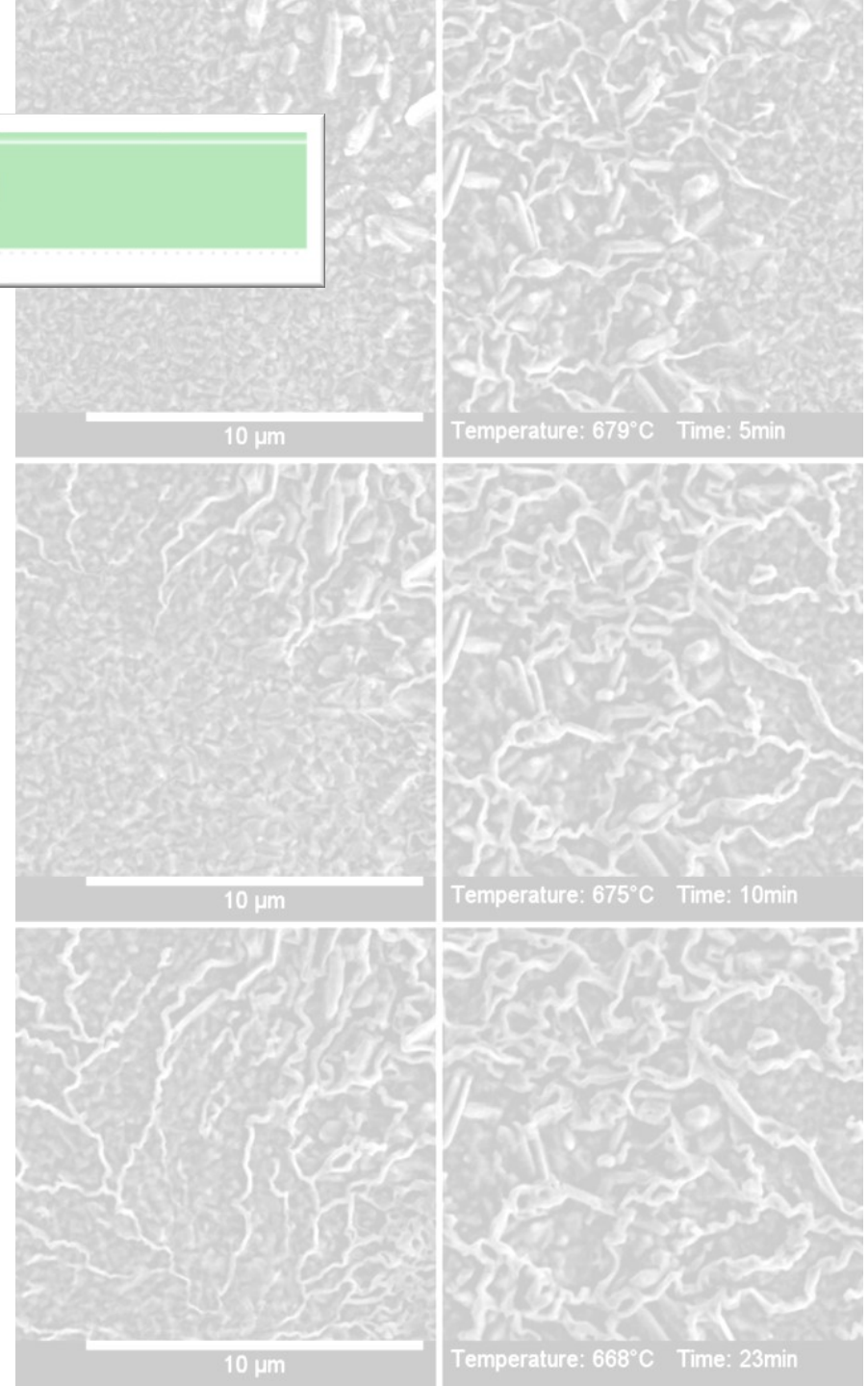
Support Meeting ELN



- From very simple to very complex
 - Area selection, background subtraction
 - Machine learning with pre- and post-processing
- On commercial softwares
 - Proprietary formats IN and sometimes IN/OUT
- On python notebooks
 - Custom code
 - Open source libraries (Hyperspy)
- On imageJ (for example)

- Researcher working in physics «typical transmission electron microscopy workflow»
 - Bulk samples, extract lamella, microscope observations
- Wants to find a tool for integrating his entire data treatment workflow
- Discussion in the team to find appropriate solutions
- Send the list of solutions



- Researcher asks for support how to publish his data
 - 1.95 Go, 87 films in 16 folders and sub-folders et sous dossiers (organization with chapters from thesis)
 - 1 script for image processing including a set of images to prove the concept
 - 3 notebooks python with test data
 - Licences?
- Demo of [Zenodo](#) and upload together with the researcher



Support in general



- **Support** covers all data-related topics, such as metadata and documentation, DMPs, ELN/LIMS, licensing (including code), preservation and ethics
- Includes not only in-person support, but also
 - Tools, such as the EPFL [Cost Calculator](#),
 - EPFL [FastGuides](#)
 - [EPFL DMP and RDM strategy templates](#)

 EPFL Library
 BIBLIO THEQUE

Cost Calculator for Data Management

Welcome to our cost calculator this tool will help researcher to have an estimate of the cost of managing, storing and publishing data.

Many providers are included in the service and you will be able to calculate a cost based on your needs.
 Total cost is calculated dynamically based on your inputs.

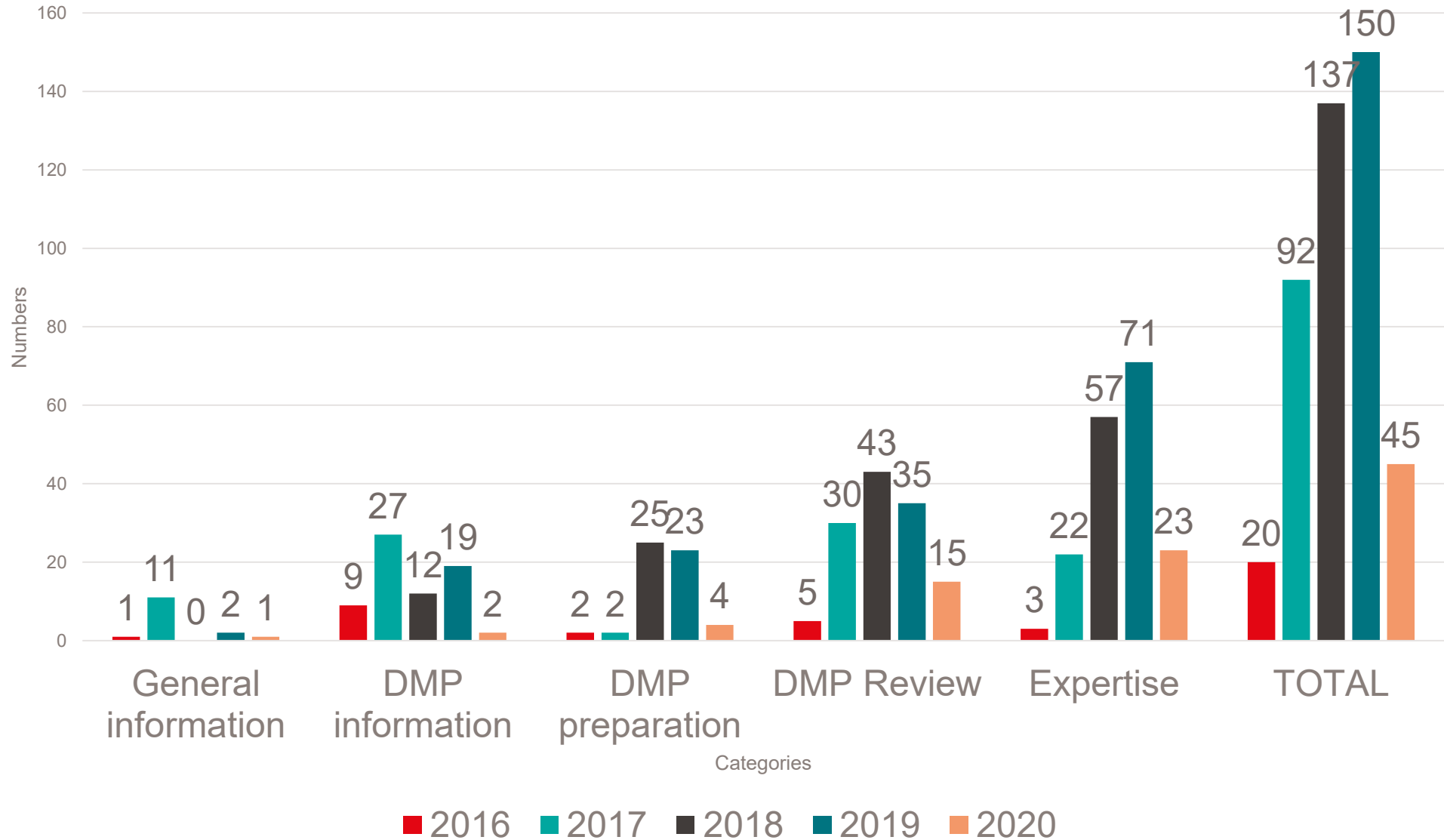
We hope you will enjoy this tool and it will be useful for you.

[To Know More \(HOWTO\)](#) [I need help with my DMP](#)

Project Name: Project Duration: Change Currency:

Line controls	Category	Provider information	Cost
<input type="text" value="1"/>	Active Storage	Select a Provider	0 CHF
<input type="text" value="1"/>	Electronic LabBook	Select a Provider	0 CHF
<input type="text" value="1"/>	Database	Select a Provider	0 CHF
<input type="text" value="1"/>	Data Repository	Select a Provider	0 CHF

Support numbers



- Prepare the objectives:
 - Finalize your data management plan according to your needs
 - Know what should happen with your DMP and linked research outputs
 - Discuss how to implement the lessons learned in research data management for your context
- Find good and bad examples
- Write to the participants for their DMPs

Venice 'time machine' project suspended amid data row

Disagreements among international partners leave plans to digitize the Italian city's history in limbo.

Davide Castelvecchi



Historians want to use archive documents to create a virtual time machine for Venice, pictured here in the 18th century. Credit: DEA/Getty

Like the city itself, an ambitious effort to digitize ten centuries' worth of documents that record the history of Venice is at risk of sinking. Two key partners have suspended the Venice Time Machine project after reaching an impasse over issues surrounding open data and methodology. The State Archive of Venice and the Swiss Federal Institute of Technology in Lausanne (EPFL) say they have had to pause data collection, and the archive's director has raised questions about the usability of the s

Source: <https://www.nature.com/articles/d41586-019-03240-w>

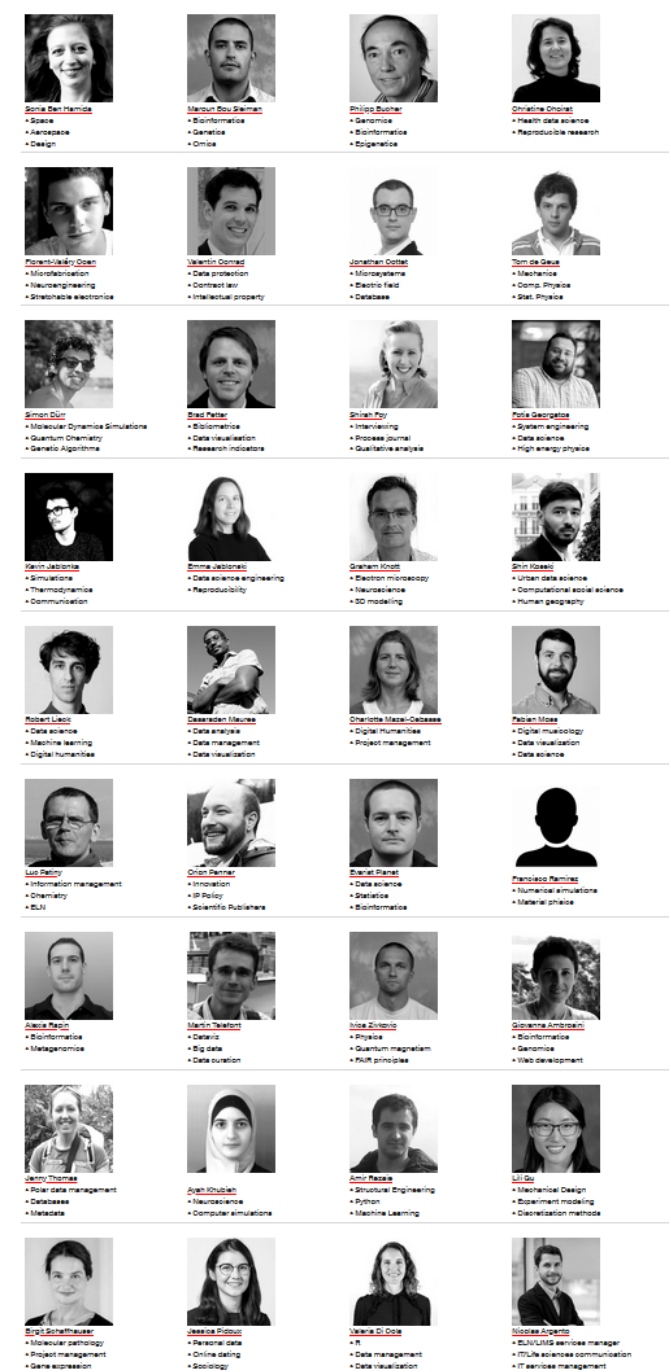
- Finalize your DMP (2h)
- RDM workshop for EPFL Staff (1 day)
- RDM workshop for EuroTech PostDoc Fellows (half a day)
- Good practices for code and data management (half a day)
- Lab data manager (1 day)
- Software Carpentry (2 days)
- Credited course RDM in chemistry (3 days)

- And **many others**... on demand or regular.

2018 : 205 participants, 14 trainings

2019 : 198 participants, 17 trainings

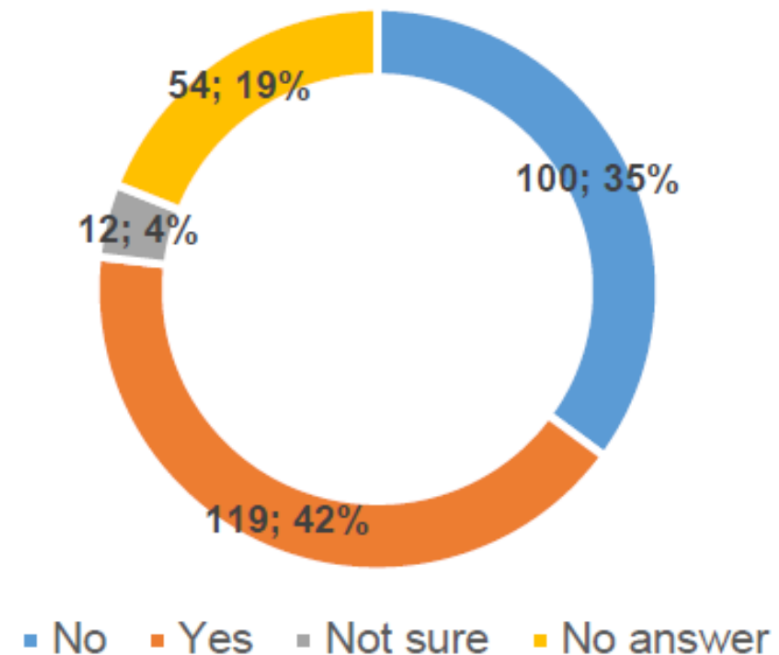
- Since 2019, 34 researchers from diverse disciplines, regular meetings
 - **Advise** researchers on data handling or redirect them to expert support on Campus
 - Act as **spokespersons** in your faculty
 - Endorse the **FAIR** (Findable, Accessible, Interoperable, and Re-usable) principles
 - Promote (or participate in) **training** or events
 - Develop or promote RDM **tools** useful for research
 - Participate to possible **publications** on RDM



Outlook

- In 2017 launched the first time, in collaboration with TU Delft (Netherlands), University of Cambridge (UK), and University of Illinois (US)
- Questions about practices and needs based on common framework
- New round in 2019

Have you published any data in the last 5 years?



Guidance for depositing data in a repository (includes choice of datasets to be deposited, choice of metadata, data preparation for depositing, link between datasets and publications) <i>(has been detailed)</i>	1 (↑4)	5
Availability of an institutional repository <i>(same question)</i>	2 (↓1)	1
Guidance for choosing a repository <i>(same question)</i>	3 (↓1)	2
Advice for storage and security issues <i>(same question)</i>	4 (-)	4
Support for data management plan implementation <i>(unclear abbreviations have been removed)</i>	5 (↑2)	7
Guidance for choosing a license to apply to your data <i>(same question)</i>	6 (↑3)	9
Advice for verifying data integrity and checking data obsolescence <i>(same question)</i>	7 (↑1)	8
Guidance for cleaning and managing data in the laboratory <i>(was “lacking “cleaning”, added as part of ELN/LIMS service reflection”)</i>	8 (↓2)	6
RDM state-of-the art in the laboratory & related recommendations <i>(replaced the question about “data curation”)</i>	9 (↓7)	2
Guidance on finding datasets and how to re-use them <i>(has been detailed)</i>	10 (-)	10
Guidance for data anonymization in case of sensitive data, commercial data or data licensed by third parties <i>(has been detailed)</i>	11 (-)	11

- Services are **systematically evaluated** and **constantly adapted** according to user feedback and their emerging needs
- Feedback survey** at the end of each year to all researchers we have been in contact for RDM

EPFL Library *Research Data Management (RDM)* support

A brief feedback (< 4 min) to help us improve the services to support the research.

* indicates mandatory questions. Answers are treated in compliance with the EU GDPR.

Has the Library RDM support brought you to some concrete *results*? *

Ex.: DMP submission, discovery of software, new data management practices, etc.

Yes
 No

Were you in contact with the Library RDM support?

How *responsive* has been the Library RDM support? *

Responsive / Fast

How *accurate* has been the Library RDM support? *

Inaccurate / Unreliable Accurate / Reliable

||

How *useful* has been the Library RDM support? *

Useful / Fruitful

||

Would you *recommend* the RDM support to a colleague? *

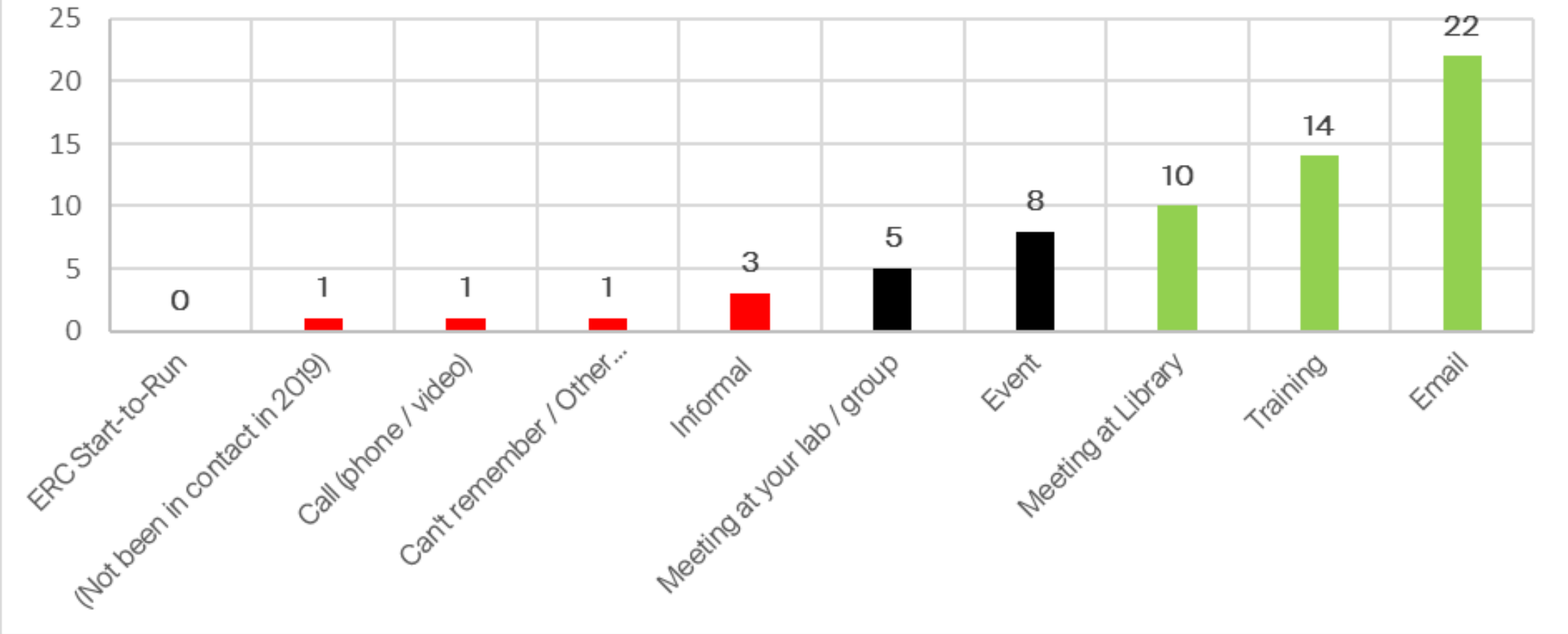
Also for subjects not directly related to writing a DMP.

No / Unlikely Likely / Yes



Service Review Results

In 2019, which kind of contact(s) did you have with the Library RDM service?



Coming soon...

- Ensure **archiving and long term preservation** of EPFL data
- **Intensifying active data management** services, i.e., ELN/LIMS, P5 EasyFair Project Renku training anonymization,
- Increasing **expert training on demand**
- Foster EPFL data and code **dissemination and reuse**
- Assist **RDM incentives promotion and Open Data awareness**, i.e. EPFL Data Policy

RESEARCH DATA MANAGEMENT

BASIC PRINCIPLES

Research data are essential components of any research project. In order to ensure the visibility, integrity and reproducibility of research, EPFL highlights the importance of sound management of data throughout their life cycle (including archiving).

Funding sources increasingly require the establishment of a Data Management Plan (DMP) as a condition for funding projects. Such a plan aims to ensure proper data management and long-term backup. In particular, this is one of the Swiss National Science Foundation's requirements for the majority of its funding programmes and one of the European Commission's requirements in the framework of the Horizon 2020 programme. Furthermore, funding source

RESPONSIBILITIES

All EPFL employees are responsible for the proper management of their own research data.

Heads of Unit have overall responsibility for the proper management of research data relating to any project in their unit.

See also chapters "[IT Security and Data Protection](#)" and "[Data Protection](#)".

LEGAL BASIS

LEX 3.3.2 "[Directive concerning research integrity](#)"



2020 Playlist

Questions?

EPFL



Thank you

Eliane Blumer

**Lausanne –
20.04.2020**

**Ecole
polytechnique
fédérale
de Lausanne**

Renku project

- RDM team participates in the Swissuniversities project EasyFAIR, one of the Swiss national projects with different HEI partners.
- Meeting with the project team
- Discuss next steps for implementing a training with Renku

RENKU

Collaborative Data Science

Renku

Reproducibility

Renku **captures the lineage** of your work so recreating a critical piece of analysis is always possible. In addition, it tracks the details of your **computational environment** so that others can reliably work with your data and code.

Shareability

Need to **share your results** with a colleague? **Make your data available** to a project partner? **Work together** on a project with a teammate? Renku makes all of these easy. And Renku lets you **track everyone's contribution** to the final result.

Federation

Collaborate across institutional boundaries, while maintaining complete control of your resources. Federation lets you **share** information, data, or code **without having to make any compromises**.

Source: <https://renkulab.io/>