

# Bottom-up Open Science shift for data and code

The case for a  
Data Champions  
community

Francesco Varrato

Research Data  
EPFL Library Team

# RDM directions

## Top-down approach

### Institutional data policies

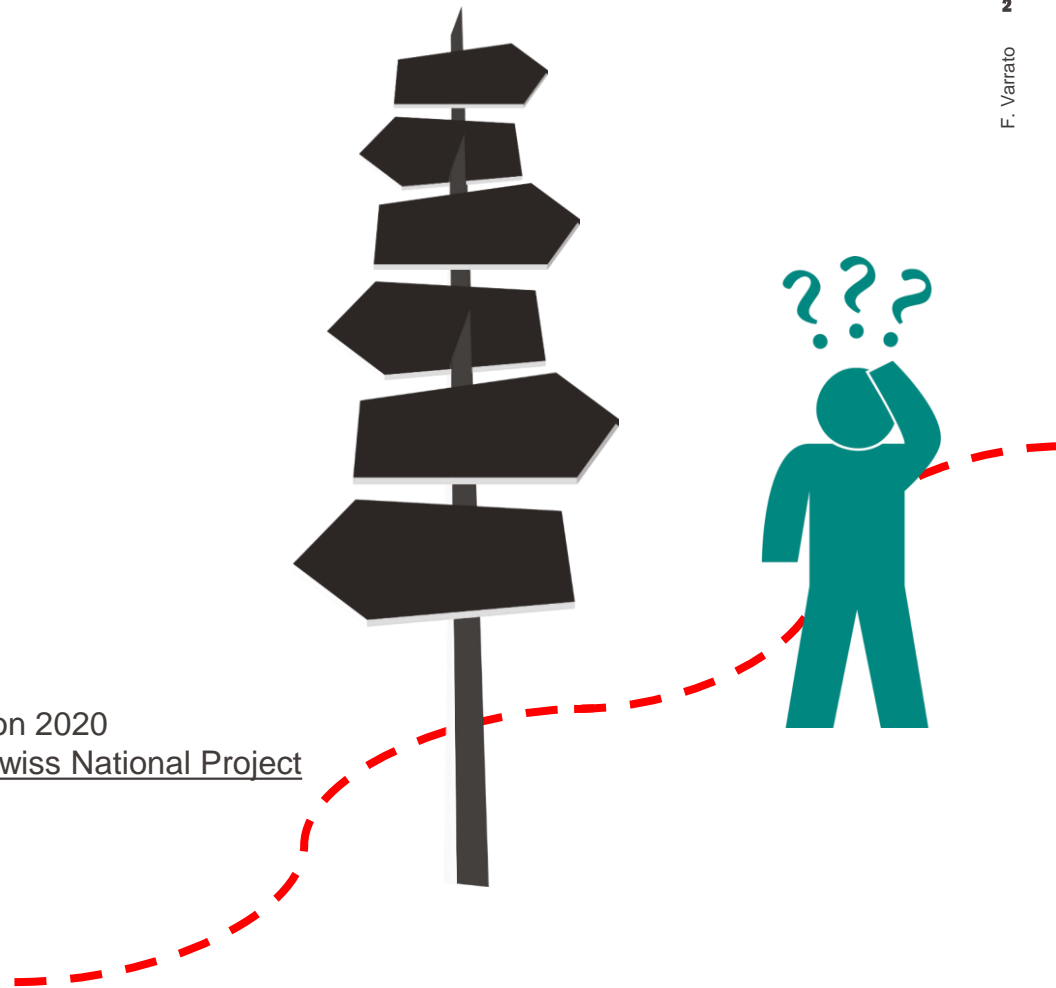
- [MIT](#)
- [UNIGE](#)
- [TU Delft](#)
- [University of Oxford](#)
- [University of Edinburgh](#)
- [University of Cambridge](#)
- [Humboldt-Universität zu Berlin](#)
- ...

### General data guidelines

- SNSF [Open Research Data policies](#)
- EC [Data Management manual](#) for Horizon 2020
- Digital Lifecycle Management (DLCM) [Swiss National Project](#)
- ...

**EPFL compliance guide** ([p.30](#))

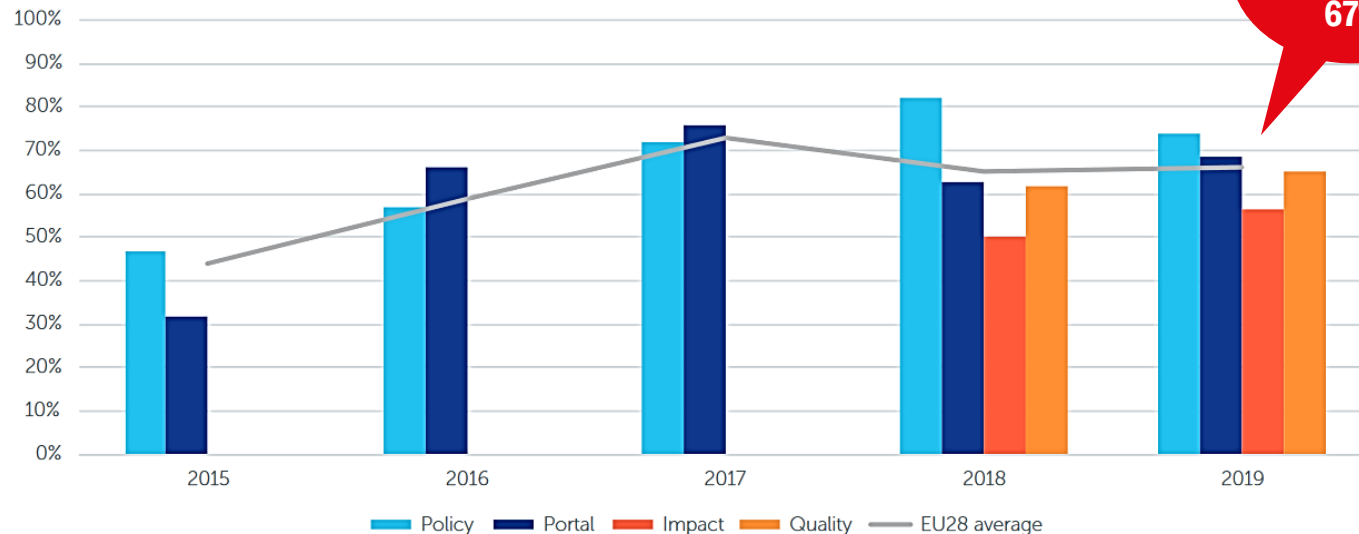
but no official Data Policy, yet!



# RDM directions

## EU28 open data maturity development 2015-2019

Room for **improvement**



But good practices **already exist** in academic research!

# RDM expertise(s)

No unique profile

Data Visualization	Digital Humanities	Bioinformatics	Data Analysis	Scientific Information Management	Data Protection	Statistical Physics
Urban Data Science	Molecular Dynamics Simulations	Data Reduction	3D Modelling	Data Science	Genetic Algorithms	Health Data Science
Computational Mechanics	Transcripts	Data Curation	Big Data	Digital Musicology	Aerospace Design	Research Indicators
Machine Learning	Computational Social Science	Reproducible Research	Bibliometrics	Qualitative Analysis	Computational Physics	... etc.

## No unique approach



### STRUCTURE

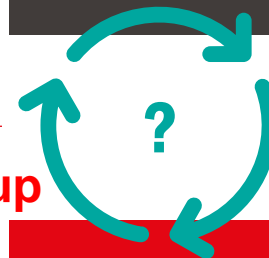
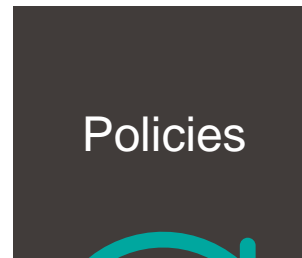
- 2 Colleges
- 5 Schools
- > 15 interdisciplinary centers
- 26 Institutes
- > 370 laboratories and research groups



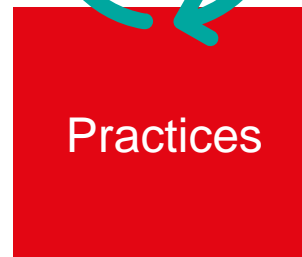
### PEOPLE

- > 8,800 students
- ~2,200 PhD students
- > 3,500 staff
- > 340 professors

Bottom-up



Top-down



# Data Champions community

## A possible approach

### Examples

- Cambridge Univ.
- TU Delft
- ...

### Volunteering

- Expertise
- Availability

### Library

- Support
- Collaboration

### Webpage

- Contacts
- Field

### What / Who

- **Interdisciplinary** community
- Good data practices **advocates**
- **Both** researchers and staff

### Shared principles

- FAIR data principles
- **Passion** for data ♥

### Mission(s)

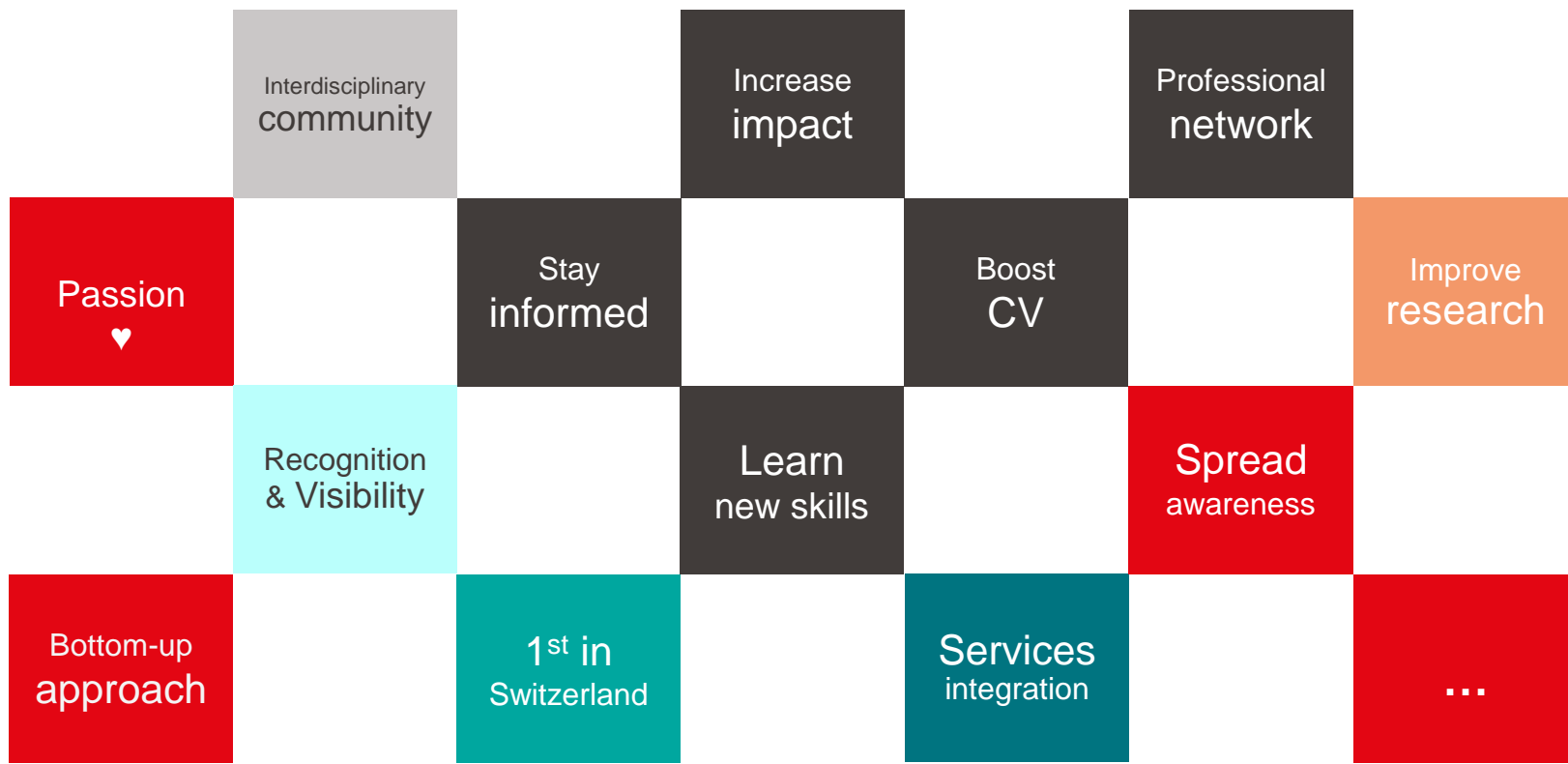
- **Advise** research community
- **Advocate** FAIR data principles
- **Promote** good data mgmt

### Platform(s)

- Webpage (Library + VPR)
- Google Drive
- Email + Slack

# Why (**top** → **bottom**)

## EPFL Data Champions





# Why (top ← bottom)

## EPFL Data Champions

To **share** my skills and **gain** new ideas

data sharing is essential to combat the **reproducibility** crisis

I am **passionate** about **rigorous methods** and the inductive / abductive process of analyzing qualitative data

I would like to **help fellow researchers** with managing, acquiring, producing and analyzing their data

Wish to help / **encourage** others in the community for a **cultural** shift

Since **similar challenges** can be expected to reoccur in different labs, it seems natural and beneficial to build a community and **share** experience

I am very much committed to **revolutionize** the **scientific community** to become truly collective

I think it is important to **share** information about data management within the School, and **train** each other

I would love to **explore new data** formats, analysis techniques, software, **best practices**

I became a scientist to **help society** by creatively **solving hard problems** [...] joining a collective, **global effort** to do so

I am **big advocate** of **data sharing** and effective data visualization as I believe it makes **science better** and **more accessible**

keen to **help** people understanding the **legal** framework

NEW  
CALLMeeting #1  
Nov. 2019**24**( **5 ♀** + **19 ♂** )  
Scientists  
PhD Students  
Specialists**12**Schools  
Institutes  
ServicesMeeting #2  
March 2020**33**( **10 ♀** + **23 ♂** )  
Scientists  
PhD Students  
Specialists**20**Schools  
Institutes  
ServicesMeeting #3  
August 2020**33**( **12** + **21** )  
Scientists  
PhD Students  
Specialists**23**Schools  
Institutes  
ServicesMeeting #4  
Nov. 2020**??**( **?? ♀** + **?? ♂** )  
Scientists  
PhD Students  
Specialists**??**Schools  
Institutes  
Services

# New DCs community 101

## 1 - Essential steps to build one



### Before

- **Draft** the roles and responsibilities
- **Ask** permissions
- **Collect** opinions in focus groups
- **Prepare** webpage and communication



### Call

- **Invite**
- **Leverage** on other DC communities
- **Leverage** on interested people
- **First meeting**: let them do the talk



### After

- **Officialize** a community coordinator
- **Organize** also informal meetings
- **Cure** webpage / tools / communication
- **Adapt & Repeat**

# New DCs community 101

## 2 - Pros & Cons



### INSTITUTION

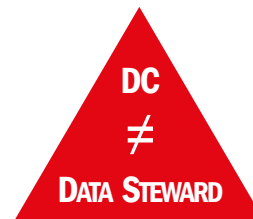
- Leverage on **motivated** persons
- Direct **influence** on practices
- Fastlane to **support** scholars

- Need **official** acknowledgment
- Need (small) dedicated **HR**
- Need **visibility**

### DATA CHAMPIONS

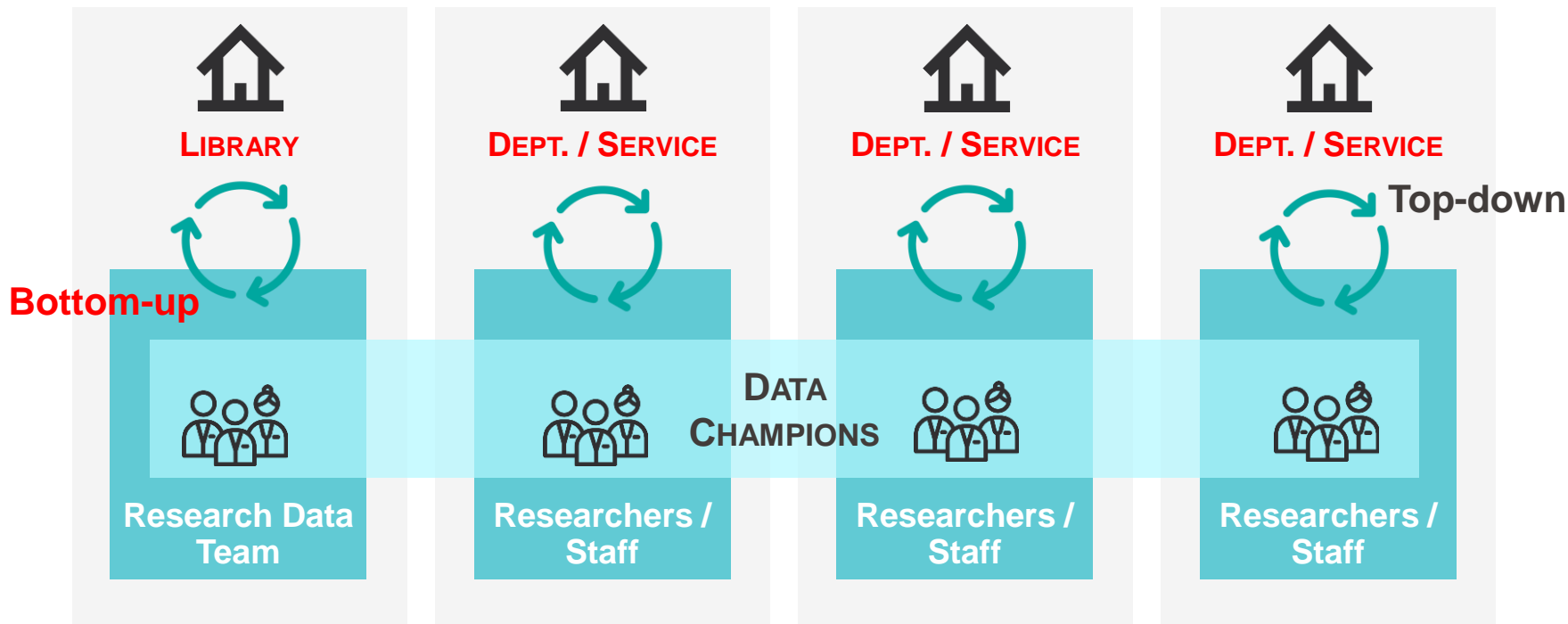
- Broader **overview** on practices
- Boost **CV** of individuals
- Professional **network**
- Improve **research**
- Stay **informed**

- Need some (little) **availability**
- Difficult to **measure** an impact



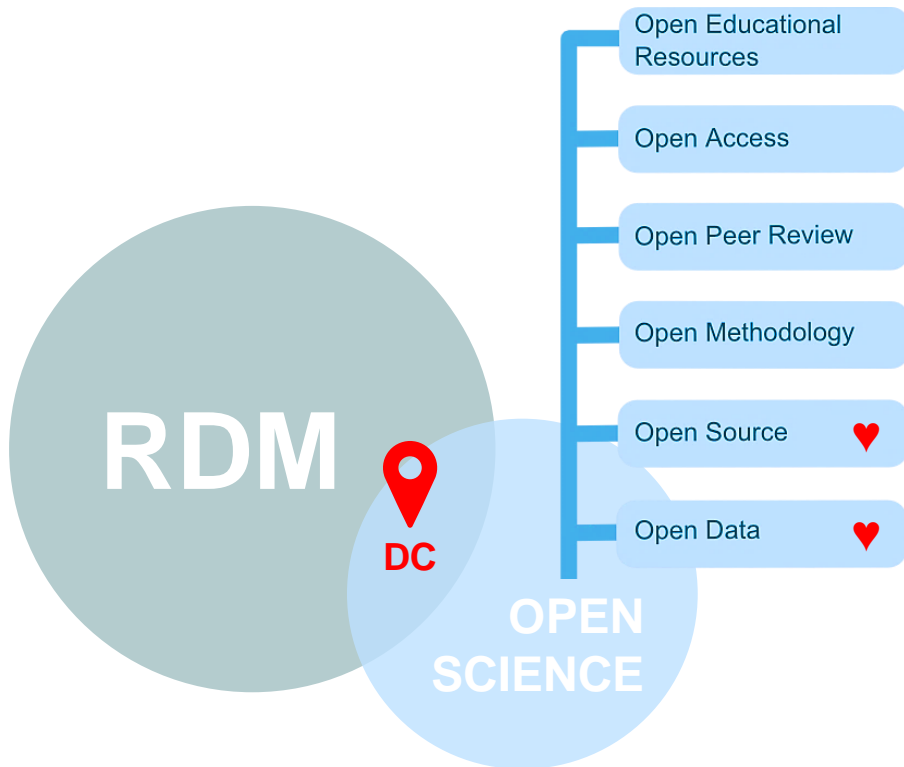
# New DCs community 101

## 3 - Synergy with central services



# New DCs community 101

## 4 - Potential in boosting Open Science



Sources: [de.wikipedia.org/wiki/Offene\\_Wissenschaft](https://de.wikipedia.org/wiki/Offene_Wissenschaft), [go.epfl.ch/datachampions](https://go.epfl.ch/datachampions)

### EPFL Data Champions

Managing data is often a fragmented, frustrating experience on the top of research activities ... But data generation, analysis, visualization, sharing, etc. greatly affect research projects. Ask an EPFL Data Champion.

Search ...



**Giovanna Ambrosini**  
 • Bioinformatics  
 • Genomics  
 • Web development



**Meroun Bou Sleiman**  
 • Bioinformatics  
 • Genetics  
 • Omics



**Philippe Bucher**  
 • Genomics  
 • Bioinformatics  
 • Epigenetics



**Christine Chirist**  
 • Health data science  
 • Reproducible research



**Florent-Valéry Coen**  
 • Microfabrication  
 • Neuroengineering  
 • Stretchable electronics



**Valentin Conrad**  
 • Data protection  
 • Contract law  
 • Intellectual property



**Valérie Di Cola**  
 • R  
 • Data management  
 • Data visualization



**Tom de Geus**  
 • Mechanics  
 • Comp. Physics  
 • Stat. Physics



**Simon Durr**  
 • Molecular Dynamics Simulations  
 • Quantum Chemistry  
 • Genetic Algorithms



**Brad Ecker**  
 • Bibliometrics  
 • Data visualization  
 • Research indicators



**Shirah Fay**  
 • Interviewing  
 • Process journal  
 • Qualitative analysis



**Estia Georgatos**  
 • System engineering  
 • Data science  
 • High energy physics



**Kevin Jablonka**  
 • Simulations  
 • Thermodynamics  
 • Communication



**Emma Jablonka**  
 • Data science engineering  
 • Reproducibility



**Graham Knott**  
 • Electron microscopy  
 • Neuroscience  
 • 3D modelling



**Shin Kosaki**  
 • Urban data science  
 • Computational social science  
 • Human geography



**Robert Liook**  
 • Data science  
 • Machine learning  
 • Digital humanities



**Maud Ehrmann**  
 • Natural Language Processing  
 • Large-scale textual data  
 • Digital Humanities



**Charlotte Mazel-Cabasse**  
 • Digital Humanities  
 • Project management



**Fabian Moss**  
 • Digital musicology  
 • Data visualization  
 • Data science

The background of the slide is a nighttime photograph of a modern building with a curved glass facade. The interior lights are visible through the glass, and the building is illuminated with a blue light. The sky is dark, and there are some light trails in the foreground.

[go.epfl.ch/datachampions](https://go.epfl.ch/datachampions)

**Thank  
you!**

francesco.varrato  
@epfl.ch

datachampions  
@epfl.ch