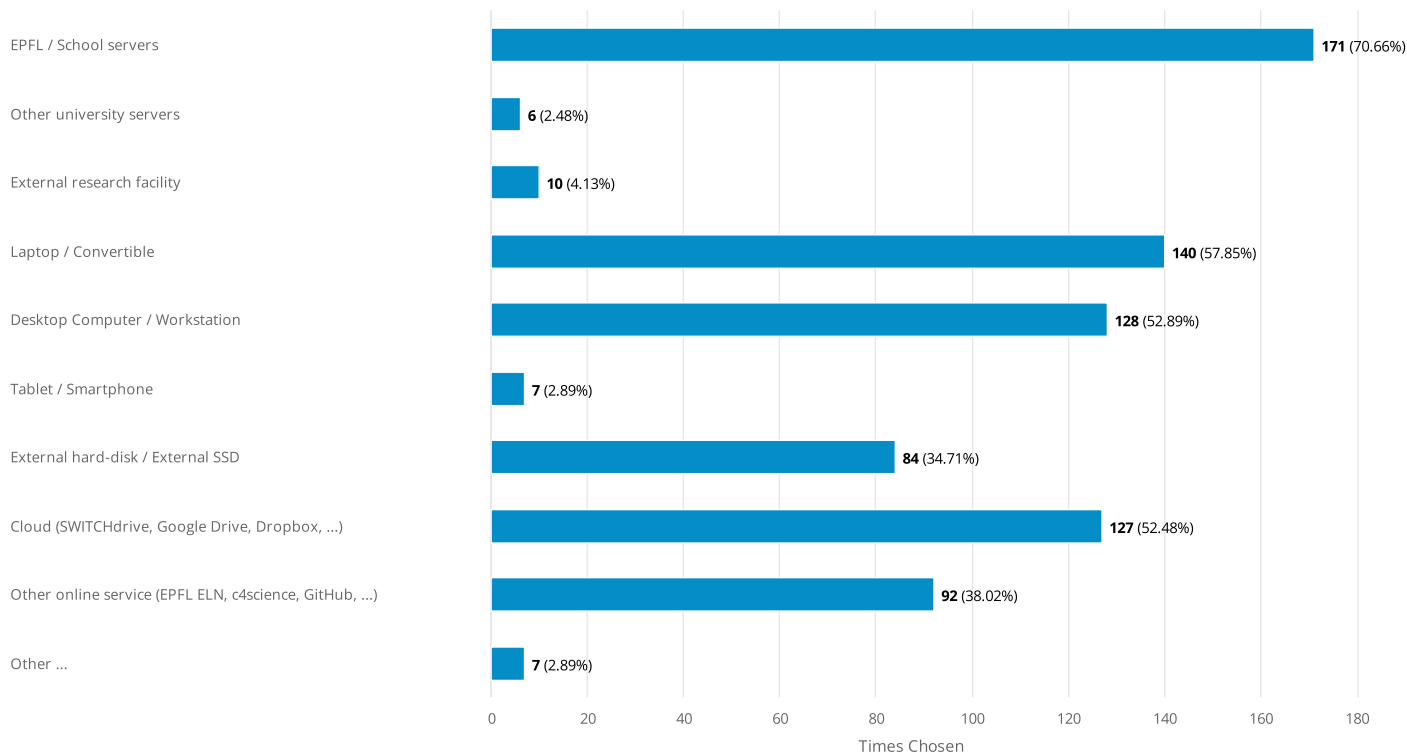


Quantitative assessment of RDM practices 2021

1. During a project, where do you store (even temporarily) your research data and code?

Number of responses: 242



"Other ..." text answers:

Paper (yes, the one you write on with a pen)

Servers of the Center where I work

Group internal NAS

<http://renkulab.io>

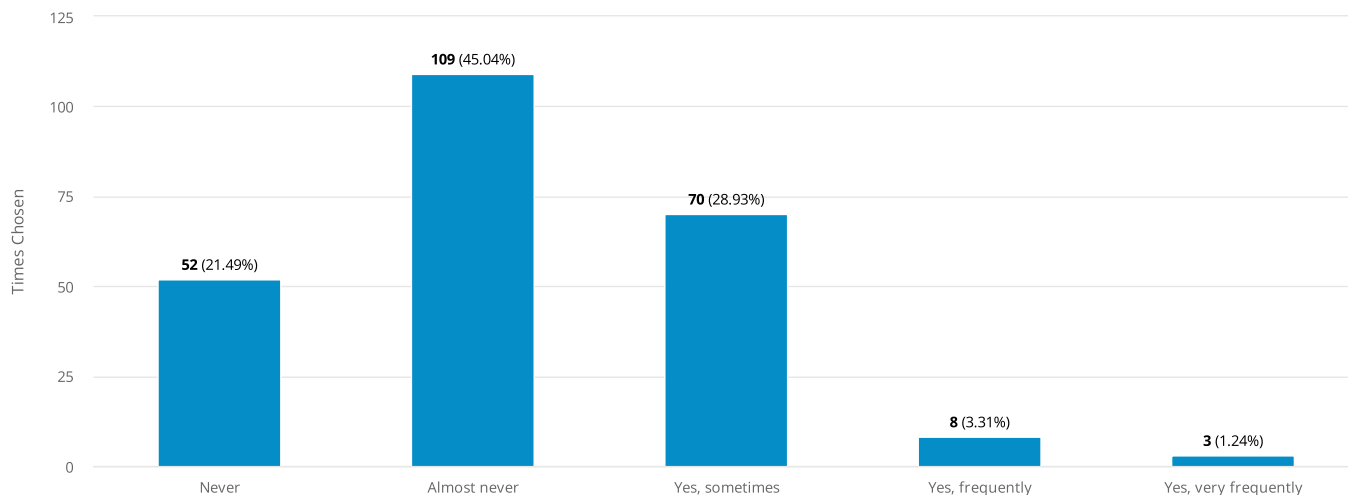
NAS

GitHub

Lab server

2. Have you ever problems finding specific file in your storage?

Number of responses: 242



2.1 Which measures have you taken (or intend to take) to overcome such findability issues?

Number of responses: 144

Text answers:

Being more disciplined and systematic; finding a way of automating the process.

Better organization

Hard to tell. There is so much data/analysis/results. I back up everything and try to document well the path to the most important results.

Better directory structure
README files that can be searched

other serve

A systematic way of naming folders based on the date of creation + creating a README file which describes shortly the content of these folders: chemical species, computationtype, specific settings,...

Name correctly the files.
Give a better structure to the folders/subfolders.

Better organization. More explicit classification of folder/files

I use git

Auto sync of important files (through SWITCHdrive for instance)

find, grep commands on Linux server

providing a common structure how research data should be stored

Don't work on the project that requires data

Proper naming, with keywords that make it easy to find.

Multiple backups

Improve my personal organization concerning data storage

Better organization of the folders and files. Deleting frequently old and unused files.

Create specific folders for tasks, topics, events etc.

Keep everything on epfl server only,
Better naming of folder
If possible create file which explain where is what but honestly this is only a dream because I dont have time

Group by project

Just restored the previous version of the folder, and I could find the file I wanted.

I try to be organized in creating folders and store almost everything on google drive and code on GitHub, such that all data is in those two places.

Precise naming of each file / folder + Precise ordering of folder

None, works fine.

git lab and new name tree structure

consitent file names and folder organisation

backup on several places, mark all data with the date, sample name etc

naming convention

write good names

create descriptive and searchable file names

More accurate records of which HDD holds which data

proper directory treeing and file naming, backup of hard drives

folder hirarchy

rethinking structure of my data folders

new reorganisation of the subfolders

Structured naming and organization of files

use the date in the file name

J'ai tellement de fichiers à organiser, je cherche la bonne méthode...

N/A

Better naming of folders for bookkeping

searching

organize in clear folders

local storage on computer and external hard disk

Writing index and/or readme files with appropriate instructions, especially for files about older projects.

using less different computers, you always forget stuff and when different version of a file exist it's hard to know which one is the latest

none

better structure to file system
more explicit file names
gitlab approach where files can be commented

Try to use a sensible file system and naming convention...

Keep track of different versions and create meaningful names

Systematic management of data

Well structured folders, main folders numbered "01_" .. "02_". Changes with life-cycle of projects. Folder-tree not deeper than 4 levels. File-names start with date (yymmdd), working-docs are stored separately from final versions
File-folder-system corresponds to mail-folders .

organizing data file systematically according to date and equipments used for measurement

Try to have an appropriate structure, placing the same type of content in the same place in a systematic way. However, it is not obvious as there are overlapping classes of content.

Using more meaningful file names

File organization and use of file indexers.

Applying more consistent naming conventions and folder structures.

Better cataloguing of data (by experiment) on workstation.

I am proficient in the use of the command line. I have all my files in Google Drive, an HDD backup, and most of my text-based research files in GitHub.

Careful naming. Use of search function

Clear file naming convention that corresponds to laboratory notebook entries.

Good memory plus, search functionality of Google Drive

We are working on a group wide standardized file naming and metadata convention.

I don't have anything planned

Relatively clear folder structure

systematically storage data by date/project

None, it can takes a little bit more to find my data, but I always find them.

Improve organization and file names.

I have a folders by year / date / project.

Classify the data in folders according to the sections of the project and use Github to synchronize code between different machines.

Consistent naming scheme to find files by search

improve organization

Good organisation in directories + meaningful naming

As it is a team/collaborative project we defined a more strict folder architecture/naming for the projects

be more organized

Better indexing and organization (a life long quest...).

adequate project foldering and dated file categorizing

Define a schematic way of naming files/data/folders

Changed the directory tree for this project

Periodically restructure my folder structures

I try to be well organized with my file names

Systematic ordering and naming of the folders, following a data management plan

Documentation

File sorting by folders, detailed naming systems/conventions, naming files based on dates, correlating file names to a log notebook of data collection

Good file-naming is a key to solve this issue

Proper structuring

None my files are well organised

Look for notebook.
Search for keywords in the folder.

Organize the storage of files and document where these storage spaces are

Clear folder structure with documentation

-backup data on EPFL server

Make some order in my folders...

Keeping proper logbooks

Revise it regularly (more or less twice a year)

Try to centralise my storage in one place and try to use a clear storage structure across my storage.
Looking for a convenient synchronised storage solution

use directories with dates

Asked colleagues to change the folders organization.

This is mainly in EPFL servers, in which searching data by name does not work well. Measures we took are organising data very rigorously.

Documentation and indexing

Naming rules for each files and folders for myself and the people I work with. Agreement at the beginning of a project (proper DMP).

Separate active storage from archive storage, adopt a robust naming convention, document folder organization

Try to enforce a systematic ordering of experiments and naming procedure to myself and to the collaborators.

Duplication

MacOS Spotlight and my personal "archiving" naming convention

data structure and file naming

order data in folders with specific names related to the content of the data

Simple folder structure and project hierarchy

Write it down in my notebook

improve use of metadata (still learning)

Organisation of data with proper names and subfolders

I have a defined file structure for all my research, including for every project.

Name them well and keep order

Versioning and deleting redundant files

Totally explicit filename, such as data_<model>_<param1>_<param2>.npz
User-friendly data format (numpy .npz)

I always found the files I was looking for in a reasonable times in the end (few minutes).
For each project, I try to keep a main document with the location reference to all the important figures and codes I might need later.

Better organized folders, and better names for files.

I searched at every file (from the more to the less relevant) ...

none, it doesn't happen that often

naming the files in a clear and specific way and putting them in an easy to understand folder structure. Add a text file with the name of every data file and for each data file a short description of the experiment for which the data were created.

Order my folders

- use MacOS indexing tools
- comprehensively name the files
- comprehensive folde/subfolder naming

Good naming, good foldering

Organize better my folders

Order better and clean regularly.

saving paper links on the notion

-

backups ...

Comment archive

keep the folders organised.

Organizing folders

Organise systematically the files, with a clear way of naming them.

order my folders with some logic

Doing specific folders for each projects and tasks of the projects

label all files starting with date YYYYMMDD
keep folder structure organised

Use a good folder structure on the server and document the data on our OneNote labbook

Better organisation

clear communication with others of projects

Know how to use git, github, and local search utilities (find, grep, etc.)

Extensive documentation where data is stored.

Use an image data base.

Better file naming and folder structure

duplication

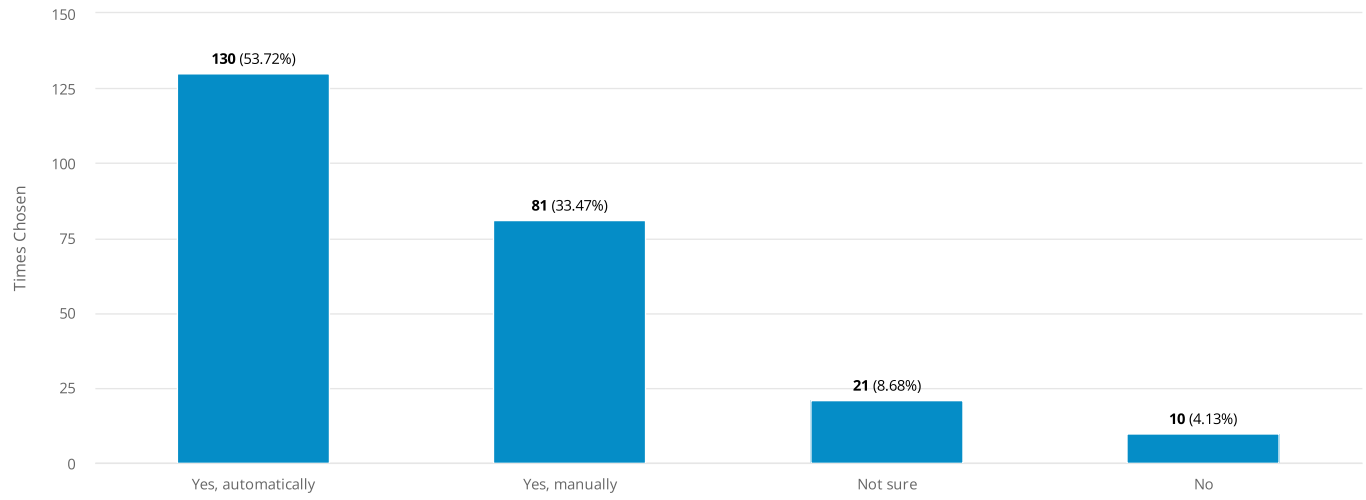
Good folder structure and documentation

I don't have such issue.

Name folders/files in a unique way. As a matter of fact, I do not have many files in my research area.

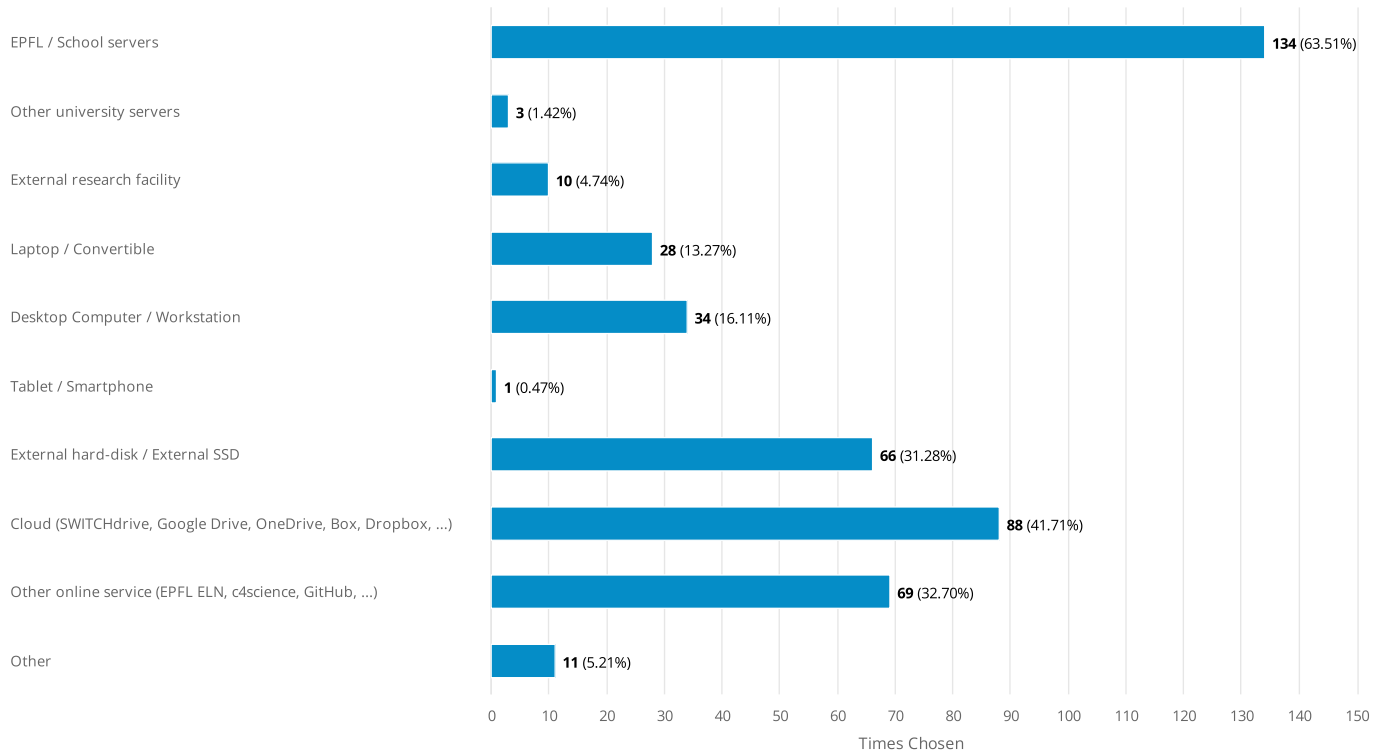
3. As of now, are your research data or code backed up?

Number of responses: 242



3.1 Where is your research data and/or code backed up?

Number of responses: 211

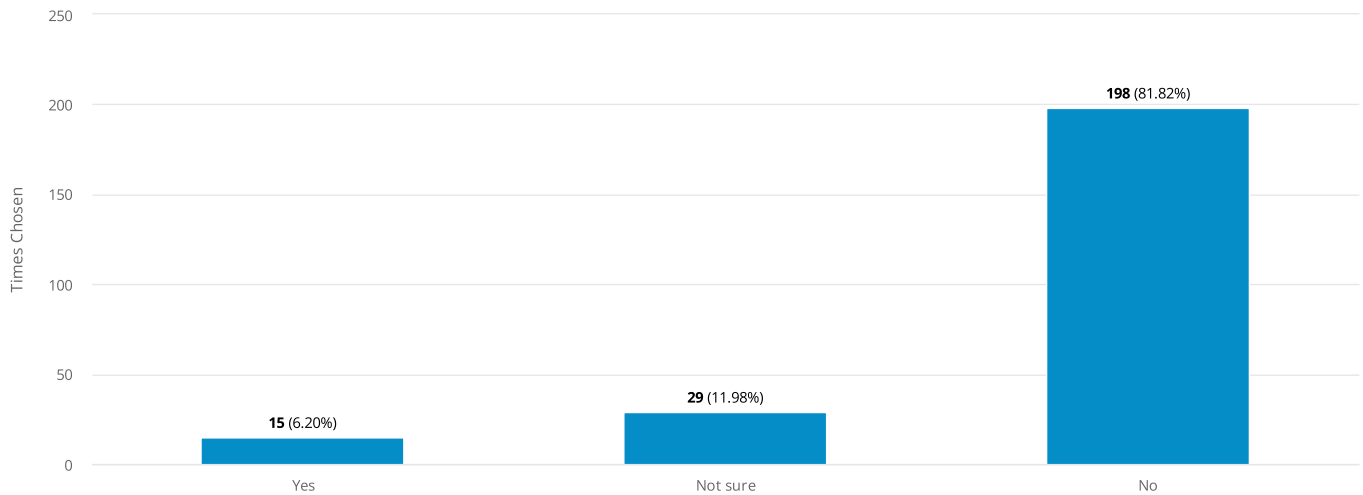


"Other" text answers:

- Lab journals in paper
- Druva
- Paper
- Servers of the Center where I work
- InSync
- EPFL S3 storage (lab server is backed up there)
- NAS
- GitHub
- Lab server
- zenodo (for open data of publications)
- gitHub

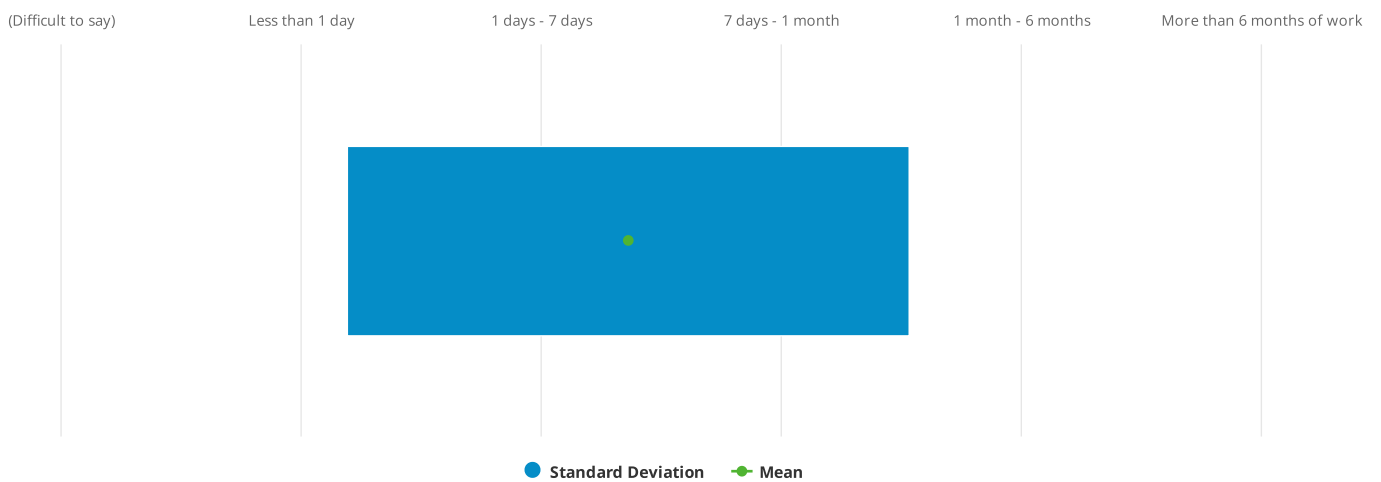
4. Have you lost any research data / code in the past two years?

Number of responses: 242



4.1 How much time did it take you initially to get these data / code?

Number of responses: 14



4.2 Could you tell what caused the loss of data / code?

Number of responses: 12

Text answers:

Accidental deleting data
Central clusters deleting idle data

Not copying the restart files from the supercomputer ontime: the scratch directory gets deleted automatically from time to time. Since this incidence, this never happened again.

Computer broke, nothing on server

i wasnt really a loss, it was a CAD model that ended havong many errors.
It only took a few days to re-do it

accidentally deleted

Mistakenly deleting directories on external HDD

accidental removal of unbacked-up data

Laptop malfunction.

Corrupt files to start with.

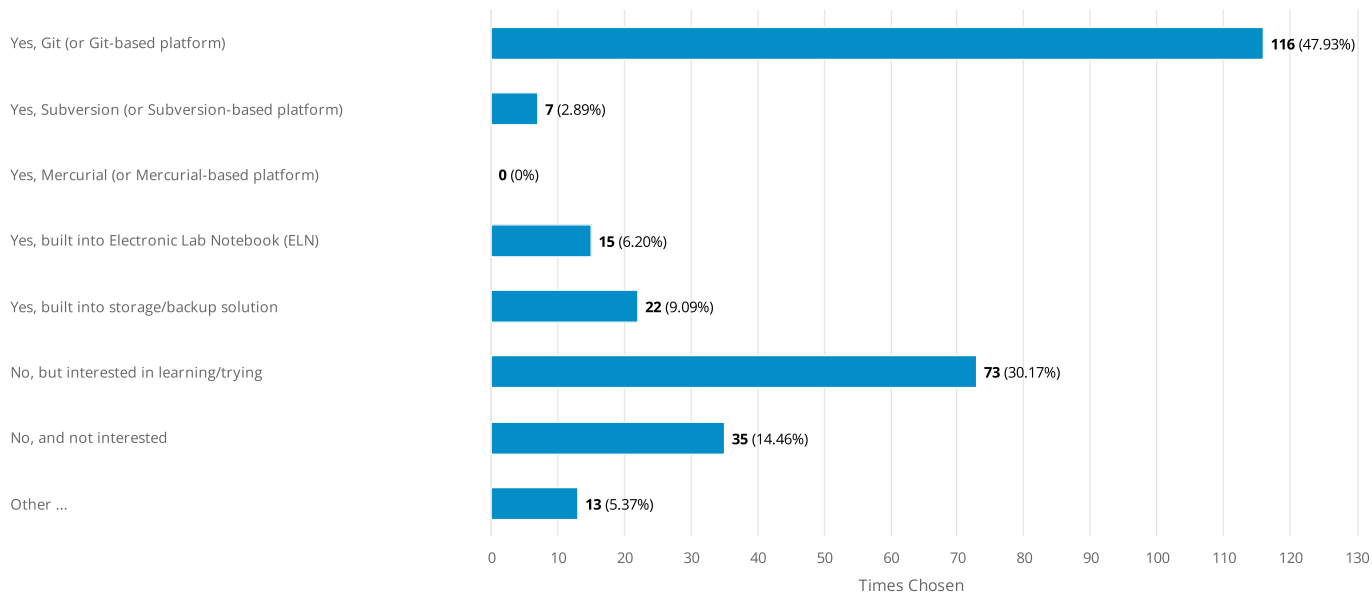
Missuse of github, oversight on my part. I didn't push data to my github cause it was too big. Then at some point I wanted to do a clean install of my repo so I deleted my local copy and cloned the github somewhere else. Obviously that deleted all data.

Lost backup

Microsoft Windows (dying/overloaded system, needed a new installation before complete sync.)

5. Do you use any tool for versioning your data or code?

Number of responses: 242



"Other ..." text answers:

partially, started also to try to implement Renku, definitely interested in further optimisation and implementation of versioning

Interested in learning about other solutions

MDS system at Swiss Plasma Center

not sure what it means...

je ne comprends pas la question

not now, but used smartgit in past

manual back-ups (R) or double-copies with date

Other ...

Not for the data.

RENU

I do not know what is meant by versioning, I do not know any of these platforms

<http://renkulab.io>

I generate updated database versions every month, the older become backups

5.1 Which Electronic Lab Notebook (ELN) do you use?

Number of responses: 13

Text answers:

workspace

EPFL ELN

Researchspace

SLIMS

SLIMS

SLIMS

One note

Elog, Oenote

Slims

slims

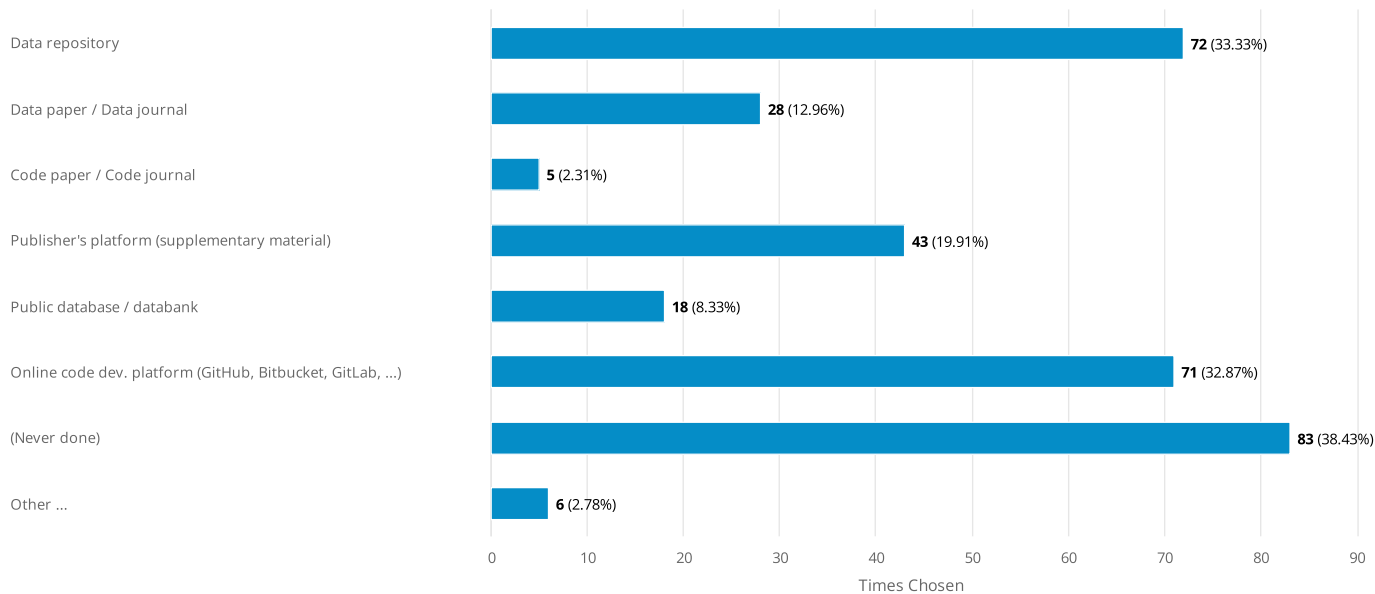
RSpace, OneNote

SLIMS

SlimS

6. Where do you publish or publicly share your research data and/or code?

Number of responses: 216



"Other ..." text answers:

And google drive for data

Platform of EU Center of Excellence for Computing Applications

Free Repositories, e.g. ArXiv, University Platforms

infoscience

I've never shared data that does not fit in a GitHub repo.

lab's space

6.1 Which tool(s)/platform(s) do you use the most to disseminate data and/or code?

Number of responses: 81

Name of tool/platform	What (Data, Code, both)
European Nucleotide Archive	Data
NCBI	genetic sequences
Github	Code
Zenodo	Data
GitHub	code
GitHub Google Drive from EPFL account	code data

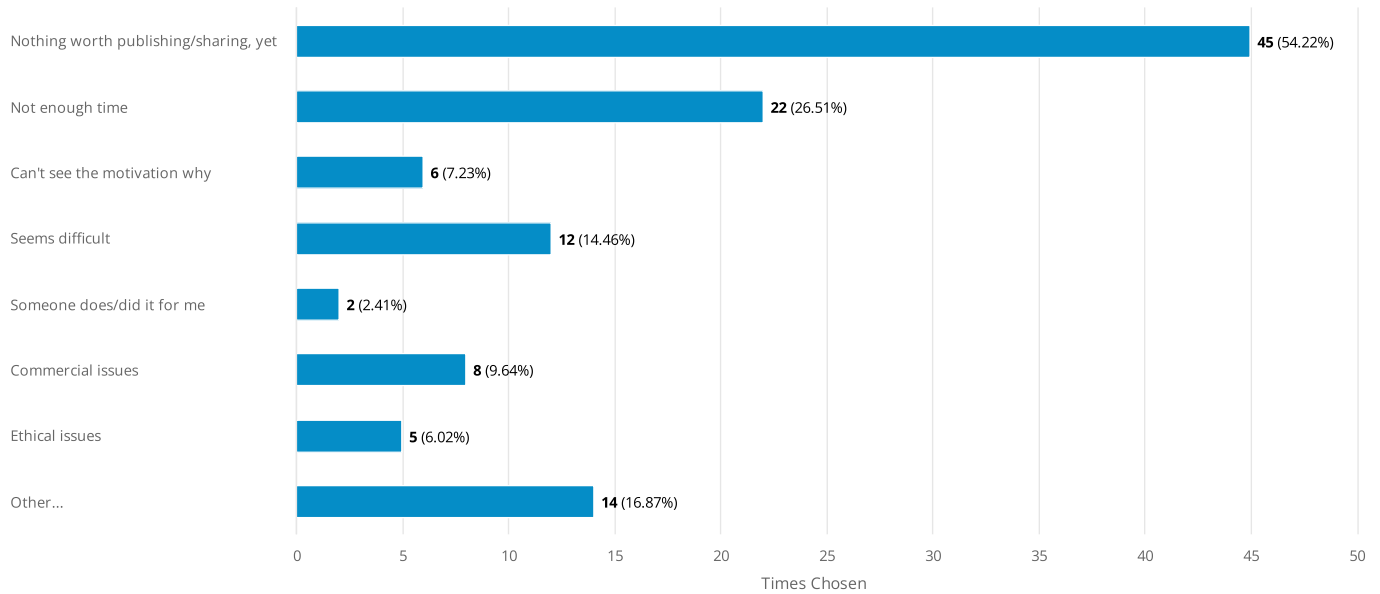
Name of tool/platform					What (Data, Code, both)				
zenodo	github	-	-	-	code	code	-	-	-
GitHub	c4Science	SwitchDrive	-	-	Code	Code	Data	-	-
GitHub	GDrive	-	-	-	code	data	-	-	-
Github	-	-	-	-	Code	-	-	-	-
Bitbucket	-	-	-	-	Code	-	-	-	-
github	zenodo	-	-	-	code	data analysis, plots, ets	-	-	-
GitHub	Zenodo	Zenodo	Infoscience	Infoscience	Code	Datasets	Papers	Datasets	Papers
Zenodo	-	-	-	-	data	-	-	-	-
materials clou	-	-	-	-	-	-	-	-	-
Hydrology and Earth Systems Science Journal (Copernicus Publishers)	Hydrological Processes journal (Wiley Publishers)	-	-	-	Data	Data	-	-	-
Bitbucket	github	-	-	-	code	code	-	-	-
c4science	gitlab	bitbucked	github	-	code	code	code	code	-
GitHub	-	-	-	-	code	-	-	-	-
GitHub	-	-	-	-	both	-	-	-	-
Gitlab, Dropbox, cernbox	-	-	-	-	Both	-	-	-	-
arXiv	-	-	-	-	papers	-	-	-	-
data repository	-	-	-	-	both	-	-	-	-
zenodo	-	-	-	-	both	-	-	-	-
frontiers	infoscience	github	-	-	both	both	both	-	-
Github	-	-	-	-	code	-	-	-	-
Zenodo	-	-	-	-	Data	-	-	-	-
Github	-	-	-	-	matlab and python	-	-	-	-
Zenodo	Github	C4science	-	-	both	code	code	-	-
c4science	zenodo	-	-	-	Code	Data	-	-	-
GitHub	Zenodo	-	-	-	mainly code; small data sets	large data sets; code versions with DOI	-	-	-

Name of tool/platform					What (Data, Code, both)				
Cambridge Structural Database (CSD)	-	-	-	-	Crystal Structure Data	-	-	-	-
GitHub	arXiv	-	-	-	Code and small datasets	Papers	-	-	-
g	-	-	-	-	r	-	-	-	-
GitHub	-	-	-	-	-	-	-	-	-
GitHub/Zenodo	-	-	-	-	both	-	-	-	-
Github	c4science	SWITCHHub	-	-	Code	Code	Data	-	-
Github	Lab website	-	-	-	Code	Data	-	-	-
zenodo	-	-	-	-	both	-	-	-	-
github	GEO	-	-	-	code	data	-	-	-
Zenodo	GitHub	-	-	-	Data	Code	-	-	-
zenodo	-	-	-	-	data	-	-	-	-
Zenodo	figshare	-	-	-	data or both	data	-	-	-
Swiss palm	journal data	-	-	-	data	excell files	-	-	-
github	materials cloud	SI	-	-	code	data	data	-	-
Zenodo	Github	c4science	-	-	data	code	both	-	-
Microsoft OneNote	Keysight Labber	Lab PC	Laptop	-	Data	Data	Data, Code	Data, Code	-
GitHub	-	-	-	-	Code	-	-	-	-
Zenodo	-	-	-	-	both	-	-	-	-
GitHub	Zenodo	-	-	-	Code	Data	-	-	-
Zenodo	-	-	-	-	data and code	-	-	-	-
c4science	Github	https://lnmcddata.bbp.epfl.ch/	-	-	code	code	data	-	-
zenodo	-	-	-	-	both	-	-	-	-
GitLab	EnviDat	-	-	-	code	data	-	-	-
GitHub	-	-	-	-	Code	-	-	-	-
Github	Infoscience	-	-	-	Code, sometimes data	Data	-	-	-
Zenodo	-	-	-	-	Both	-	-	-	-

Name of tool/platform					What (Data, Code, both)					
zenodo	GitLab	google sites	-	-	data & code	code	code	-	-	
GEO server	-	-	-	-	Data	-	-	-	-	
Mendeley	Zenodo	Github	-	-	Data	Data	Code	-	-	
Github	-	-	-	-	code	-	-	-	-	
zenodo	GitHub	-	-	-	Data	Code	-	-	-	
zenodo	-	-	-	-	data	-	-	-	-	
Github	Zenodo	-	-	-	Code	Datasets	-	-	-	
zenodo	gitHub	-	-	-	data	code	-	-	-	
Github	Bitbucket	google drive	dropbox	-	code	code	data	data	-	
GitHub	-	-	-	-	Code	-	-	-	-	
GitHub	Zenodo	-	-	-	Code	Data	-	-	-	
framagit.org	-	-	-	-	code	-	-	-	-	
zenodo	-	-	-	-	data	-	-	-	-	
repository	-	-	-	-	both	-	-	-	-	
github	-	-	-	-	code	-	-	-	-	
github	notion	-	-	-	code	classified papers	-	-	-	
ProteomeXchange	-	-	-	-	data	-	-	-	-	
Zenodo	-	-	-	-	both	-	-	-	-	
Zenodo	-	-	-	-	both	-	-	-	-	
github.com	-	-	-	-	Code	-	-	-	-	r
IDR	GIT	-	-	-	images	code	-	-	-	
Datadryad	NCBI SRA	-	-	-	both	data	-	-	-	
zenodo	-	-	-	-	everything	-	-	-	-	t
GitHub	-	-	-	-	both	-	-	-	-	

6.1.1 Why haven't you?

Number of responses: 83



"Other..." text answers:

Datasets are usually large. Also, they follow a format that is conventional in the field and it is understood that the data can be available upon request.

Mon travail est en mathématique pure, mes résultats sont publiés sous formes d'articles

no publication at EPFL yet

I mostly work on experiments. so no codes to publish

new here :-)

Not a practice in my lab

Other...

I am a student, never needed yet

no publishing data myself (platform)

data is available upon request

ownership issues

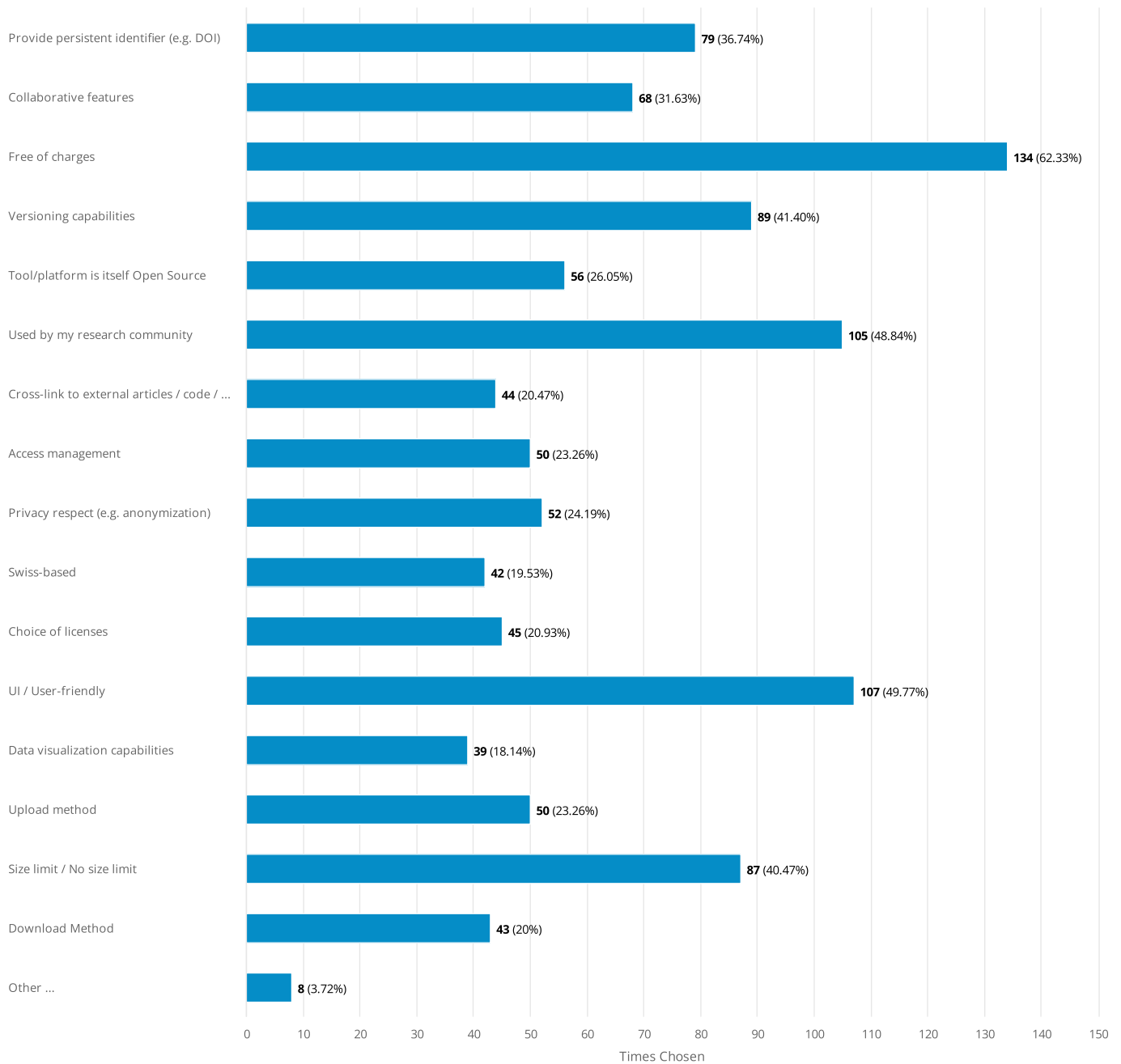
Did not appear necessary

Public data are a problem when people use it for hypothesis testing. This increases the type I error rate.

Not relevant in my field

6.2 Which of the these features do you consider important for such tool(s)/platform(s)?

Number of responses: 215



"Other ..." text answers:

clients for cloud storages for Linux and windows (!)

no opinion, never used

0

Rich metadata

Other ...

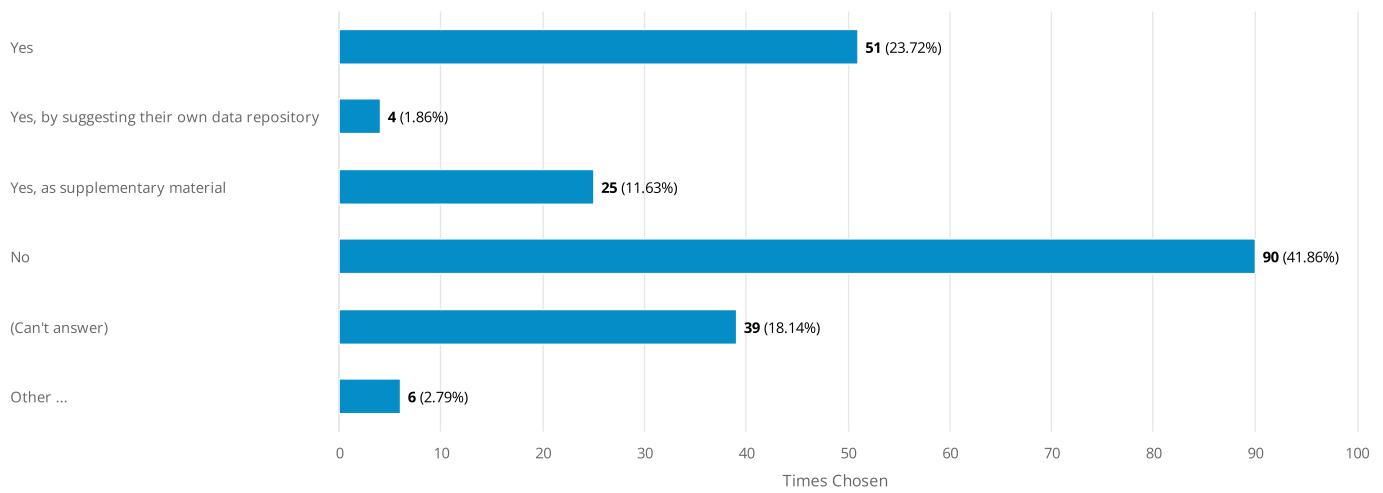
?

comply with FAIR principles

guaranteed availability in future

7. When submitting a scientific article for publication, have you ever been asked by an editor to also publish the underlying data / code?

Number of responses: 215



"Other ..." text answers:

I have been asked, but it was OK to mention that the data was available upon request.

No, because I always do it.

Been asked to provide a data accessibility statement, not explicitly to publish the data/code.

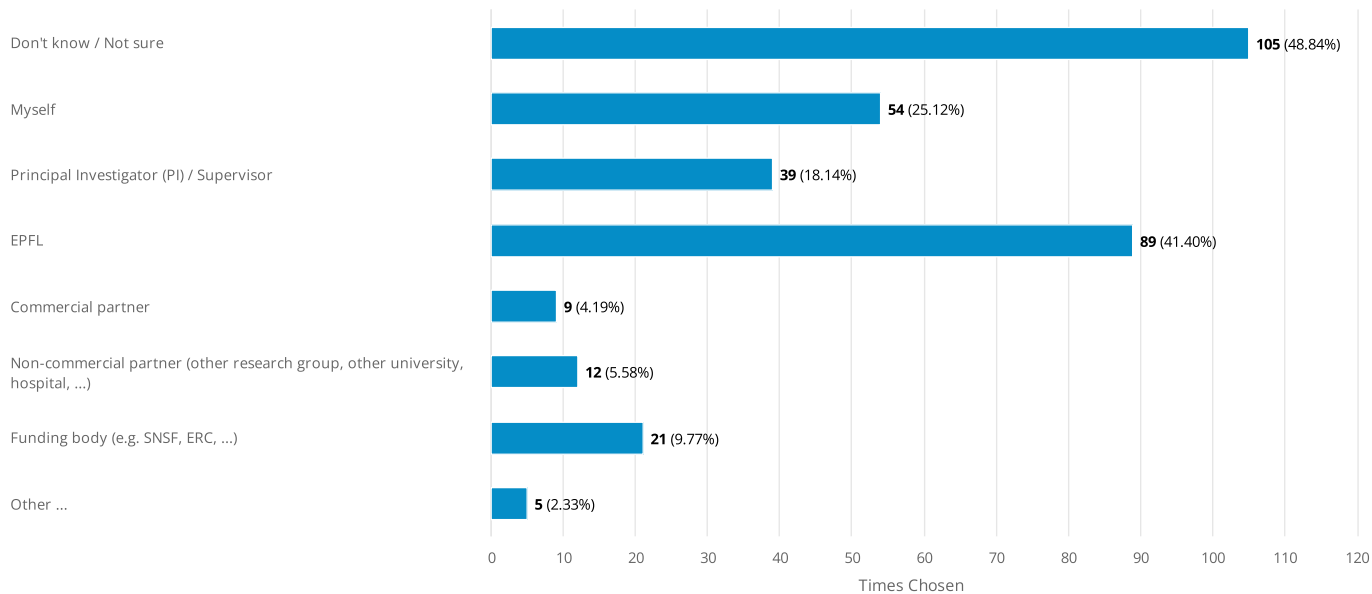
Never submitted yet

Never submitted

Recommended but not required

8. Who owns the research data/code of your project?

Number of responses: 215



"Other ..." text answers:

What does it mean "owns"? The code is shared on an MIT licence, with people who developed it as the copyright owners. The repository is owned by the lab/epfl

i would have guessed EPFL

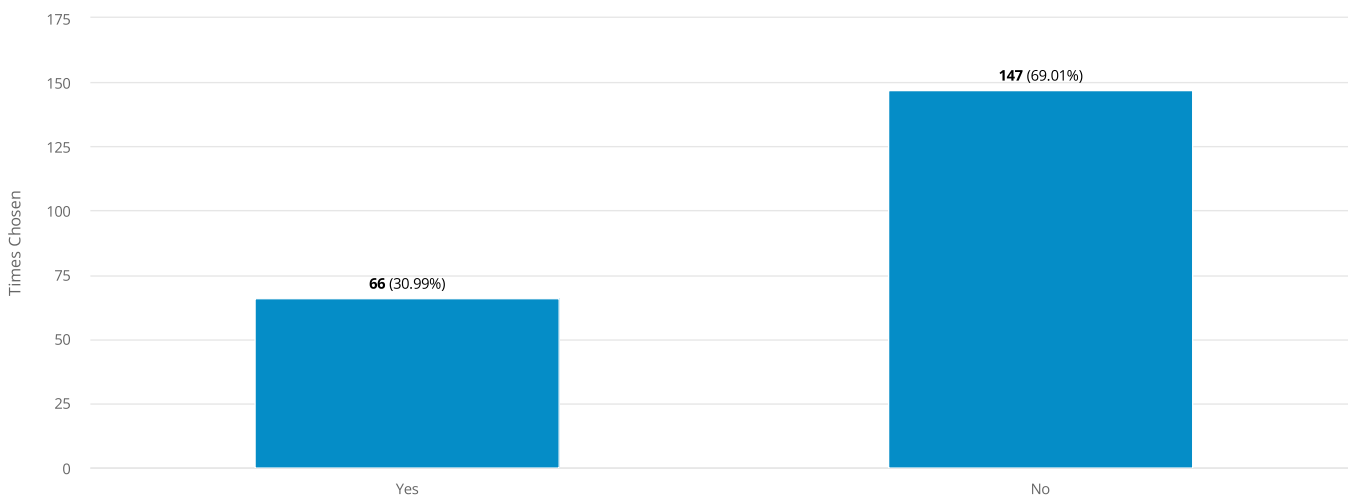
Ne s'applique pas

experimental collaboration / CERN

Depend of project, could be SNSF, EPFL, PI, Me...

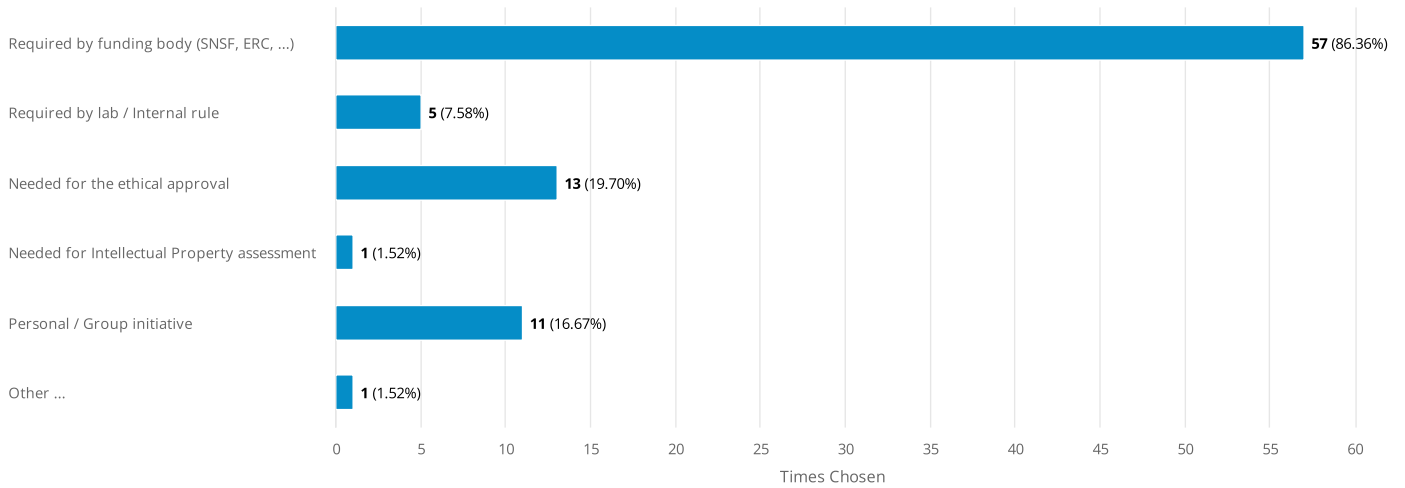
9. Have you ever written a DMP?

Number of responses: 213



9.1 What are the reasons for you to write a DMP?

Number of responses: 66

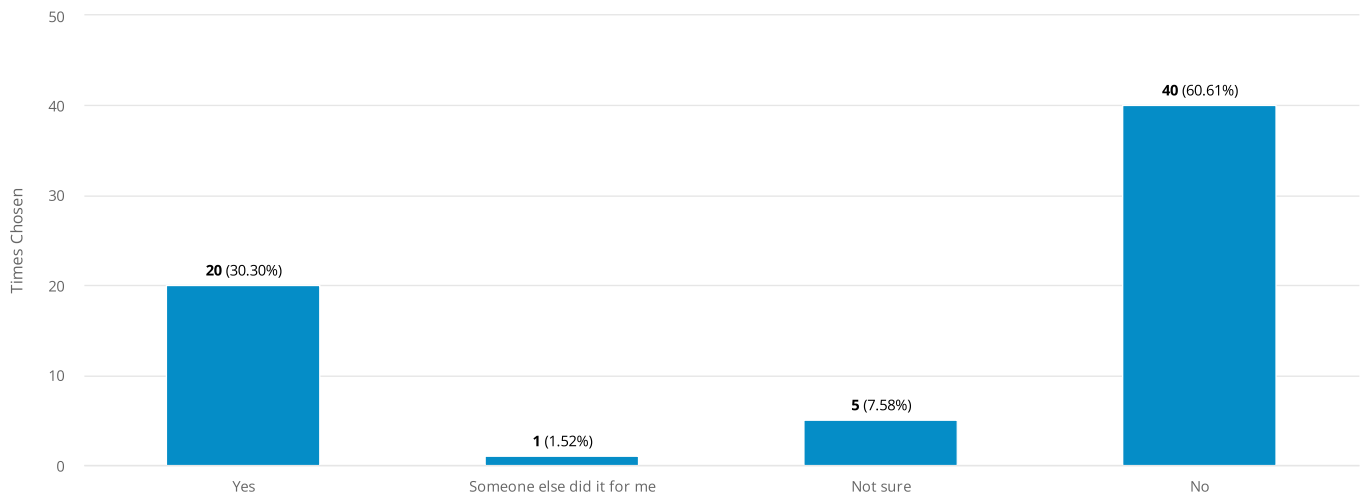


"Other ..." text answers:

Useful to plan and actuate data management

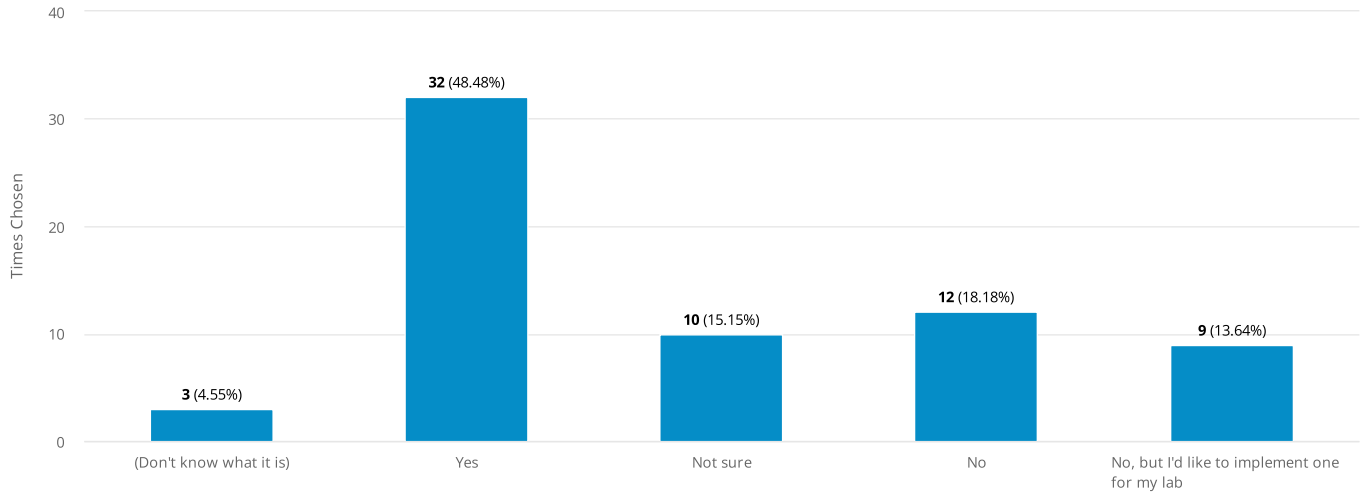
9.2 Have you ever updated a DMP since its first valid implementation?

Number of responses: 66



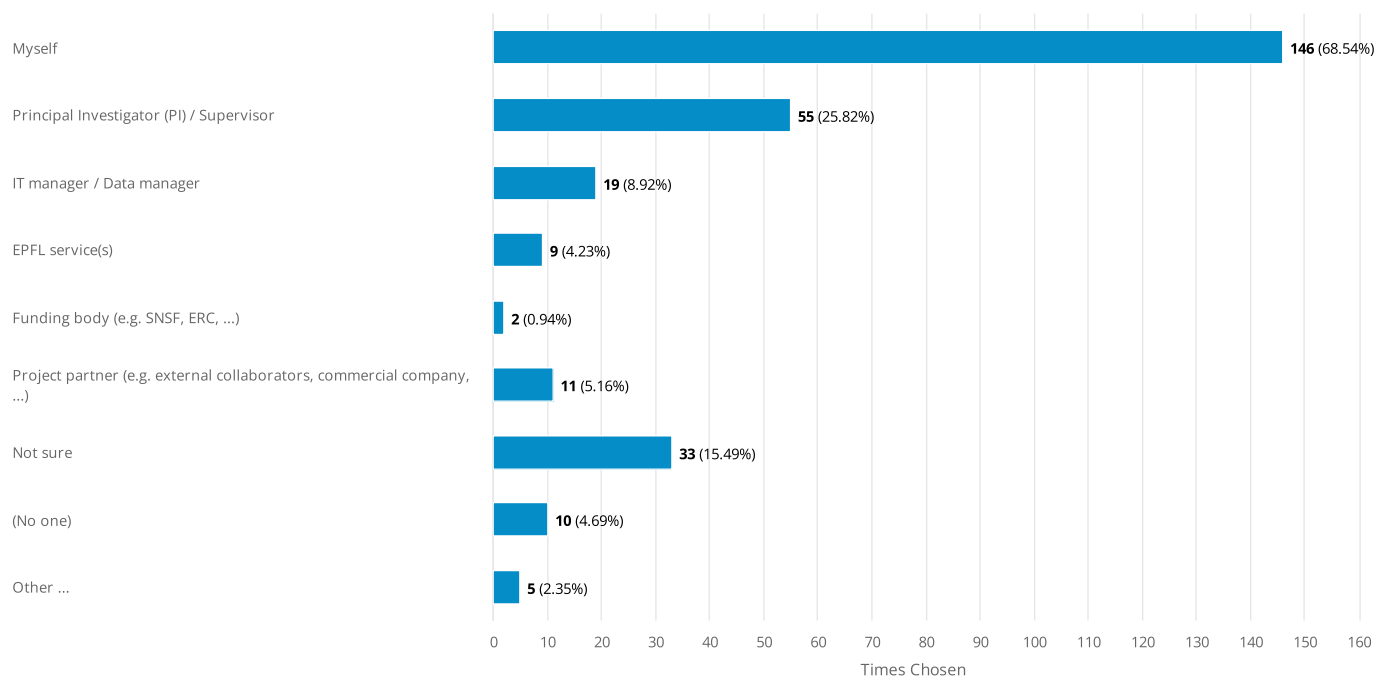
9.3 Do you (or your research group) also have a data management strategy?

Number of responses: 66



10. Who is responsible for managing the research data / code used for your project?

Number of responses: 213



"Other ..." text answers:

dedicated lab members

PhD student, Postdocs who generate and use the data

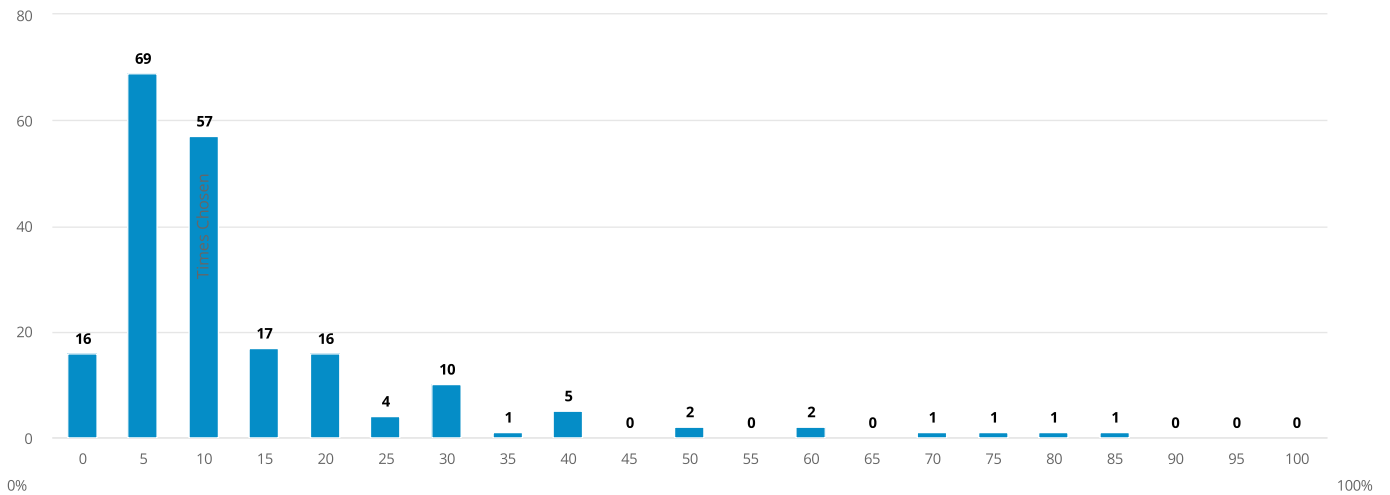
ne s'applique pas

my research group

Colleagues

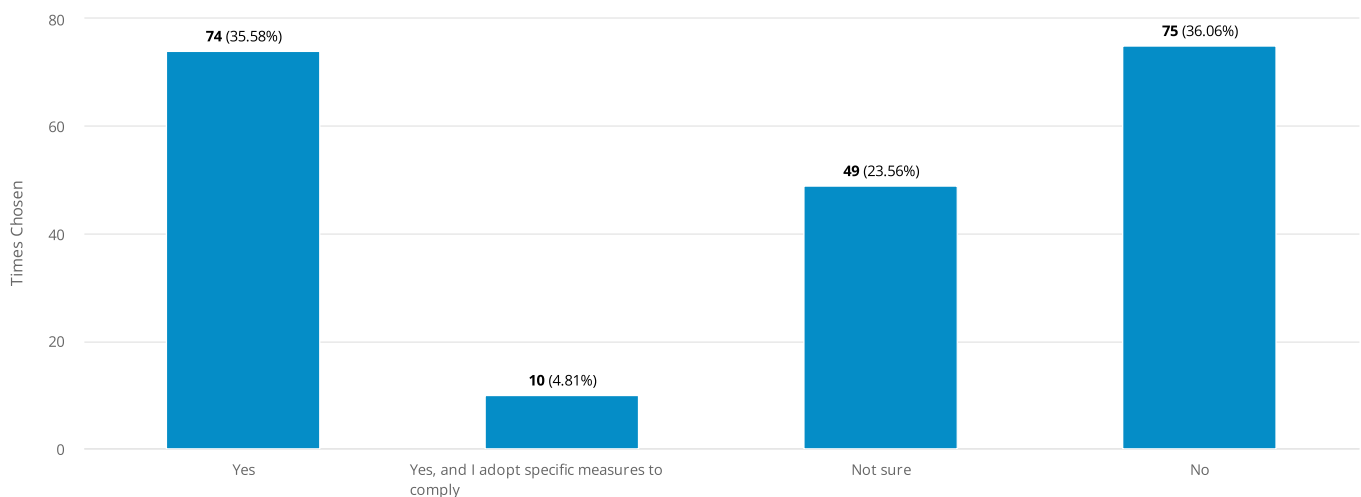
10.1 What percentage of work does this management task represent? (Estimate)

Number of responses: 203



11. Are you aware of Findable, Accessible, Interoperable and Reusable (FAIR) data principles?

Number of responses: 208



11.1 Which measures do you adopt?

Number of responses: 10

Text answers:

Use of AiiDA

Public database for code and (in the future) data, adding metadata to make findable, all data is ascii text to be read by any editor

Publishing code on GitHub; adding version tags associated with specific publications; assign a DOI to versions by synchronising between GitHub and Zenodo; ideally (not always because very time consuming) add proper unittests and documentation to the code.

Perserving all metadata in imaging files
Keeping to standard saving formats, json, xml etc

reusability, documentation

Reusable

File types being open, naming conventions, etc. as recommended by EPFL library consultants

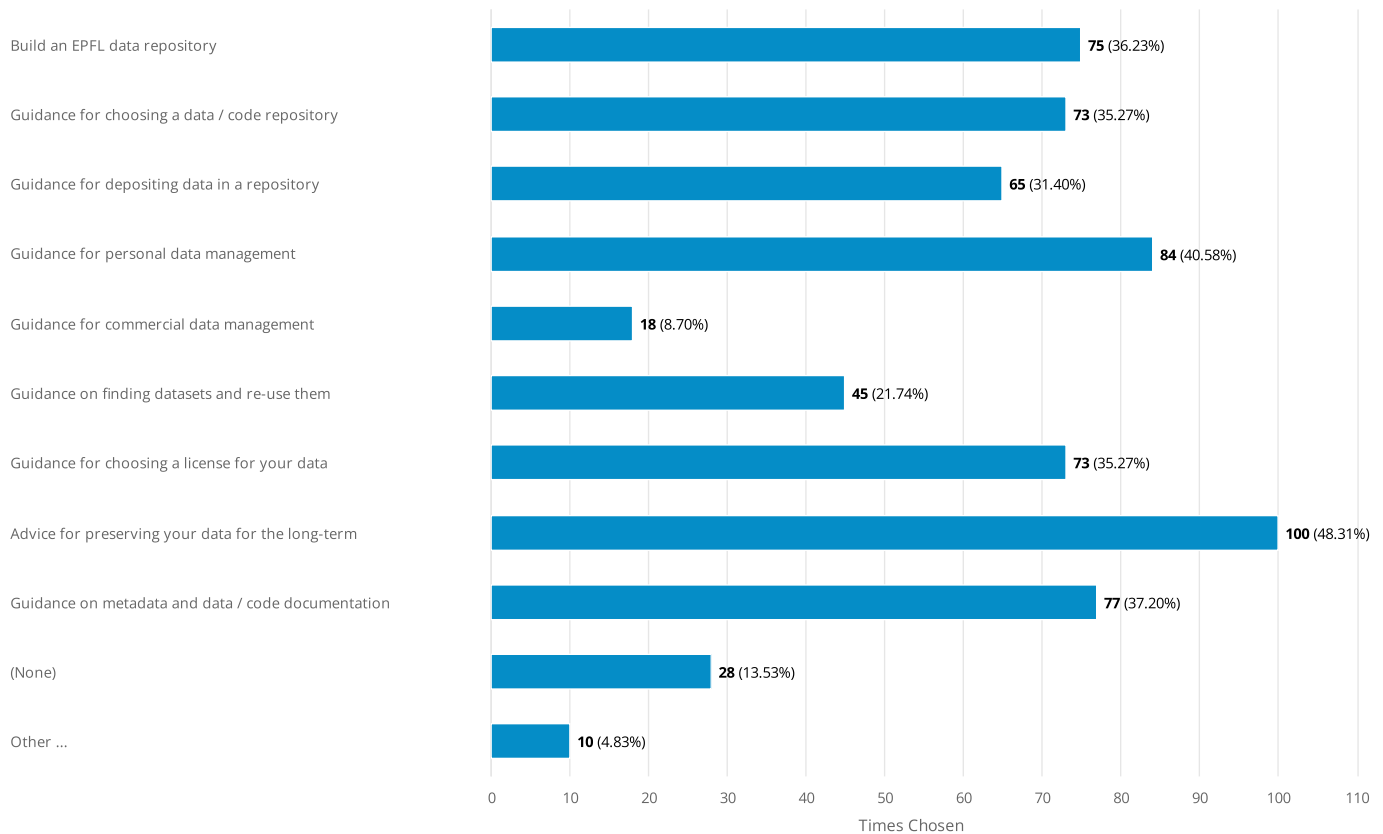
Public as Open Access in Data Repository with DOI. Try to use open formats (not so easy) and open source software for interoperability.

I write my code in Python, so that it is open and can be easily distributed (interoperable).

Open code and data

12. Which of these support offerings would be helpful for your data management practices?

Number of responses: 207



"Other ..." text answers:

None, but larger ssd (at least 4To) to employee would help us not to copy stuff a thousand time everywhere (and larger backup space too)

Consulting on correct data-management for different moments of life-cycle of project

(no comment)

Guidance for sharing data with Project partners (other than epfl)

I believe you need to put more text-based tutorials out there. It is not always efficient to sign up for one of your events. An easy to access (hopefully text-searchable and that it pops up when I ask a question to google and attach EPFL at the end) text-based tutorial listing resources would be immensely useful.

Other ...

automatic tools to help you write DMPs, e.g. forms with boxes to tick

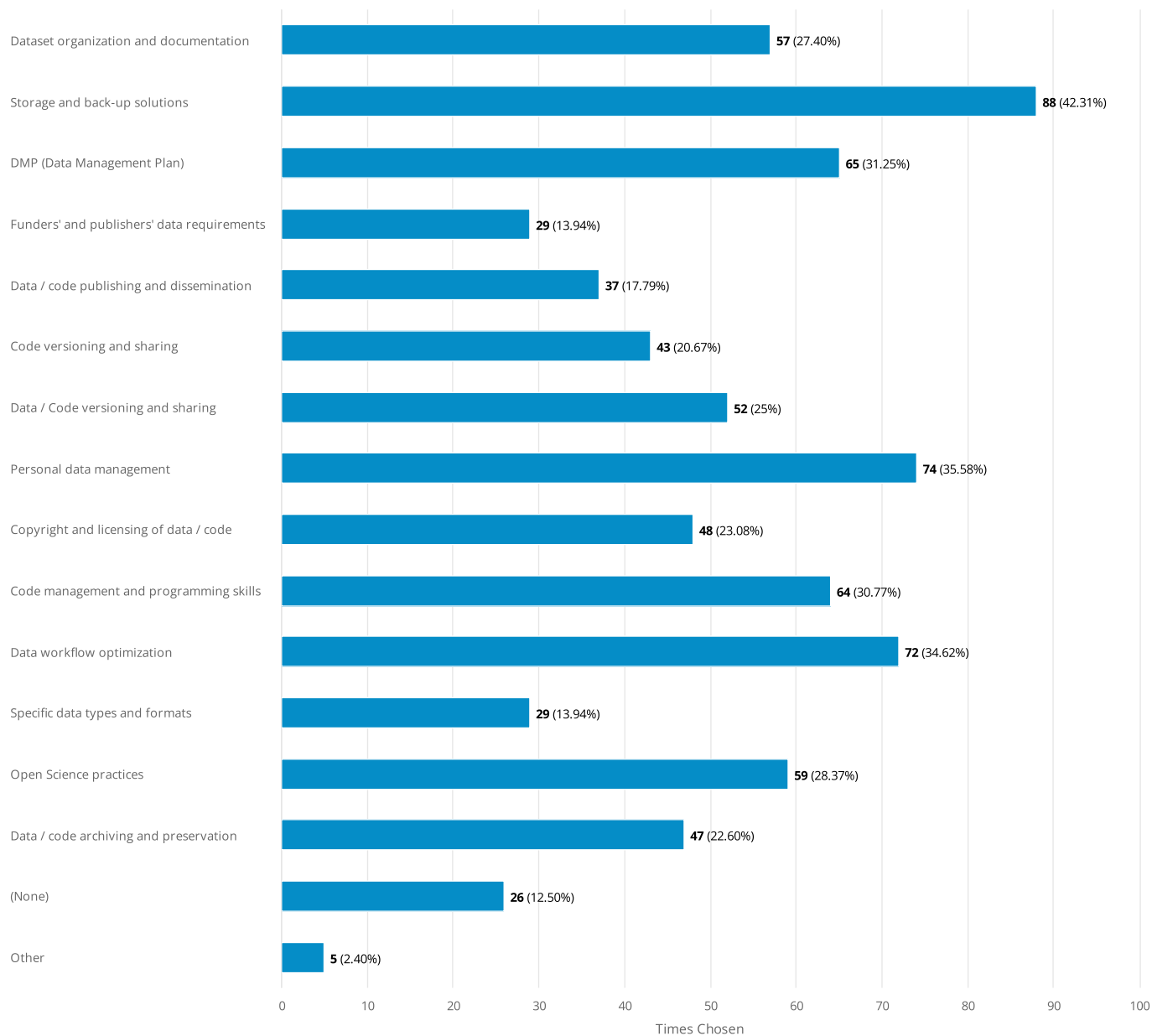
Personalized training to a laboratory

better IT environment at remote collaborating institution

Guidance to establish collective practices of data management within a lab

13. To which of these specialized training offerings would you dedicate at least 1 hour of your time?

Number of responses: 208



"Other" text answers:

I'd happily spend half an hour on many of them.

Tutorial on how to use GitHub effectively (branching, etc.)

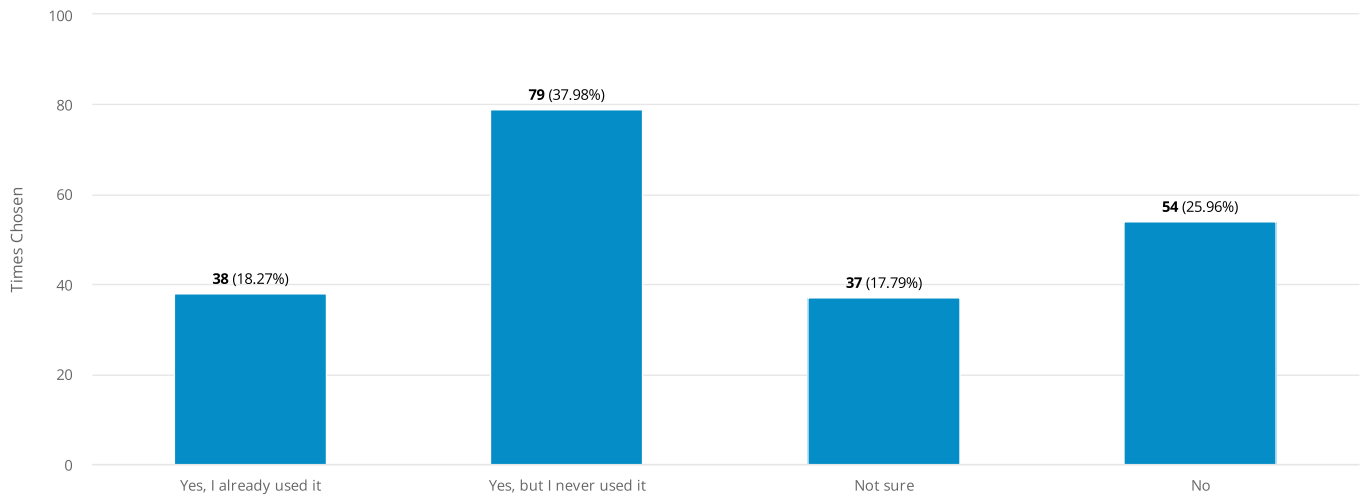
As I said, I do not think taking a course is the most efficient. The information should be there for us to reach out to it when we need it.

Other

No course, just better tools that do some of the thinking for you

14. Have you heard about EPFL Library's dedicated support for Research Data Management, researchdata@epfl.ch?

Number of responses: 208



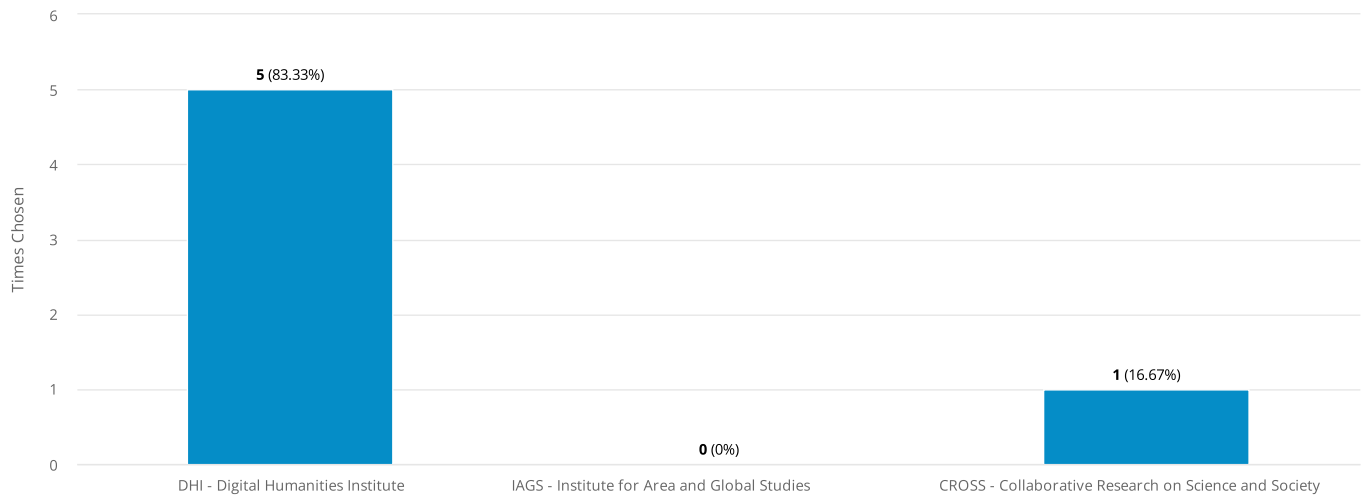
15. In which EPFL School or service do you work?

Number of responses: 206

Answer	Times Chosen	Percentage
(NOT AT THE EPFL)	0	0%
CDH - College of Humanities	6	2.91%
CDM - College of Management of Technology	0	0%
ENAC - Architecture, Civil and Environmental Engineering	25	12.14%
IC - Computer and Communication Sciences	10	4.85%
SB - Basic Sciences	67	32.52%
STI - Engineering	60	29.13%
SV - Life Sciences	31	15.05%
EPFL Middle East	1	0.49%
Other EPFL affiliation (specify)	6	2.91%

15.1 Which CDH Institute?

Number of responses: 6



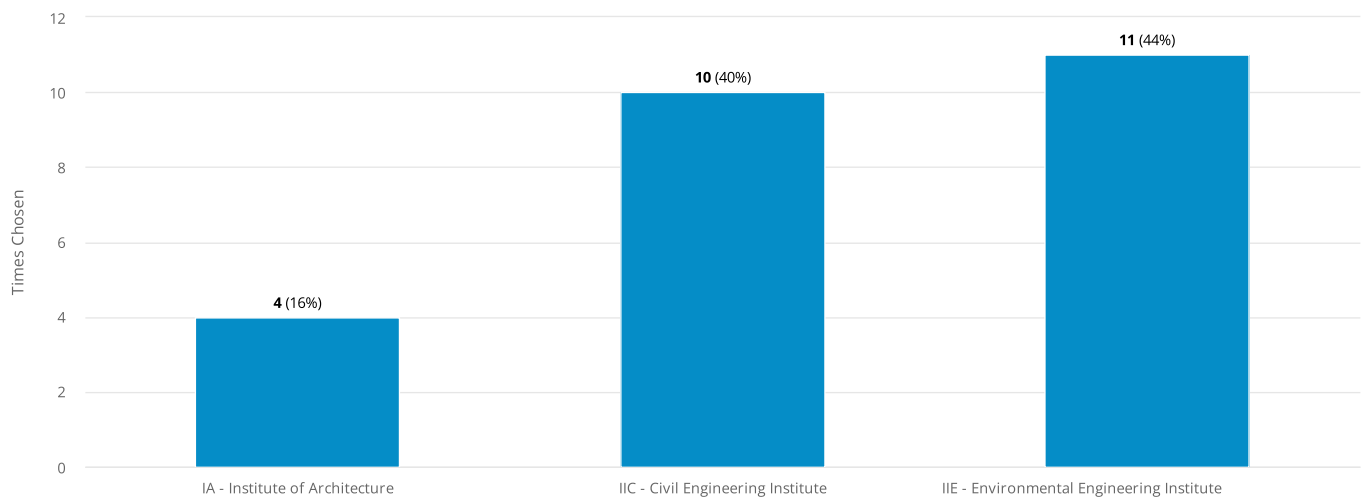
15.1 Which CDM Institute?

Number of responses: 0

No answers collected yet

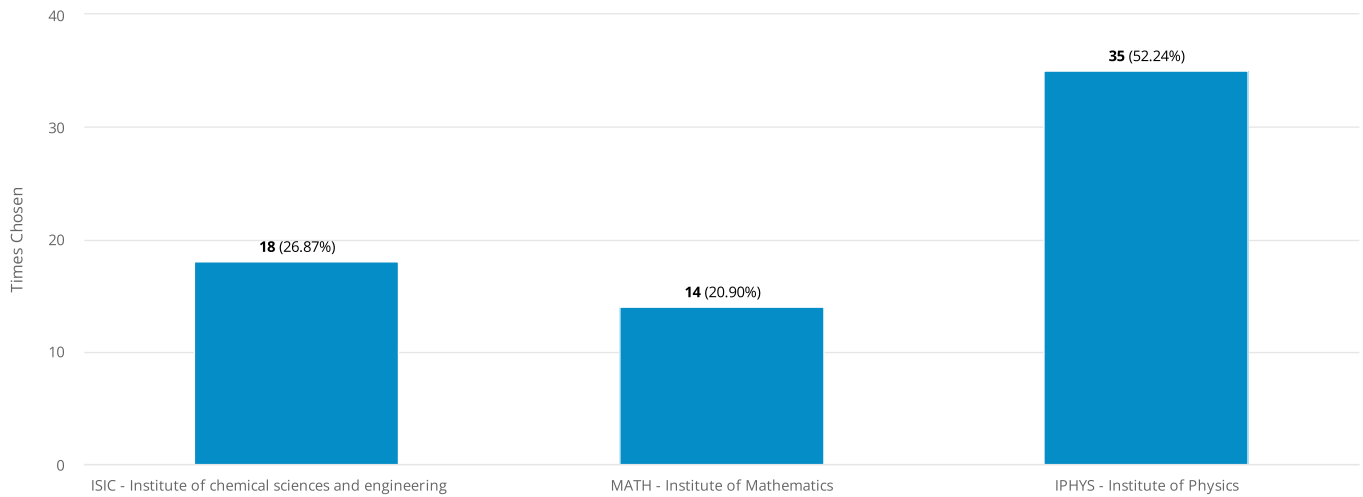
15.1 Which ENAC Institute?

Number of responses: 25



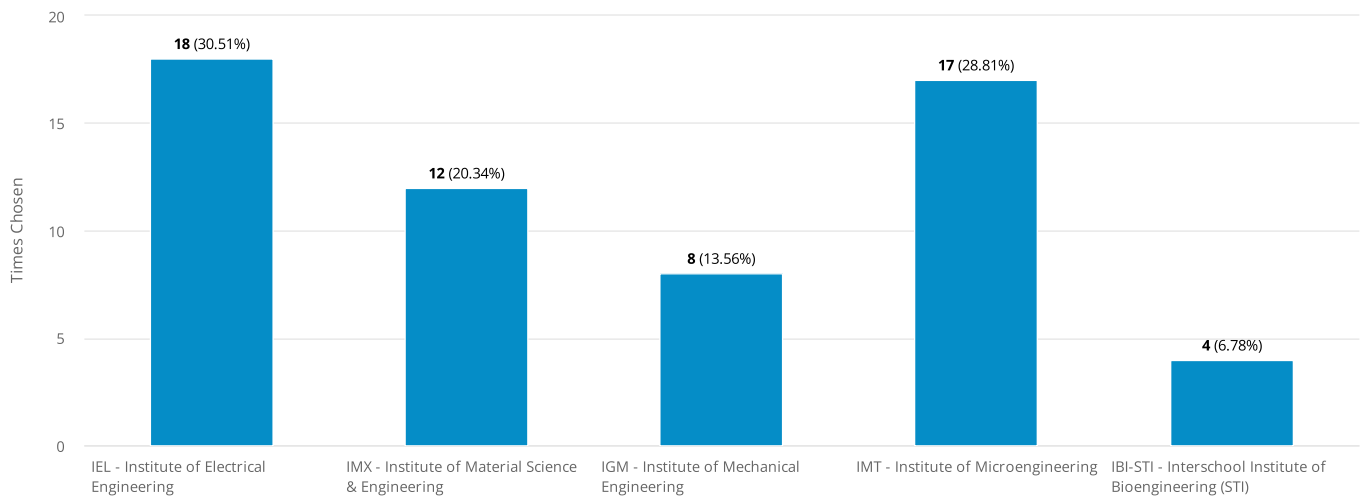
15.1 Which SB Institute?

Number of responses: 67



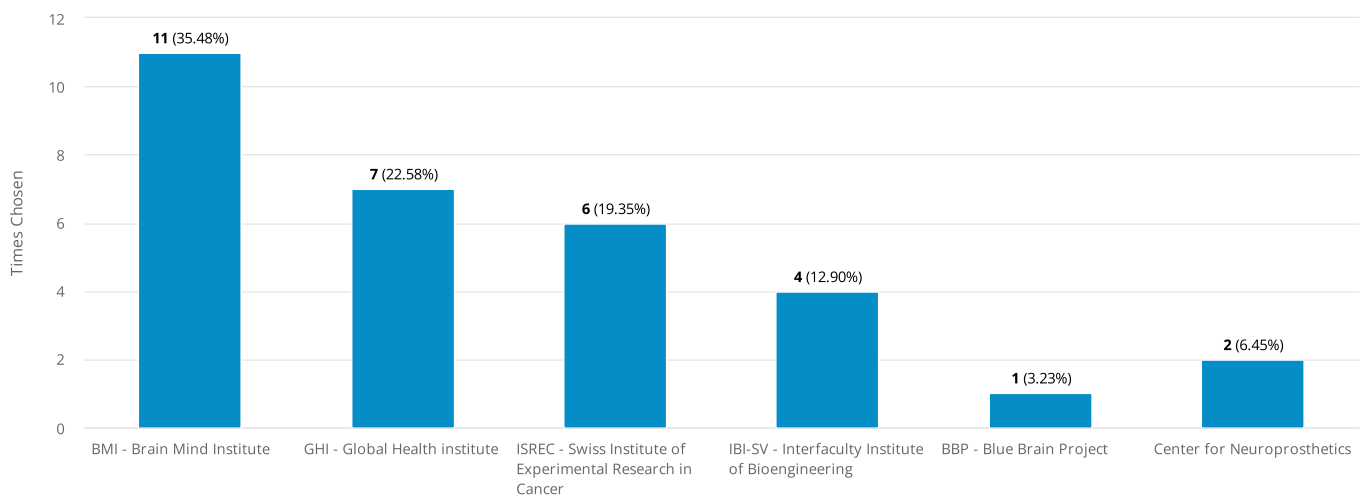
15.1 Which STI Institute?

Number of responses: 59



15.1 Which SV Institute?

Number of responses: 31



15.1 What is your EPFL affiliation?

Number of responses: 6

Text answers:

VPI
AVP CP
retired from STI
EPFL Library
Central Services
dhc

16. In which campus are you located?

Number of responses: 203

Answer	Times Chosen	Percentage
Lausanne (main campus)	181	89.16%
EPFL Neuchâtel	10	4.93%
EPFL Geneva	8	3.94%
EPFL Fribourg	1	0.49%
EPFL Valais Wallis	9	4.43%
Other ...	5	2.46%

"Other ..." text answers:

Zurich
CERN
external research centre (CERN)
Winterthur Company Office
AGORA

17. What is your role?

Number of responses: 202

