

FAIR Research Data Management & Data Management Plan



Dr Vassilios Ioannidis

Lead Computational Biologist - SIB/Vital-IT group
Data Steward, FAIR Specialist - UNIRIS - Direction UNIL
<https://orcid.org/0000-0002-4209-2578>



Dr Cécile Lebrand

Responsable du service de soutien pour la gestion des publications
et des données de recherche FBM UNIL/CHUV - BiUM
Data Steward – UNIRIS - Direction UNIL
<https://orcid.org/0000-0002-2750-3164>

Dr Grégoire Rossier

Project Manager, SIB/Training/Vital-IT group
<https://orcid.org/0000-0002-0065-5053>







Vital-IT
Competence Centre in Bioinformatics and Computational Biology

Services ▾ R&D ▾ About ▾

Welcome to Vital-IT

Vital-IT is a Competency Centre in Bioinformatics and Computational Biology that provides support and technological R&D for life science and clinical research in Switzerland and internationally.


 search for people, tools, blast databases or projects...



Team

The Vital-IT team is composed of a director and 44 scientists and technical staff with expertise in software development, data management, biological data analysis, web technology and trans-disciplinary research.


[learn more...](#)



Services

Vital-IT is installing software and databases available to the research community of the schools of higher education.

[learn more...](#)



Research and development

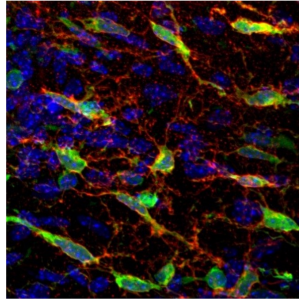
Vital-IT participates into collaborative R&D projects, in which the involvement of its scientists has been highly customizable (from pure IT service to project leaders).

[learn more...](#)

<https://www.sib.swiss>



Home > Publication & Open Access > Data management & Open Research Data



Tools and services for research data
management (DMP and ORD)

Workshops on DMP and FAIR data/ORD

Tools and services for Long Term Storage
(LTS) /archiving of your data at UNIL

Journal and funding agencies Open Data
policies

Where to publish your datasets: Data
repositories

Metadata standards

File formats for long-term preservation
and re-use

Citation for a dataset

Research data confidentiality

Intellectual property for datasets

Publish your dataset in data journals

Find datasets on Open Data repositories

Data management & Open Research Data

The content and the structure of this Web site is licensed under the Creative Commons License (CC BY NC ND Lebrand C.-BIUM library-2016) unless otherwise noted.

In order to build on previous findings, improve transparency and increase results reproducibility, it is important for researchers to be able to re-use research data. For all these reasons, the notion of publication has been evolving over the last ten years and today includes not only the results, but also the essential research data needed to validate the results.

Definition of research data

+

Data life cycle

+

Source of information

+

Head of Data Management Service

Dr. Cécile Lebrand

Tél. +41 (0)21 314 50 81

Cecile.lebrand@chuv.ch

<https://www.bium.ch/en/publication-open-access/data-management/>

Objectives

During this workshop you will learn

- best practices in data management
- FAIR data principles
- how to collect, describe, store, secure and archive data
- file formats for long term datasets storage and re-use
- metadata standards for datasets,
- policies on confidentiality and intellectual property
- journal & funding agencies policies on data sharing
- how to share your datasets related to your publication
- how to fill a DMP corresponding to your project research

Outline

1. Introduction

- a. Open Research, Reproducibility & Funding Agencies
- b. Research Data Management (RDM)
FAIR principles - Data Life Cycle

2. Planning Phase

- a. Data Management Plan (DMP)
- b. Ethical & legal aspects

3. Active phase: Collect, Process, Analyze

- a. Collect, Process, Analyze Data
- b. Data Documentation & Metadata
- c. Data Access Management & Storage

4. Concluding Phase: Preservation & Sharing

- a. Data Preservation (cold storage)
- b. FAIR Principles In Depth
- c. Data Sharing & Reuse (repositories, licenses)

1. INTRODUCTION

a. Reproducibility and Open Research Data

Poll 1: Open Research Data

What are the advantages for you, researchers and the scientific community to make published works and accompanying datasets freely accessible and reusable through Open Access (OA)?

Select the 3 answers that best correspond to you

- Verification of published research
- Preserving accessibility to data
- Allowing reuse and repurposing of data
- Improved methodologies
- Higher diffusion and visibility for the authors' work
- Higher citation rate of your publications
- Fulfillment of funding mandate
- Foster collaboration
- Accelerate innovation

Join at:
vevox.app

ID:
193-360-850



Advantages and disadvantages of Open Research Data

Editorial

Anaesthesia 2023, 78, 551-556

Table 1 Advantages and drawbacks of the open science movement.

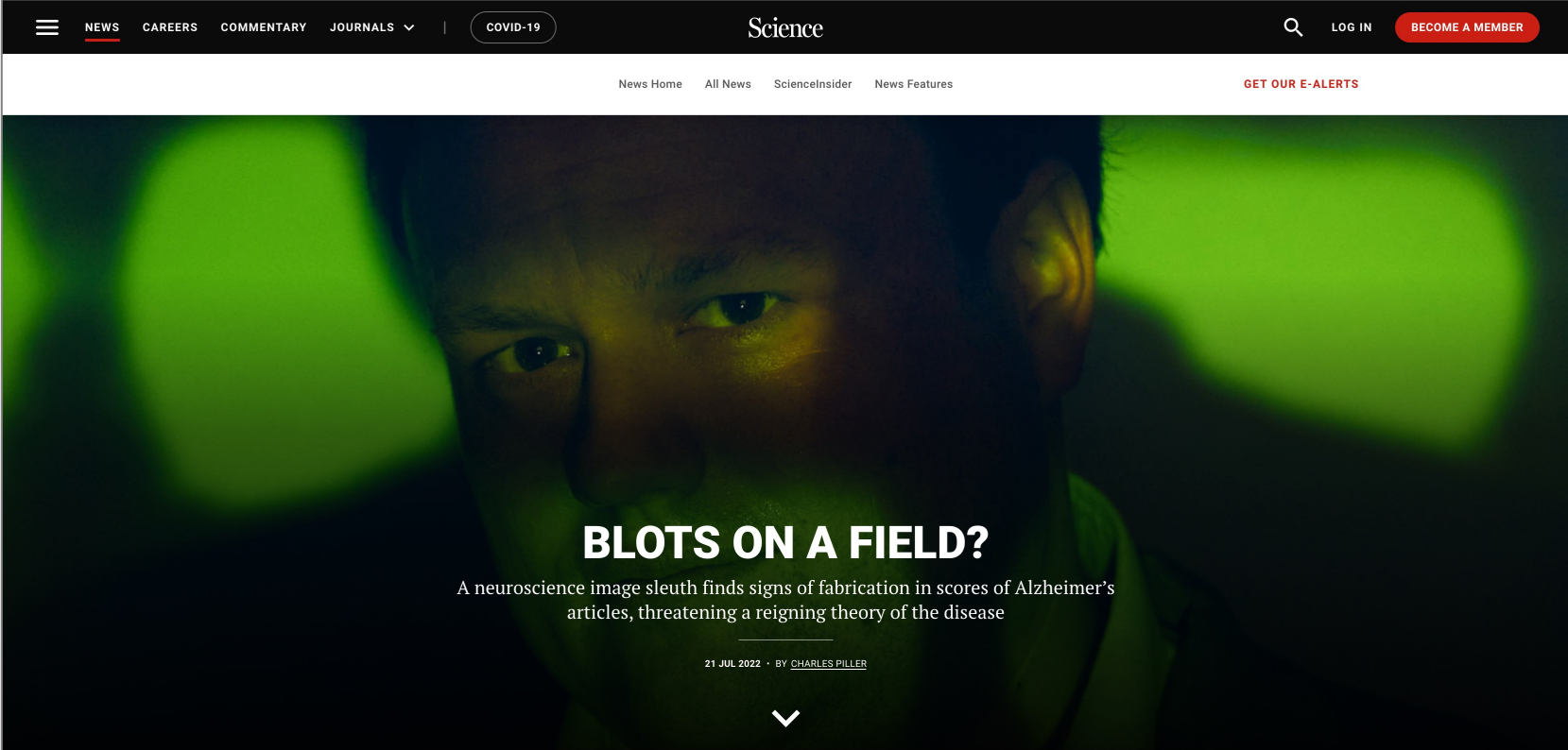
| Advantages | Disadvantages |
|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| Transparency, reproducibility and mitigation against data manipulation | Increased workload for authors to supply repositories with appropriately deidentified data |
| Increased rigour in peer review | Unclear which repositories most appropriate |
| Public access to publicly funded research | Risk of complex data misinterpretation |
| Acceleration of discoveries | Risk of data misuse (dual-use dissemination) |
| Financial savings in access, labour and transaction costs | Additional administration for authors to obtain ethical and institutional data sharing agreements |
| Increased impact of publications | Loss of data ownership |

Open science should be a pleonasm

A. Dos Santos Rocha, E. Albrecht, K. El-Boghdadly

**Context: basic, translational
and clinical research validity
under controversy**

Falsification of results



The screenshot shows the Science website's homepage. At the top, there is a navigation bar with a menu icon, links for NEWS, CAREERS, COMMENTARY, JOURNALS, and a COVID-19 button. The Science logo is centered, with search and login options on the right. Below the navigation, there are links for News Home, All News, ScienceInsider, and News Features, along with a 'GET OUR E-ALERTS' button. The main content area features a large image of a man's face with a green glow. Overlaid on this image is the article title 'BLOTS ON A FIELD?' in large white text, followed by a sub-headline: 'A neuroscience image sleuth finds signs of fabrication in scores of Alzheimer's articles, threatening a reigning theory of the disease'. Below the sub-headline, the date '21 JUL 2022' and author 'BY CHARLES PILLER' are displayed. A white downward-pointing arrow is centered at the bottom of the image area.

NEWS CAREERS COMMENTARY JOURNALS COVID-19

Science

LOG IN BECOME A MEMBER

News Home All News ScienceInsider News Features GET OUR E-ALERTS

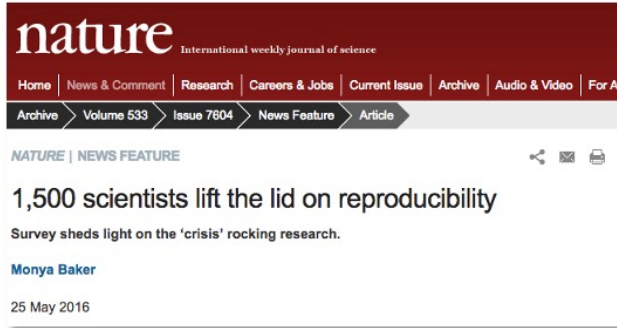
BLOTS ON A FIELD?

A neuroscience image sleuth finds signs of fabrication in scores of Alzheimer's articles, threatening a reigning theory of the disease

21 JUL 2022 · BY CHARLES PILLER

<https://www.science.org/content/article/potential-fabrication-research-images-threatens-key-theory-alzheimers-disease>

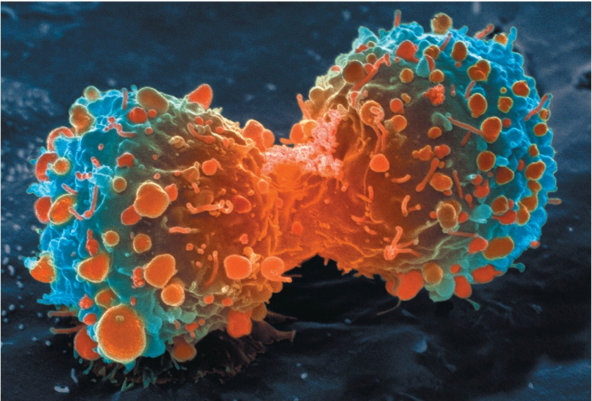
Reproducibility crisis: researchers' view



*(image under © 2016 Macmillan Publishers Limited. All Rights Reserved.
Follow the link below to access the complete article.)*

*(image under © 2016 Macmillan Publishers Limited. All Rights Reserved.
Follow the link below to access the complete article.)*

Preclinical research validity under controversy

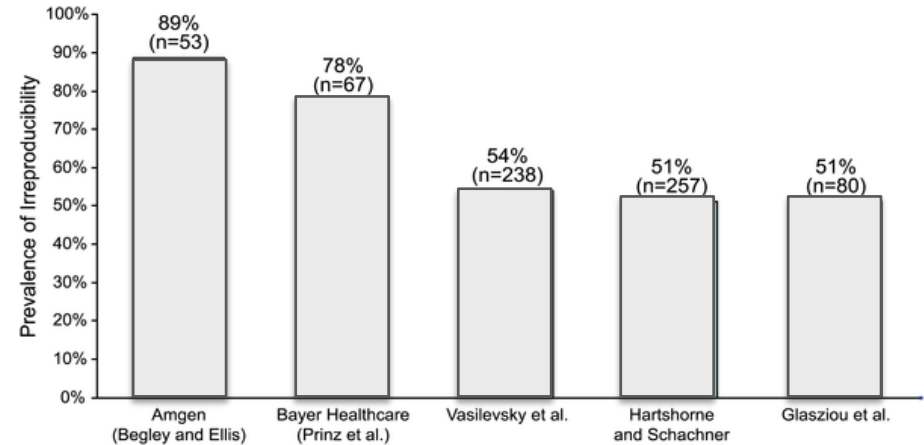


Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

<https://www.nature.com/articles/483531a>



CC-BY. Freedman LP, Cockburn IM, Simcoe TS (2015) PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165
<http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165>

nature reviews drug discovery

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature reviews drug discovery](#) > [news in brief](#) > [article](#)

NEWS IN BRIEF | 13 January 2022

Preclinical cancer research suffers another reproducibility blow

<https://elifesciences.org/articles/71601>



Investigating the replicability of preclinical cancer biology

eLife | Home Magazine Community About Search Alerts [SUBMIT YOUR RESEARCH](#)

Edited by
Roger J Davis et al.

Reproducibility Project: Cancer Biology

Investigating reproducibility in preclinical cancer research.

Collection · Dec 10, 2014

Views 46,472

The Reproducibility Project: Cancer Biology was an initiative to independently replicate selected experiments from a number of high-profile papers in the field of cancer biology. In the end 50 experiments from 23 papers were repeated. The final two outputs from the project recount in detail the challenges the project team encountered while repeating these experiments ('Challenges for assessing replicability in preclinical cancer biology'), and report the results of a meta-analysis that combined the results from all the experiments ('Investigating the replicability of preclinical cancer biology'). The project was a collaboration between the Center for Open Science and Science Exchange.

RELATED

[Cancer Biology](#)

Reproducibility in Cancer Biology: What have we learned?

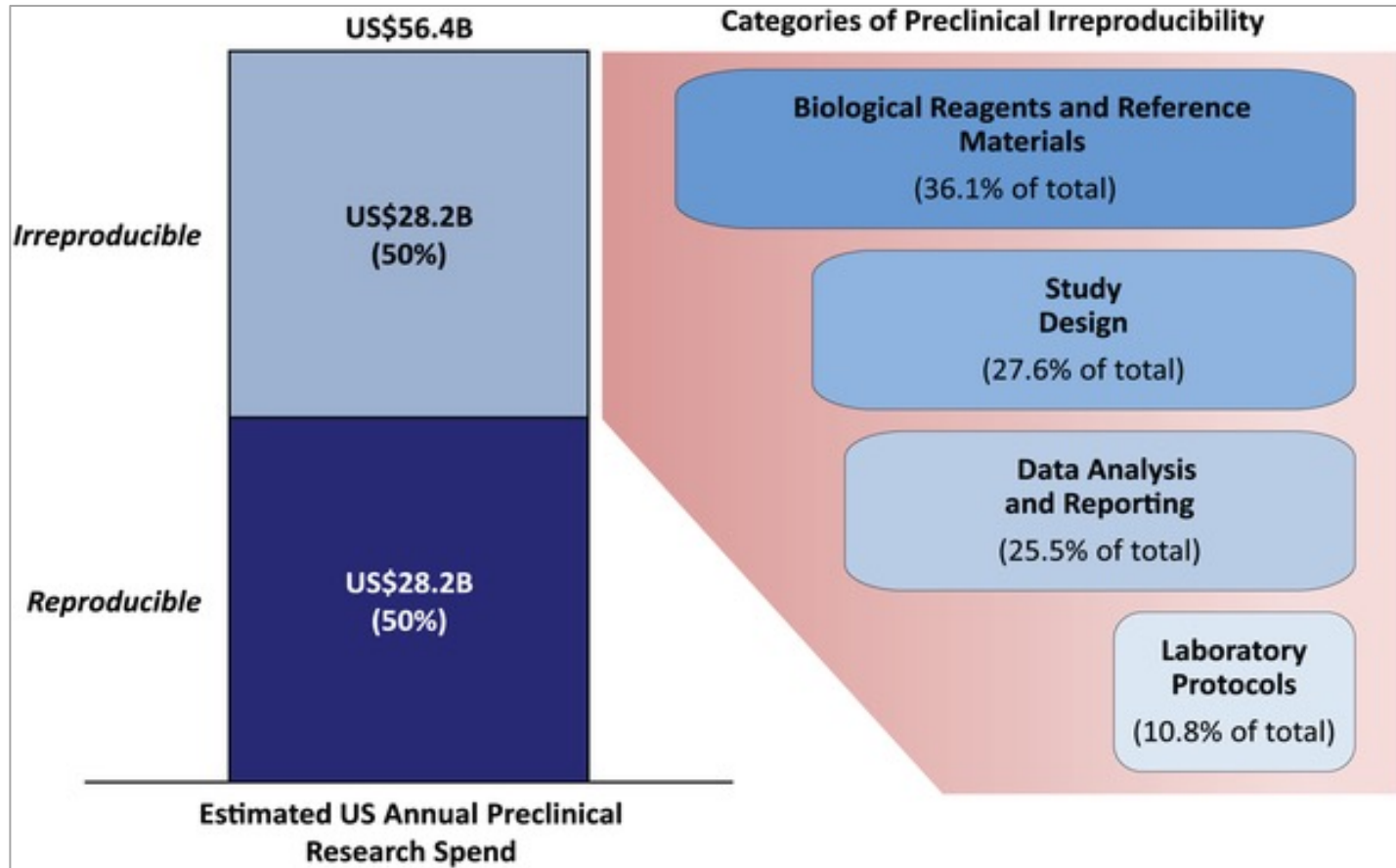
Peter Rodgers, Andy Collings

Editorial · Dec 7, 2021

<https://elifesciences.org/articles/71601>

- 50 experiments from 23 papers were repeated, replicability of 158 effects.
- 92% of replication effect sizes were smaller than the original.
- 46% of replications (51/112) succeeded.

Research expenses and errors contributing to irreproducibility



Journals: Open Data policies & reporting standard guidelines

THIS WEEK

EDITORIALS

CONSERVATION Saving species is far from a walk in the park **p.8**

WORLD VIEW Psychology gears up to check its workings **p.9**



BREAKFAST Chimps plan days to ensure they nab tastiest figs **p.11**

Journals unite for reproducibility

Consensus on reporting principles aims to improve quality control in biomedical research and encourage public trust in science.

Reproducibility, rigor, transparency, and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not necessarily make it right, and just because it is not reproducible does not necessarily make it wrong. A transparent and rigorous approach, however, can almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come from retractions and the objective examination of the resulting data.

It was with the goal of strengthening such approaches in the biomedical sciences that a group of editors representing over 30 major journals, representatives from funding agencies, and scientific leaders assembled at the AAAS headquarters in June of 2014 to discuss principles and guidelines for preclinical biomedical research. The gathering was convened by the U.S. National Institutes of Health, *Nature*, and *Science*.

The discussion ranged from what journals were already doing to address reproducibility and the effectiveness of those measures, to the magnitude of the problem and the cost of solutions. The attendees agreed on a common set of Principles and Guidelines in Reporting Preclinical Research (www.nih.gov/about/reporting-preclinical-research.htm) that list proposed journal policies

EDITORIAL

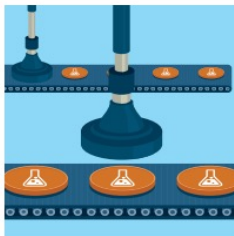
Journals unite for reproducibility

Reproducibility, rigor, transparency, and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not necessarily make it right, and just because it is not reproducible does not necessarily make it wrong. A transparent and rigorous approach, however, can almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come from retractions and the objective examination of the resulting data.

It was with the goal of strengthening such approaches in the biomedical sciences that a group of editors representing over 30 major journals, representatives from funding agencies, and scientific leaders assembled at the AAAS headquarters in June of 2014 to discuss principles and guidelines for preclinical biomedical research. The gathering was convened by the U.S. National Institutes of Health, *Nature*, and *Science*.

The discussion ranged from what journals were already doing to address reproducibility and the effectiveness of those measures, to the magnitude of the problem and the cost of solutions. The attendees agreed on a common set of Principles and Guidelines in Reporting Preclinical Research (www.nih.gov/about/reporting-preclinical-research.htm) that list proposed journal policies

menters were blind to the conduct of the experiment, how the sample size was determined, and what criteria were used to include or exclude any data. Journals should recommend the deposition of data in public repositories where available and link data bidirectionally to the published paper. Journals should strongly encourage, as appropriate, that all materials used in the experiment be shared with those who wish to replicate the experiment. Once a journal publishes a paper, it assumes the obligation to consider publication of a retraction of that paper, subject to its usual standards of quality.



“...scientific journals are standing together in their conviction that reproducibility and transparency are important...”



Marcia McNutt
Editor-in-Chief
Science Journals

The more open-ended portion of the guidelines suggests that journals establish best practices for image-based data (such as screening for manipulation and storing full-resolution archival versions) and how to describe experiments more completely. An example for animal experiments is reporting the source, species, strain, sex, age, husbandry, inbred and strain characteristics, or transgenic animals, etc. For cell lines, one might report the source, authentication, and mycoplasma contamination status. The existence of these guidelines does not obviate the need for replication or independent verification of research results, but should make it easier to perform such replication.

Some of the journals at the meeting already had implemented all or most of these principles and guidelines. But the important point is that a

• Transparency and Openness Promotion (TOP) guidelines adoptées par 560 journaux et 49 associations. *B. A. Nosek et al. Science 2015;348:1422-1425*

• Principles and Guidelines in Reporting Preclinical Research set up by NIH, Nature and Science.

• All papers with Data policies: NPG, Cell Press, PLoS, Science, EMBO, PNAS, Lancet, BMJ, BMC, ...

Downloaded from www.sciencemaq.org on August 28, 2015

<http://www.nature.com/news/journals-unite-for-reproducibility-1.16259>

<http://www.sciencemaq.org/content/346/6210/679.long>

No raw data, no science: another possible source of the reproducibility crisis



No raw data, no science: another possible source of the reproducibility crisis

Tsuyoshi Miyakawa

Molecular Brain 13, Article number: 24 (2020) | Cite this article

59k Accesses | 92 Citations | 2158 Altmetric | Metrics

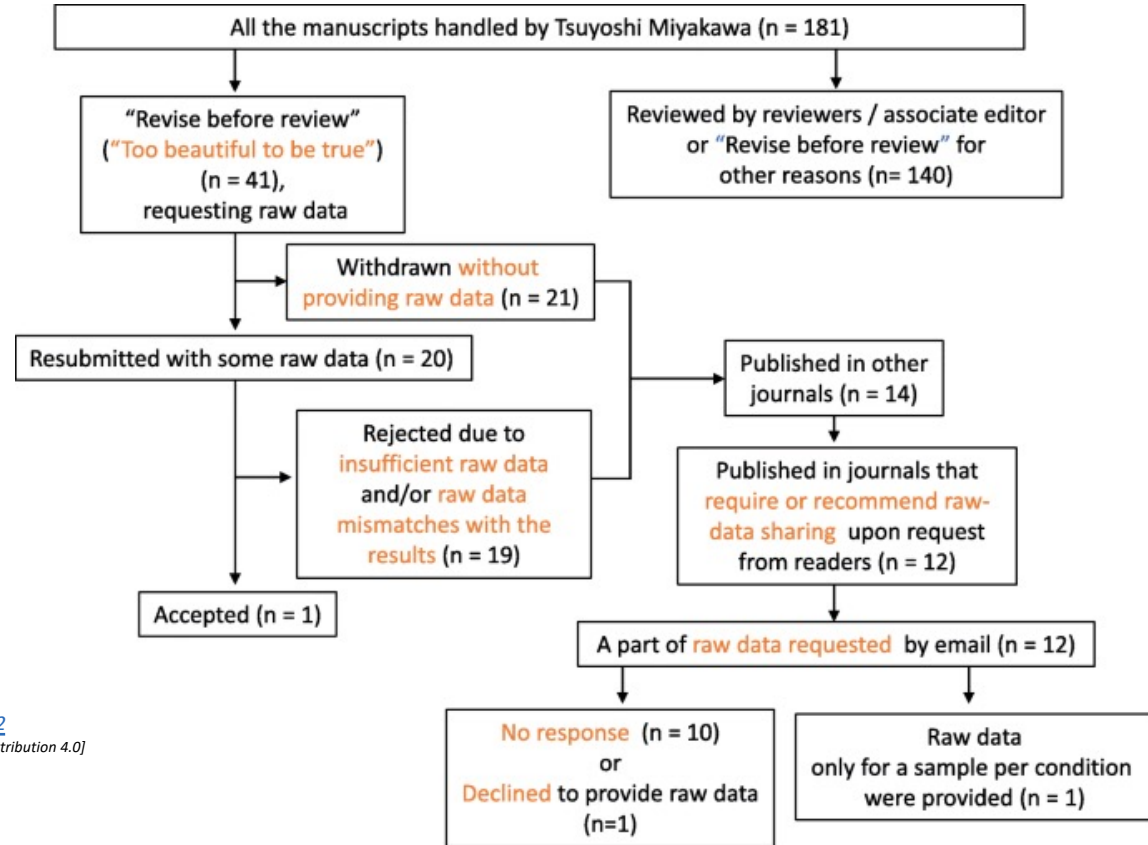
Abstract

A reproducibility crisis is a situation where many scientific studies cannot be reproduced. Inappropriate practices of science, such as HARKing, p-hacking, and selective reporting of positive results, have been suggested as causes of irreproducibility. In this editorial, I propose that a lack of raw data or data fabrication is another possible cause of irreproducibility.

As an Editor-in-Chief of *Molecular Brain*, I have handled 180 manuscripts since early 2017 and have made 41 editorial decisions categorized as “Revise before review,” requesting that the authors provide raw data. Surprisingly, among those 41 manuscripts, 21 were withdrawn without providing raw data, indicating that requiring raw data drove away more than half of the manuscripts. I rejected 19 out of the remaining 20 manuscripts because of insufficient raw data. Thus, more than 97% of the 41 manuscripts did not present the raw data supporting their results when requested by an editor, suggesting a possibility that the raw data did not exist from the beginning, at least in some portions of these cases.

<https://molecularbrain.biomedcentral.com/articles/10.1186/s13041-020-0552-2>

© The Author(s). 2020 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0



Swiss Funding Agency policy on Open Research Data



SNSF values research data sharing as a fundamental contribution to the impact, transparency and reproducibility of scientific research.

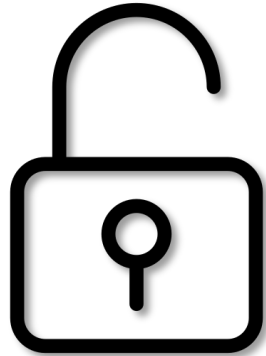
Since October 2017:

- Researchers have to **provide a DMP** as an integral part of their research proposal.
- SNSF expects that metadata and data generated by funded projects will be **publicly accessible in non-commercial, public repositories** provided there are no legal, ethical, copyright or other issues.

1. INTRODUCTION

b. Research Data Management (RDM) FAIR principles - Data Life Cycle

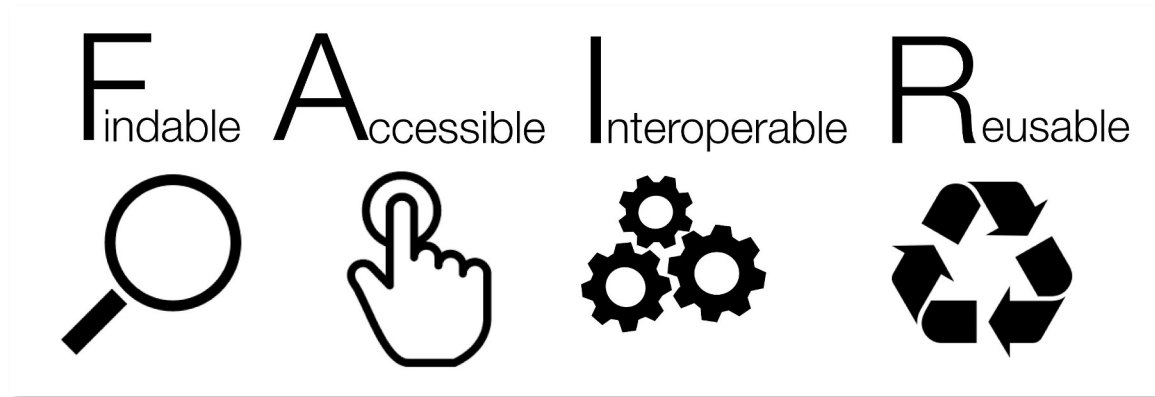
Open Research Data (ORD) Principles



flaticon.com

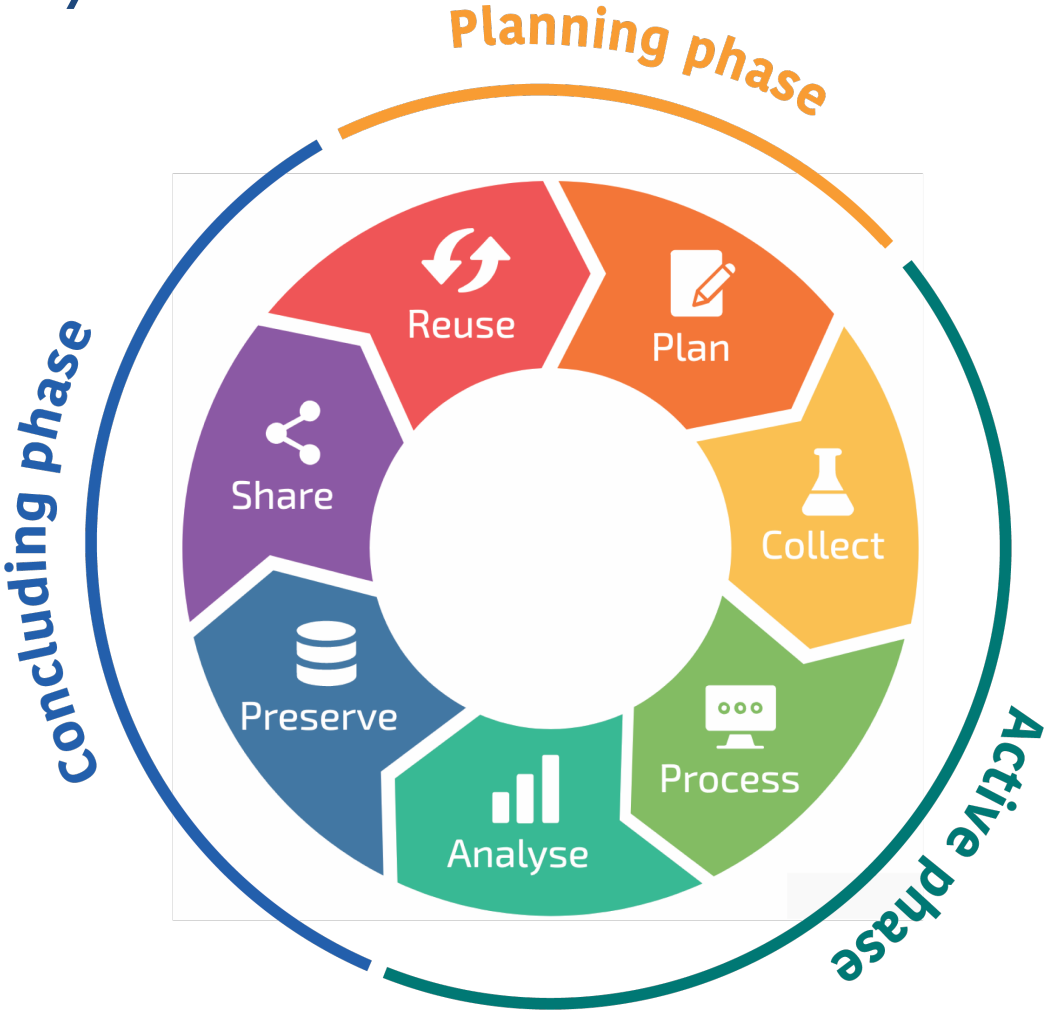
ORD values research data sharing as a fundamental contribution to the impact, transparency and reproducibility of scientific research.

≠ The FAIR Principles

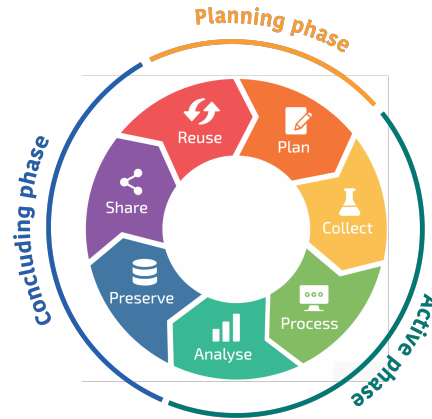


https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

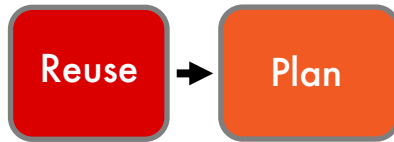
Data Life Cycle



Overview of a project's phases



Planning Phase



Reuse of previous data

Requirement from funding agencies

Research integrity

Legal framework
Ethical issues

Active Phase



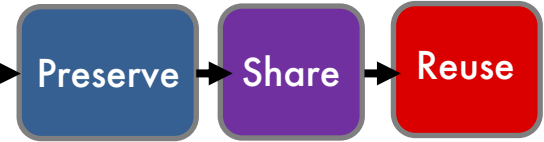
Storage and backup

Data access management- security

Data access management- sharing

Metadata and Documentation

Concluding Phase



Long term preservation

Publication in Repositories

DOI and licence for reuse

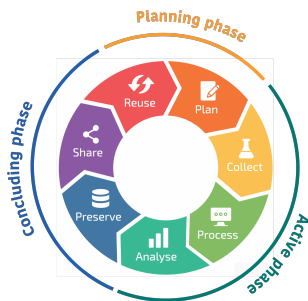
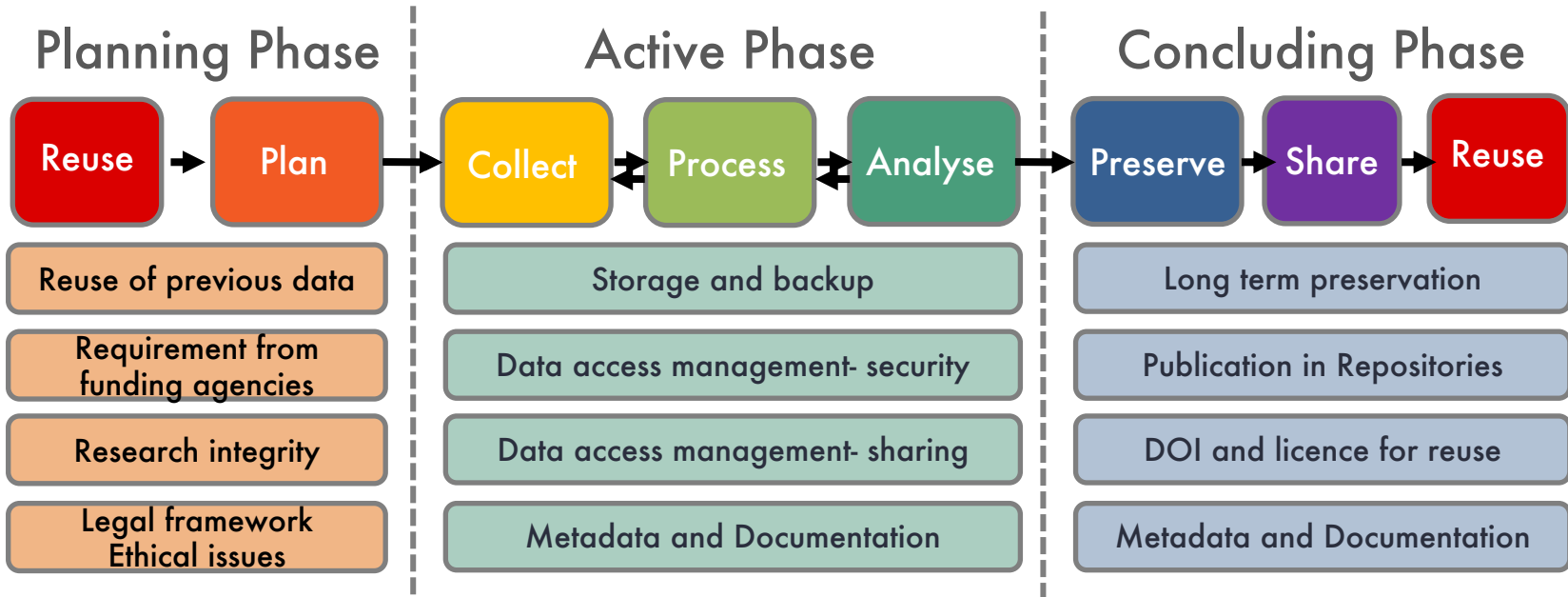
Metadata and Documentation



2. PLANNING PHASE

a. Data Management Plan

Data Management Plan



Content

- anticipate in detail the management of your research data (analyses, organization, storage, security and sharing)
- specify the type of data
- budget, intellectual property, and monitoring over time.

Video “The what, why and how of data management planning”

SNFS DMP

Practical aspects

Information: SNSF guidelines for researchers

SNSF guidelines for researchers

The SNSF has elaborated guidelines for researchers concerning the Data Management Plans (DMPs).

[Guidelines](#) and [Content of the DMP](#)

SNSF regulations

The regulations related to the SNSF policy on Open Research Data can be found in the Funding Regulations and in the General Implementation Regulations.

[FR Art. 47](#) and [IR Art. 2.13](#)

Eligible Costs for SNSF grants

The **costs of enabling access to research data** that is collected, observed or generated under an SNSF grant are eligible if the research data is deposited in recognised scientific, digital data archives (data repositories) that meet the FAIR principles and do not serve any commercial purposes ([IR 2.13](#)). It is permissible to upload data to commercial repositories, but only the data preparation costs will be covered by the SNSF.

Applications and Projects

Grant application 8

1. Personal data

- Responsible applicant
- Other applicants
- Applicants' employment
- Project partners

2. Application data

- Basic data I
- Basic data II
- Use-inspired project
- Re-submission
- Continuation of
- Link to other SNSF projects
- Further requested and available funds (not from the SNSF)
- University or research institution
- Requested funding

4. Data management plan (DMP)

- Research requiring authorisation or notification
 - Exclusion of external reviewers
 - General remarks on the project
- ### 3. Annexed documents (upload)
- Research plan
 - CV and research output list
 - Quotes
 - Cover letter
 - Official certificates
 - Lead Agency and International Co-Investigator Scheme
 - Other annexes
 - Administrative part of the



Grant application 8

New application

Project funding in biology and medicine (division III)

Deadline: 03 April 2018 17:00 Swiss local time

Start: -

In preparation

Data management plan (DMP)

Import DMP

Please describe how you plan to make the research data Findable, Accessible, Interoperable and Reusable ([FAIR data principles](#)) in the following sections. Each of the four topics should be addressed with a level of detail appropriate to the project and research field. Sub-questions and help texts are available for each issue. The "questions you might want to consider" will help you to complete the form. However, depending on the project and research field, you may not need to address each of these questions in your DMP.

Complete the DMP form in the same language as your research plan.

The information provided in this template is not part of the scientific evaluation and will not be shared with external reviewers. Note, however, that the final version of the DMP will be published on P3 ([public database of the SNSF](#)) at the end of the project.

Detailed [guidelines](#) are available about the DMP. Furthermore, answers to a set of [frequently asked questions \(FAQs\)](#) about open research data (ORD) are also available.

I do not submit a DMP for the following reason:

1. Data collection and documentation

- ✔ 1.1 What data will you collect, observe, generate or reuse?
- ! 1.2 How will the data be collected, observed or generated?
- ! 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- ! 2.1 How will ethical issues be addressed and handled?
- ! 2.2 How will data access and security be managed?
- ! 2.3 How will you handle copyright and Intellectual Property Rights issues?

3. Data storage and preservation

- ! 3.1 How will your data be stored and backed-up during the research?
- ! 3.2 What is your data preservation plan?

4. Data sharing and reuse

- ! 4.1 How and where will the data be shared?
- ! 4.2 Are there any necessary limitations to protect sensitive data?
- ✔ 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- ✔ 4.4 I will choose digital repositories maintained by a non-profit organisation.

SNSF DMP – content of the mySNF form



SWISS NATIONAL SCIENCE FOUNDATION

Data Management Plan – content of the mySNF form

| Question | Help text |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 Data collection and documentation | |
| 1.1 What data will you collect, observe, generate or reuse? Questions you might want to consider: - What type, format and volume of data will you collect, observe, generate or reuse? - Which existing data (yours or third-party) will you reuse? | Briefly describe the data you will collect, observe or generate. Also mention any existing data that will be (re)used. The descriptions should include the type, format and content of each dataset. Furthermore, provide an estimation of the volume of the generated data sets. (This relates to the <i>FAIR Data Principles</i> F2, I3, R1 & R1.2) |
| 1.2 How will the data be collected, observed or generated? Questions you might want to consider: - What standards, methodologies or quality assurance processes will you use? - How will you organize your files and handle versioning? | Explain how the data will be collected, observed or generated. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated measurements, data recording standards, usage of controlled vocabularies, data entry validation, data peer review, etc. Discuss how the data management will be handled during the project, mentioning for example naming conventions, version control and folder structures. (This relates to the <i>FAIR Data Principle</i> R1) |
| 1.3 What documentation and metadata will you provide with the data? Questions you might want to consider: - What information is required for users (computer or human) to read and interpret the data in the future? - How will you generate this documentation? - What community standards (if any) will be used to annotate the (meta)data? | Describe all types of documentation (README files, metadata, etc.) you will provide to help secondary users to understand and reuse your data. Metadata should at least include basic details allowing other users (computer or human) to find the data. This includes at least a name and a persistent identifier for each file, the name of the person who collected or contributed to the data, the date of collection and the conditions to access the data. Furthermore, the documentation may include details on the methodology used, information about the performed processing and analytical steps, variable definitions, references to vocabularies used, as well as units of measurement. Wherever possible, the documentation should follow existing community standards and guidelines. Explain how you will prepare and share this information. (This relates to the <i>FAIR Data Principles</i> I1, I2, I3, R1, R1.2 & R1.3) |

Eligible costs for SNSF grants

Project partners

2. Application data

Basic data I

Basic data II

Use-inspired project

Re-submission

Continuation of

Link to other SNSF projects

Further requested and available funds (not from the SNSF)

University or research institution

Requested funding

Data management plan (DMP)

Research requiring

Finances

+ Add expense

Travel

Conferences and workshops

Additional project costs (incl. consumables)

Scientific open access e-publications

Direct costs of infrastructure use

Computing time and data **+**

Costs for granting access to research data (Open Research Data)

The costs of computing time and cloud computing as well as for data acquisition or for gaining access to data are eligible.

The costs must have been incurred specifically for services provided in the context of the approved research project and **may not include** any general costs for service and maintenance of the infrastructure.

Costs for granting access to research data (Open Research Data) (Research funds) ?

General information

Designation *

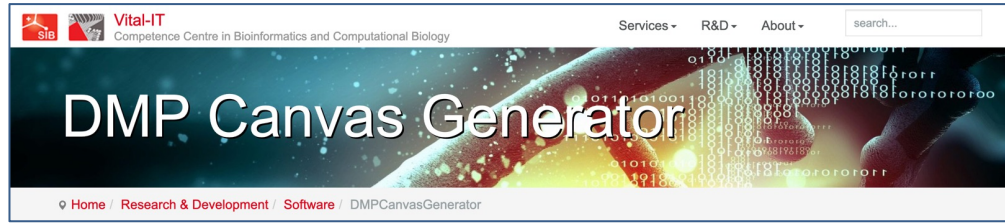
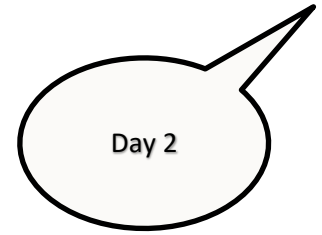
| | |
|---------|---------|
| 1. Year | 02/2018 |
| | 01/2019 |

Amount CHF

Relation to research plan / Comments / Additions *

The costs of granting access to research data that was collected, observed or generated under an SNSF grant are in principle eligible if they are deposited in a recognised scientific repository according to [art. 47](#) of the funding regulations and [art. 2.13](#) of the implementation regulations.

DMP tools and resources



<https://dmp.vital-it.ch/>



<https://dmp.unil.ch/>



<https://dmponline.dcc.ac.uk/>



<https://ds-wizard.org/>



<https://rdmkit.elixir-europe.org/>



<https://fairplus-project.eu/>



2. PLANNING PHASE

b. Ethical & Legal aspects

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?
- 2.3 How will you handle copyright and Intellectual Property Rights issues?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

2.1. How will ethical issues be addressed and handled?

Questions you might want to consider

- What is the relevant protection standard for your data? Are you bound by a confidentiality agreement?
- Do you have the necessary permission to obtain, process, preserve and share the data? Have the people whose data you are using been informed or did they give their consent?
- What methods will you use to ensure the protection of personal or other sensitive data?

SNSF says...

Ethical issues in research projects demand for an adaptation of research data management practices, e.g. how data is stored, who can access/reuse the data and how long the data is stored. Methods to manage ethical concerns may include: **anonymization of data; gain approval by ethics committees; formal consent agreements**. You should outline that all ethical issues in your project have been identified, including the corresponding measures in data management. (This relates to the *FAIR Data Principle A1*)

Poll 2: Ethical issues

What is the relevant protection standard for sharing your data?

Select the right answers

- Informed consent indicating that data are used for research purposes
- Provision for data sharing in the informed consent
- Approval by ethics committees
- Data pseudo-anonymization
- Data anonymization
- Ensure that there are no direct identifiers
- Ensure that there are no indirect identifiers
- Unlimited indirect identifiers
- Remove any nonessential identifying details
- Aggregate data if required
- Reduce the precision of a variable if required

Join at:
vevox.app

ID:
193-360-850



Information: Research on human subjects in Switzerland

Data concerning humans made publicly available must remain totally confidential and be anonymized at best. Researchers should include a provision for data sharing in the informed consent.

- Consult the [Commission cantonale d'éthique de la recherche sur l'être humain](#) before planning research, data use and data sharing.
- Comply with the general privacy protection law according to the [Federal Act on Data Protection](#) ([Loi fédérale sur la protection des données, LPD](#)).
- Adhere to the [Federal Act on Research involving Human Beings](#) ([Loi relative à la recherche sur l'être humain, LRH](#)) created to protect the dignity, privacy and health of human beings involved in research.

General consent for data sharing of sensitive human data

General consent for research allows you to decide whether or not you wish to :

- Make your clinical data and residual biological samples available for research.
- Contribute to the CHUV Genomic Biobank by donating a small amount of blood for genetic analysis for research purposes.
- Can be use by researchers for ORD

<https://www.chuv.ch/fr/consentement-general/cg-home>



The image shows a screenshot of the CHUV website's general consent page for research. The page has a dark blue header with the CHUV logo and the text "Consentement général pour la recherche". Below the header, there are three navigation links: "COMPRENDRE", "PARTICIPER", and "PROTÉGER", each with a downward arrow. The main content area is a large blue rectangle with the text "Consentement général pour la recherche" in white. The CHUV logo is also visible in the bottom right corner of the main content area.

Specific individual consent for data sharing of sensitive human data

Data privacy for sensitive information related to personal and private information will be handled carefully.

Indeed, we will divulgate and provide open-access to sensitive human data only **after receiving explicit consent for publishing data openly** of the individuals:

Proposition from CER-VD:

«Afin de promouvoir la recherche, les journaux et organismes de financement demandent parfois le partage des données utilisées pour les publications. Il est possible que vos données soient mises à disposition sur un site dédié pour l'utilisation par d'autres chercheurs et la validation des résultats publiés. Dans le cas d'un tel partage, vos données seront toujours codées de façon à ce qu'un utilisateur de ces données ne puisse pas remonter à votre identité. Vous devez être d'accord avec cela si vous participez à l'étude.»

<https://www.cer-vd.ch/>



Research on human subjects in Switzerland

Encoding/de-identification / anonymization of data; gain approval by ethics committees; formal consent agreements:

- Manage the above three title points according to your situation.
- Consult a specialized service for correct encoding/de-identification /anonymization of your data and ask funding agencies for cost coverage.

For research data sharing: If some data cannot be shared publicly because they are bound by legal, ethical or confidentiality criteria, you should explain their specific constraints.

Data Protection Impact Assessment (DPIA)

Process designed to identify risks arising out of the processing of personal data and to minimise these risks as far and as early as possible

Tool: hands-on guidance how to de-identify health-related data based on a subjective risk assessment approach done in accordance with Swiss law requirements.

- "Template de-identification use case evaluation and risk assessment"
- "Data de-identification - phased approach, Guidance for de-identification of health-related data in compliance with Swiss law requirements"

It is recommended to complete the assessment in the concept phase of the research project prior submission to the ethics committee.

<https://sphn.ch/document/template-use-case-evaluation-and-risk-assessment/>



Best Practices for Sharing Human Subjects Data

Dryad Requirement

- Ensure that there are **no direct identifiers**.
- **Limit indirect identifiers**. (Dryad allows a maximum of 3.)
- Remove any nonessential identifying details.
- **Aggregate data** – variables which may be potentially revealing, e.g., age, can be grouped.
- **Reduce the precision of a variable** – e.g., remove day and month from year of birth; use country instead of city; add or subtract a small, randomly chosen number.
- Restrict the upper or lower ranges of a continuous variable to hide outliers by collapsing them into a single code.
- Provide good documentation of your data in a **README file**.

<https://datadryad.org/docs/HumanSubjectsData.pdf>



Data sharing of sensitive human data

Take home message

- Ensure privacy protection through proper data anonymization.
- Ensure that shared data (used in conjunction with other publicly available data) do not contain information that allow the identification of an individual.

**Be conscious that no perfect anonymization is feasible!!!!
Make sure to get the appropriate patient consent before starting
your research on Human!!!!**

Other sensitive data: animal or species sensitive research data

Permission and data management of animal data throughout the project

- Manage animal data in accordance with the « Ordonnance sur la protection des animaux, section 6. Documentation et statistique » (Art 143 Registre des animaux ; Art 144 Procès-verbaux de l'expérience).
- Form A and Form B on eTV (Swiss Federal electronical TierVersuch software)

For research data sharing

- In this case the project will not necessitate specific limitations on data sharing.
- Restriction in the case of endangered species data may apply (see dryad guidance).

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?
- 2.3 How will you handle copyright and Intellectual Property Rights issues?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

2.3. How will you handle copyright and Intellectual Property Rights issues?

Questions you might want to consider

- Who will be the owner of the data?
- Which licenses will be applied to the data?
- What restrictions apply to the reuse of third-party data?

SNSF says...

Outline the **owners of the copyright and Intellectual Property Right (IPR) of all data that will be collected and generated, including the licence(s). For consortia, an IPR ownership agreement** might be necessary. You should comply with relevant funder, institutional, departmental or group policies on copyright or IPR. Furthermore, clarify what permissions are required should third-party data be re-used. (This relates to the FAIR Data Principles I3 & R1.1)

Poll 3: Copyright

Who will be the owner of the data?

Select the right answer

- Your university
- The PI
- The collaborators
- The SNSF

Join at:
vevox.app

ID:
193-360-850



Copyright and intellectual property for datasets

- In the European Union and USA, data collection **involving creativity is copyrightable**.
- The structural elements of a database involving originality is **not explicitly covered by copyright in Switzerland** (see code des obligations).
- **Check your contract with your university** (e.g: according to the UNIL-CHUV contract any data created or modified in the course of the professional activity by a UNIL collaborator **belongs to the UNIL** (except for copyrights belonging to the creator). For any use other than for scientific and academic purposes, the institution approval is necessary (eg for patenting and commercialization).

How will you handle copyright and Intellectual Property Rights issues?

Legal data agreement

- Collaboration agreement in the project ([Data agreement wizard](#)).
- Consortium Agreement (CA), Data Transfer and Use Agreement (DTUA) and Data Transfer and Processing Agreement (DTPA) of the Swiss Personalized Health Network ([SPHN](#))

Open licenses for data

Promote sharing and unlimited use of the data that you produced using [explicit open licences](#).

Restrictions applying to the reuse of third-party data?

If you cannot share some data because of commercial and patenting issues you should contact your technology transfer office at your University and explain the specific constraints.



3. ACTIVE PHASE

a. Collect, Process, Analyse

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 2.3 How will you handle copyright and Intellectual Property Rights issues?
- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

1.1. What data will you collect, observe, generate, reuse?

Questions you might want to consider

- What type, format and volume of data will you collect, observe, generate or reuse?
- Which existing data (yours or third-party) will you reuse?

SNSF says...

Briefly describe the **data you will collect, observe or generate**. Also **mention any existing data that will be (re)used**. The descriptions should include the type, format and content of each dataset. Furthermore, provide an **estimation of the volume** of the generated data sets. (This relates to the *FAIR Data Principles* F2, I3, R1 & R1.2)

Type, format and volume of data

Typical data used in a research project

- Raw data
- Protocols
- Codes/scripts/programs
- Processed data
- Results
- Notes/notebooks

Type, format and volume of data

Types

- **State:** raw versus analysed data, qualitative vs quantitative
- **Nature:** (un)structured text, spreadsheet, image, video, audio, database, map, etc...
- **Mode of collection:** experimental, observational

Type, format and volume of data

Files format

Appropriate file format crucial to access and reuse your data.

- **Open file** formats: used by anyone (file specifications publicly available).
- **Proprietary file** formats: only with software provided by the vendor.

*Prefer **open file formats** to share or archive data, they are commonly used and less prone to obsolescence*

<https://www.bium.ch/en/publication-open-access/data-management/#5>

Type, format and volume of data

Data format (examples)

- Cell microscopy images (TIFF)
- Sequencing data (FASTQ)
- Scripts (.sh, .r, .py)
- MS data (mzXML, **PKL***)
- Interview videos (MP4)
- Texts accompanying videos (PDF)

** Proprietary format*

(https://en.wikipedia.org/wiki/Mass_spectrometry_data_format)

Type, format and volume of data

Scale

- **Estimate** the volume during the planning phase.
- Volume will influence on the **costs** (storage and backup).
- Which data to keep for **long-term** preservation (archive)?

Type, format and volume of data

Data reuse

Consider **re-using** existing data from earlier projects

- Avoid unnecessary experiments
- Make research more robust by aggregating results

Provenance

- From **your** group
- From **another** group (what permission?)

Example

Data generated will be in various **types**, **formats** and **sizes** of datasets, all of which will be accessible using common software allowing easy access and long-term validity during and after the project, thus facilitating data sharing. The format/types of data include:

- i) **Cell images** e.g. phase and fluorescence, and electron micrographs (**~5,000 images over project**). Software used includes OpenLab, Softworx and IN Cell Investigator, with data saved as **software-specific files e.g. liff and lg3 files**, as well as generic formats such as **jpeg, tiff** etc.
- ii) **Mass spectrometry spectra** (from <50 samples). MS data will be analysed using Bruker Data Analysis or Thermo Excalibur software (generating **xml** and **raw files**) and proteins will be matched to the *T. brucei* genome dataset using the Matrix Science Mascot search engine. **Each LC-MS data file is between 1-2GB.**

Example

To ease the process, list in a table: type of experiment, **types** of data, equipment, software, **file formats** and **volume**

| Types | Equipment | Software | Data storage format | Data archiving / sharing format | Volume |
|-----------------------------------------------------------------|----------------------|---------------------------------------------------------|----------------------------|----------------------------------------|----------------|
| Microscopy images | | | | | |
| Raw data: microscopy cell images | Zeiss LSM 710 Quasar | ZEN lite software | .liff | .tiff uncompressed, JPEG2000 | 500 GB |
| Secondary data: 3D Z-stack reconstructions and processed images | | Imaris 7.2.1 software; Fiji/ImageJ; Adobe Photoshop CS5 | .ims, .tif series, .PSD | .tiff uncompressed, JPEG2000 | 1 TB |
| Analysed data: cell quantifications | | Imaris 7.2.1 software, Excel | .ims, .xlsx | .xlsx; .csv | 3 GB |
| Raw data :time lapse video microscopy | Leica SP5 | LAS AF Lite 4.0.11706 | .czi files; .avi, .mov | MPEG-4; Motion JPEG 2000 | 500 GB |
| Analysed data: tracking function | | Metamorph software 6.0 | .xlsx | .xlsx; .csv | 2 GB |
| Western Blots | | | | | |
| Raw data: cell images | | | | | 1 GB |
| Analysed data: quantification | | | | | 500 MB |
| | | | | | TOTAL = |

[Bium recommended File formats for long-term preservation and re-use](#)

[Dryad recommended formats if you plan to publish your data on this Data repository](#)

Hands-on
Data collection table

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 2.3 How will you handle copyright and Intellectual Property Rights issues?
- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

1.2. How will the data be collected, observed or generated?

Questions you might want to consider

- What standards, methodologies or quality assurance processes will you use?
- How will you organize your files and handle versioning?

SNSF says...

Explain **how the data will be collected, observed or generated**. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated measurements, data recording standards, usage of controlled vocabularies, data entry validation, data peer review, etc. Discuss **how the data management will be handled during the project**, mentioning for example naming conventions, version control and folder structures. (This relates to the *FAIR Data Principle R1*)

“Experimental records that aren’t being digitized
account for 17% loss of all research data

T. H. Vines et al, Cell Current Biology 2014

Data management software

Will you use animal or human samples?

For Humans...



...and Animals



Electronic Laboratory Notebooks (ELN)



User friendly ***

Easy set-up and use

Open source (MPL)

Unlimited project users

Free 10GB storage

Workflow visualization



User friendly **

Easy set-up and use

Open source (MIT)

Single project user ?

Free 2GB storage

Molecular biology suite



User friendly *

Complicated to use

Not open source

Single project user

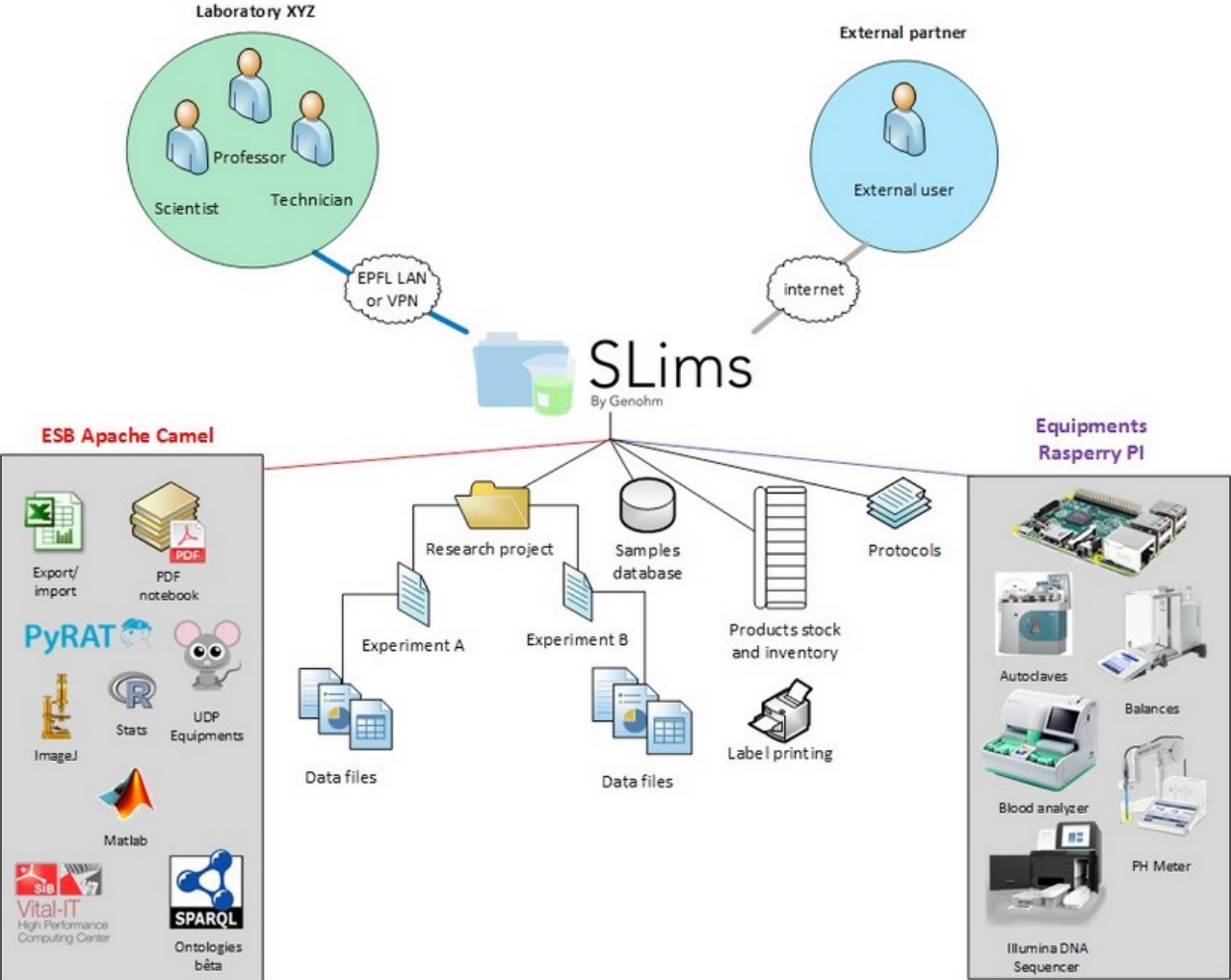
Free 25MB storage

GraphPad Prism
interface

Protocols sharing



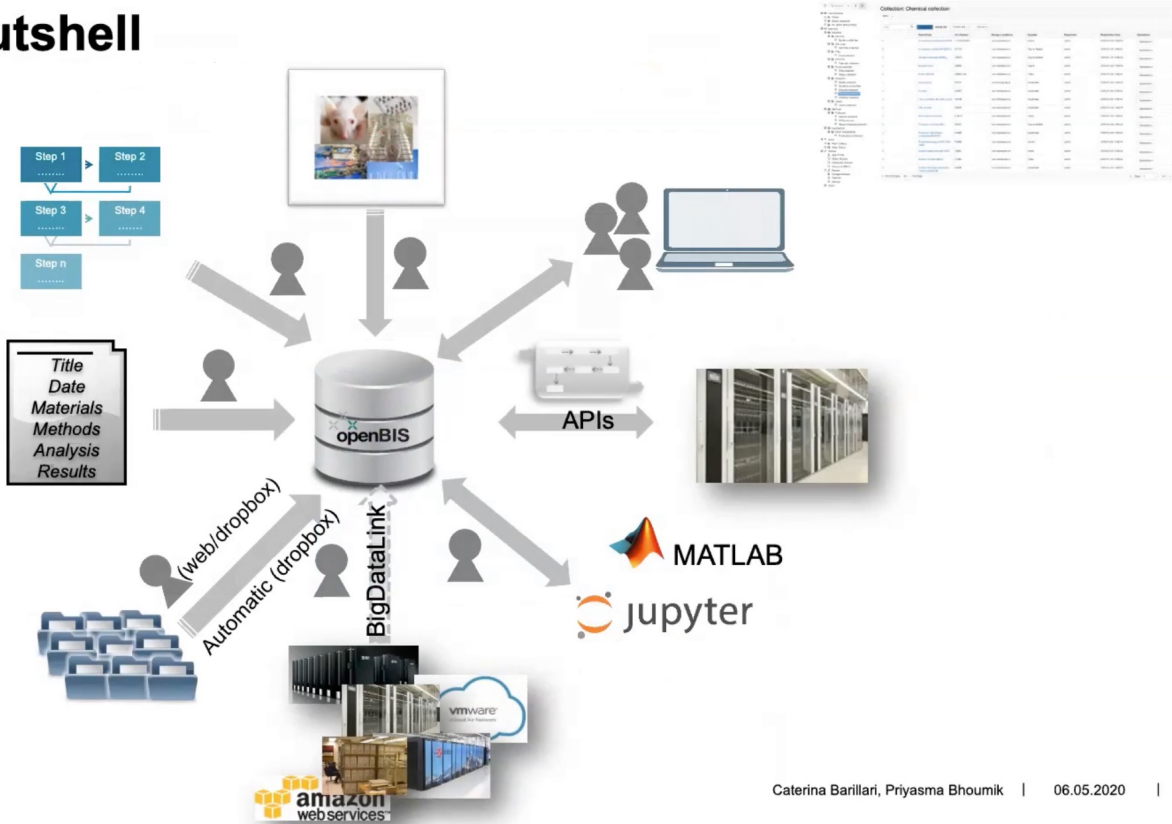
Laboratory Information Management System (LIMS)



SLims is a customizable platform providing LIMS, ELN, sample and order management features!

Laboratory Information Management System (LIMS)

openBIS in a nutshell



Data entry validation

Text file

```
GeneNames_Excel_Change.txt — Edited
ENSG00000180096 SEPT1
ENSG00000168385 SEPT2
ENSG00000100167 SEPT3
ENSG00000108387 SEPT4
ENSG00000184702 SEPT5
ENSG00000125354 SEPT6
ENSG00000122545 SEPT7
ENSG00000164402 SEPT8
ENSG00000184640 SEPT9
ENSG00000186522 SEPT10
ENSG00000138758 SEPT11
ENSG00000140623 SEPT12
ENSG00000154997 SEPT14
ENSG00000145416 MARCH1
ENSG00000099785 MARCH2
ENSG00000173926 MARCH3
ENSG00000144583 MARCH4
ENSG00000198060 MARCH5
ENSG00000145495 MARCH6
ENSG00000136536 MARCH7
ENSG00000165406 MARCH8
ENSG00000139266 MARCH9
ENSG00000173838 MARCH10
ENSG00000183654 MARCH11
ENSG00000173077 DEC1
```

Gene names

Imported into Excel

| | A | B | C | D |
|----|-----------------|---|--------|---|
| 1 | ENSG00000180096 | | Sep.01 | |
| 2 | ENSG00000168385 | | Sep.02 | |
| 3 | ENSG00000100167 | | Sep.03 | |
| 4 | ENSG00000108387 | | Sep.04 | |
| 5 | ENSG00000184702 | | Sep.05 | |
| 6 | ENSG00000125354 | | Sep.06 | |
| 7 | ENSG00000122545 | | Sep.07 | |
| 8 | ENSG00000164402 | | Sep.08 | |
| 9 | ENSG00000184640 | | Sep.09 | |
| 10 | ENSG00000186522 | | Sep.10 | |
| 11 | ENSG00000138758 | | Sep.11 | |
| 12 | ENSG00000140623 | | Sep.12 | |
| 13 | ENSG00000154997 | | Sep.14 | |
| 14 | ENSG00000145416 | | Mar.01 | |
| 15 | ENSG00000099785 | | Mar.02 | |
| 16 | ENSG00000173926 | | Mar.03 | |
| 17 | ENSG00000144583 | | Mar.04 | |
| 18 | ENSG00000198060 | | Mar.05 | |
| 19 | ENSG00000145495 | | Mar.06 | |

Dates!

Poor data layout

Poor data layout

| S1Sh.cuo | | | | |
|----------|-----------|--------|--------|---|
| | A | B | C | D |
| 1 | | Group1 | Group2 | |
| 2 | | Day 0 | | |
| 3 | Sodium | 139 | 142 | |
| 4 | Potassium | 3.3 | 4.8 | |
| 5 | Chloride | 100 | 108 | |
| 6 | BUN | 18 | 18 | |
| 7 | Creatine | 1.2 | 1.2 | |
| 8 | Uric acid | 5.5* | 6.2* | |
| 9 | | Day 7 | | |
| 10 | Sodium | 140 | 146 | |
| 11 | Potassium | 3.4 | 5.1 | |
| 12 | Chloride | 97 | 108 | |

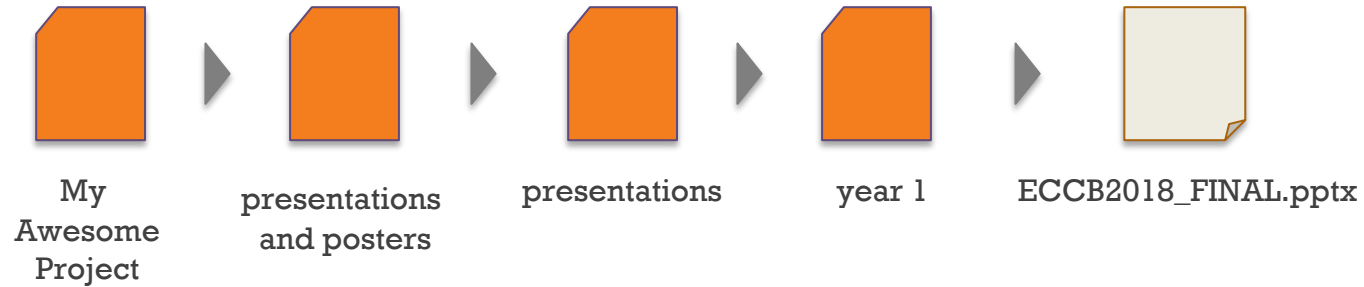
Improved data layout

Table-2_blood-population-1.csv

| Parameter | Day | Control | Treatment | Units | P value |
|-----------|-----|---------|-----------|-------|---------|
| Potassium | 0 | 55 | 120 | mg/l | 0.8 |
| NaCl | 0 | 3.8 | 3 | mg/l | 0.7 |
| Zinc | 0 | 4 | 8 | mg/l | 0.84 |
| Uric acid | 0 | 12 | 20 | mg/l | 0.88 |
| Potassium | 2 | 123 | 99 | mg/l | 0.75 |
| NaCl | 2 | 5 | 2.8 | mg/l | 0.77 |
| Zinc | 2 | 6 | 14 | mg/l | 0.8 |
| Uric acid | 2 | 12 | 15 | mg/l | 0.74 |

Data organization – folders structure

Create directory hierarchy in advance
...and make it as simple as possible



Data organization – folders structure



MAP

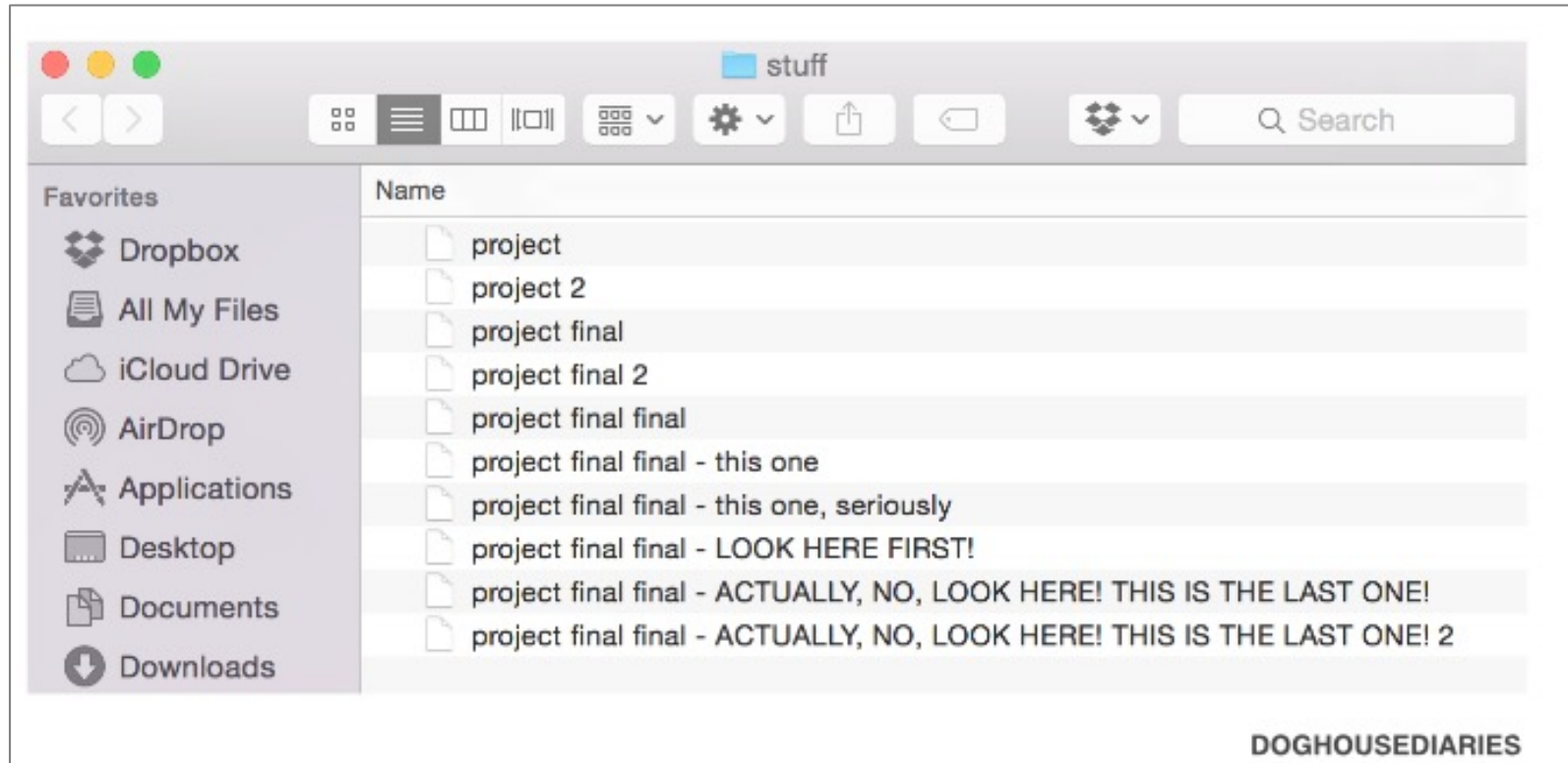


communications



20180909_MAP_ECCB18presentation.pptx

Data organization – file naming



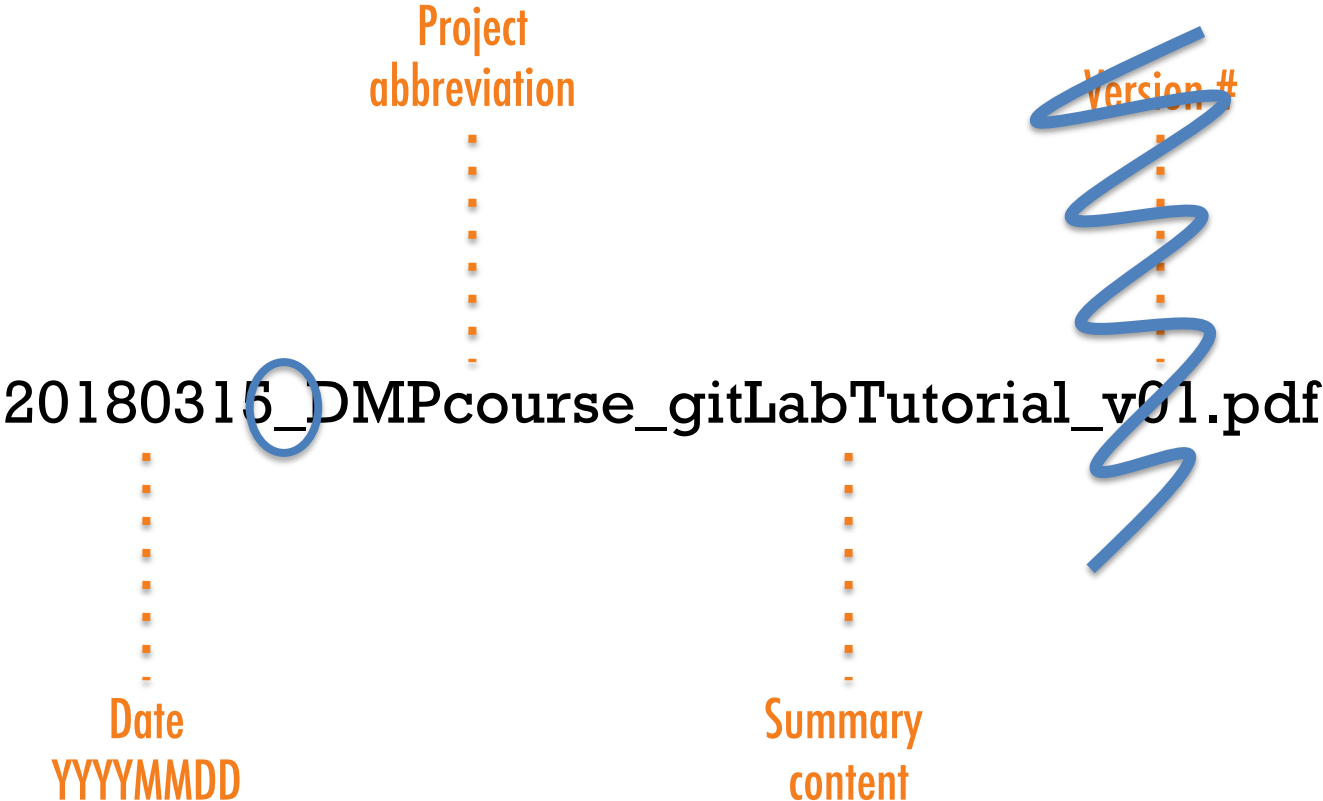
Source: thedoghousediaries.com/5964

Data organization – file naming

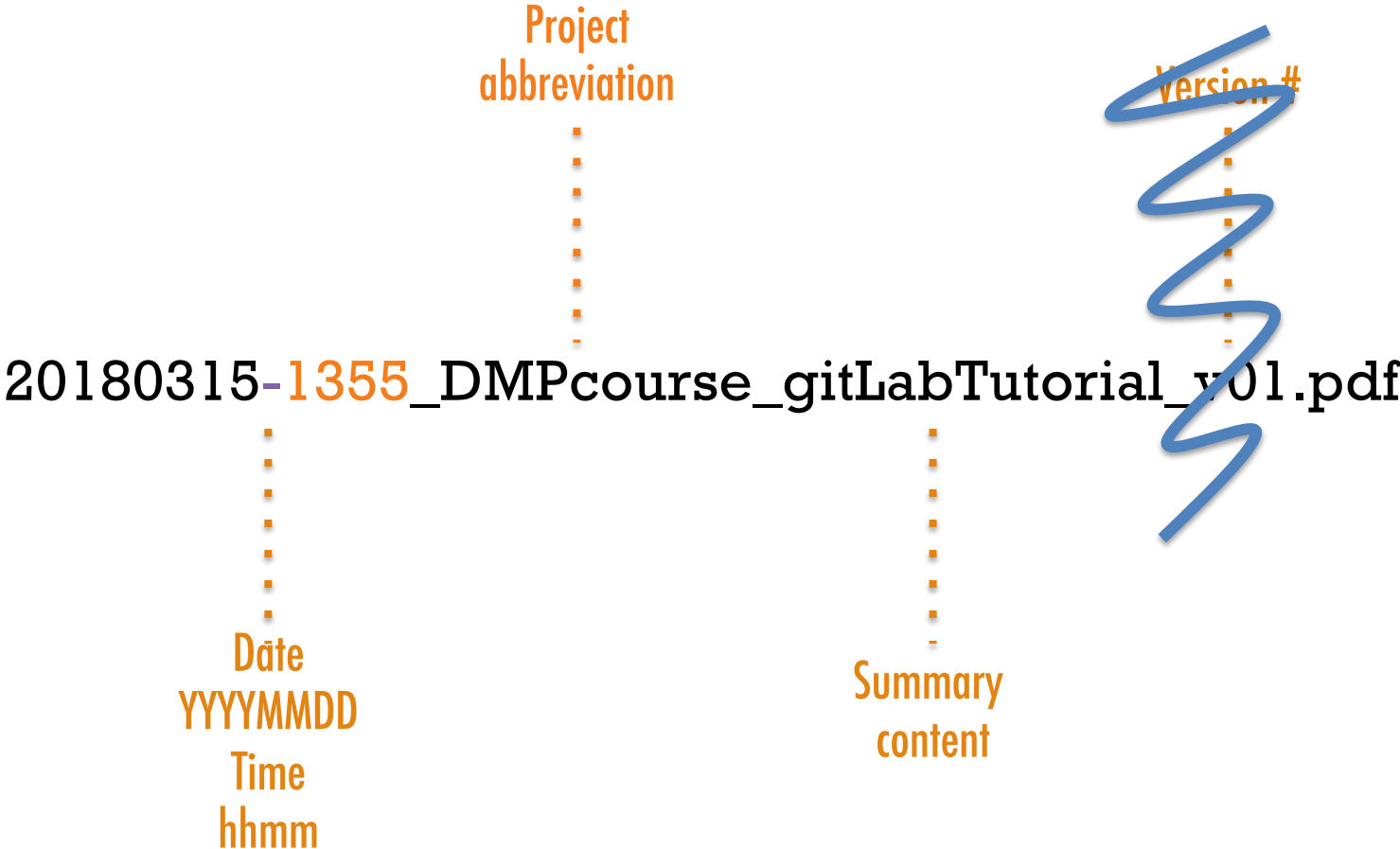
Rule of thumb in file naming

- Keep filenames short, but meaningful
- Do not use spaces
- Avoid special characters `?!+%&#()<>${}*`
- Use - _ or capital letters **toDelimitWords**
- Include numbers in at least two-digit format , i.e. 01-99
- For dates use this format: YYYYMMDD

Data organization – file naming



Data organization – file naming



Version control

Use **Version Control** systems to keep track of all modifications done during the project lifetime

Version control of your documents

Real-time collaborative tools



Version control of your documents

The screenshot shows a Google Docs interface for a document titled "Data Management Plan (DMP)". The document content includes a main title, a subtitle, and several sections of text. On the right side, a "Version history" sidebar is open, displaying a list of document versions with their respective dates, times, and authors.

Data Management Plan (DMP)
Texts needed for the automatic Word file creation

Data that will be collected / generated
A few lines for each potential category

This will be the general intro:
[Questions you might want to consider: What type and format of data will you collect/generate?]


Genomic data:
Genomics data from genome-wide association studies (GWAS) will be stored in PLINK binary format. Processed exome sequencing data will be stored in VCF (variant call format). Other genomic sequencing data, for example whole genome bisulfite sequencing (WGBS) and ChIP-Seq will be stored as raw sequence (FASTQ) and alignments stored as compressed binary (BAM) files.

Proteomic data:
Proteomic data will include peptide sequences that will be produced using LC-MS/MS, and analyzed using the following search algorithms: *[Please specify. For example: NCBI's OMSSA, the GPM's X!Tandem, X!Hunter, K-score]*. All data will be quantified using spectral counting.

Version history

Only show named versions

September 2017

- September 27, 12:16 PM** 
Current version
 - All anonymous users
 - Anastasia Chasapi
- add data type texts
September 8, 1:01 PM
 - All anonymous users
 - Anastasia Chasapi
- corrections on metadata
September 7, 7:18 PM
 - All anonymous users
 - Anastasia Chasapi
- Transcriptomics repo
September 6, 1:43 PM
 - Anastasia Chasapi
- add ELN
September 5, 6:25 PM
 - All anonymous users
 - Anastasia Chasapi

August 2017

- add instructions
August 31, 5:20 PM
 - All anonymous users

Show changes

Version control of your documents

Google Terms of Service

Rights

This license allows Google to:

- host, reproduce, distribute, communicate, and use your content – for example, to save your content on our systems and make it accessible from anywhere you go
- publish, publicly perform, or publicly display your content, if you've made it visible to others
- modify your content, such as reformatting or translating it
- sublicense these rights to:
 - other users to allow the services to work as designed, such as enabling you to share photos with people you choose
 - our contractors who've signed agreements with us that are consistent with these terms, only for the limited purposes described in the [Purpose](#) section below

Where is your data stored?

Locally?

On servers in Switzerland?

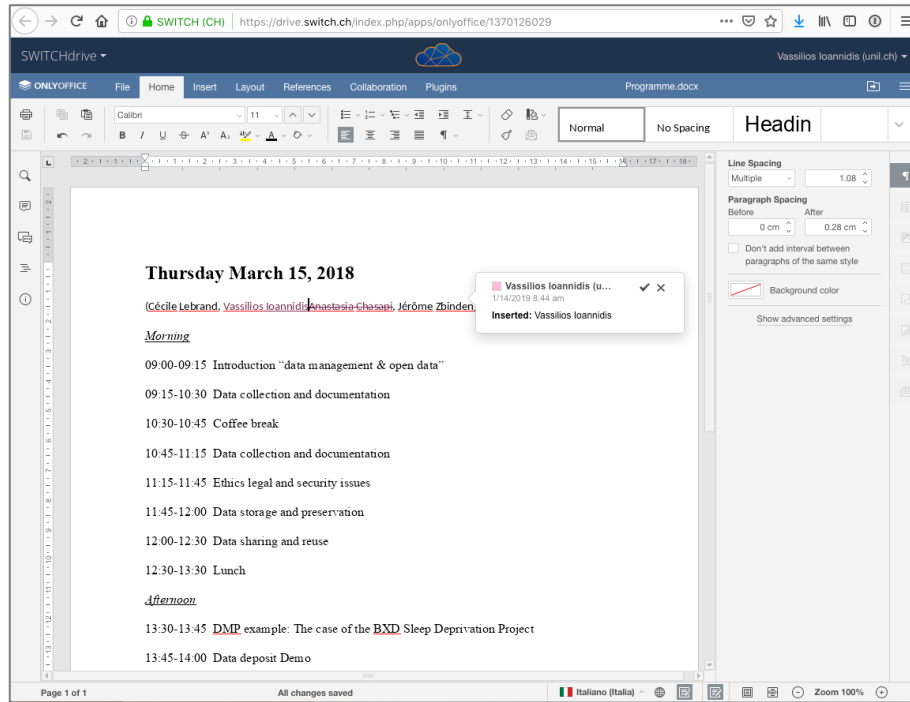
On servers elsewhere?

Check with your institution for clinical data!

SWITCH collaborative tools - OnlyOffice

OnlyOffice integration allows many collaboration functions

- Real-time collaboration on shared documents - work with others on a document and track all changes live
- No local software installation required
- Access via desktop client, browser or mobile application



SWITCH collaborative tools - SWITCHdrive

SWITCHdrive

Store, synchronise, share and edit files quickly and reliably - all in the secure SWITCH cloud.

Services at a glance

- 100 GB of storage space per user
- Online file storage
- Synchronisation of file across several devices
- File sharing and collaboration
- Access via desktop client, browser or mobile app
- SWITCH cloud service

<https://www.switch.ch/drive/>

SWITCH collaborative tools – sharing (internal)

The screenshot displays the SWITCHdrive web interface. The browser address bar shows the URL `https://drive.switch.ch/index.php/apps/files/?dir=/&fileid=53851937`. The interface includes a sidebar on the left with navigation options: All files, Favourites, Shared with you, Shared with others, and Shared by link. The main content area shows a folder named 'DMP-course' with a size of 88.8 MB. A red circle highlights the 'Shared' status icon next to the folder name. A blue arrow points from this icon to the sharing options panel on the right. The sharing panel is titled 'DMP-course' and shows the folder size and age. It has tabs for 'Activities', 'Comments', and 'Sharing'. Under the 'Sharing' tab, there are sections for 'User and Groups' and 'Public Links'. The 'User and Groups' section shows a user named 'Cécile Lebrand (unil.ch)' with options to 'notify by email' and 'can share'. The 'Public Links' section has a text input field for sharing with users, groups, or remote users, and a 'can edit' checkbox.

SWITCH collaborative tools – sharing (external)

The image shows two overlapping windows from the SWITCHdrive interface. The left window is a modal titled "Create link share: /DMP-course". It contains the following fields and options:

- Link name:** A text input field containing "DMP-course link (2)".
- Permissions:** Three radio button options:
 - Download / View:** Selected. Description: "Recipients can view or download contents."
 - Download / View / Upload:** Description: "Recipients can view, download, edit, delete and upload contents."
 - Upload only (File Drop):** Description: "Receive files from multiple recipients without revealing the contents of the folder."
- Password:** A text input field with the placeholder "Choose a password".
- Expiration:** A text input field with the placeholder "Choose an expiration date".
- Buttons:** "Cancel" and "Share".

The right window shows a file sharing card for "DMP-course" (88.8 MB, 3 days ago). The "Sharing" tab is active, showing "User and Groups" and "Public Links" sub-tabs. A blue arrow points to the "Public Links" sub-tab. Below the sub-tabs is a search input field with the placeholder "Share with users, groups or remote use". Underneath, the user "Cécile Lebrand (unil.ch)" is listed with the following sharing settings:

- notify by email
- can share
- can edit

Version control of your code



git



Next SIB Git course: [Version Control with Git, 2-4.10.2023](#)

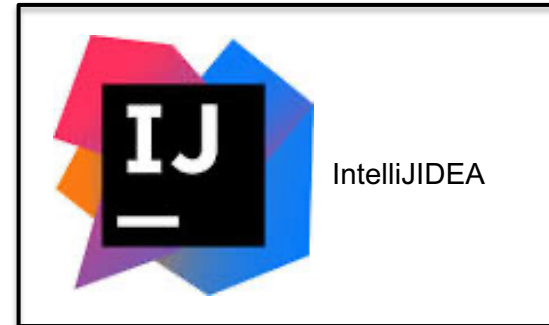
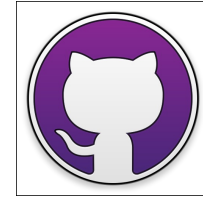
Version control of your code

The screenshot shows a GitLab commit page for a commit with ID `cb153dc6` authored 4 days ago by Vassilios Ioannidis. The commit message is "fixed raxml test. chrome is now used by default". The diff view shows changes to the file `tests/raxml/raxml.php`, with 19 additions and 58 deletions. The code changes are as follows:

```
... .. @@ -30,6 +30,7 @@ use FacebookWebDriver\Remote\RemoteWebDriver;
30 30 use FacebookWebDriver\WebDriverSelect;
31 31 use FacebookWebDriver\WebDriverBy;
32 32 use FacebookWebDriver\WebDriverExpectedCondition;
33 33 + use FacebookWebDriver\Remote\DesiredCapabilities;
34
35 35 class RaxmlTesting extends PHPUnit\Framework\TestCase
... .. @@ -57,11 +58,19 @@ class RaxmlTesting extends PHPUnit\Framework\TestCase
57 58 {
58 59     # Select the browser to be used for the tests
59 60     # firefox or chrome
60 60 - $browser_type = 'firefox';
61 61 + // Important: use chrome, as the 'setCapability("overlappingCheckDisabled",true)' does not work with firefox on this version
62 62 + // we need to disable the 'overlapping check' because of the following error:
63 63 + //FacebookWebDriver\Exception\UnknownServerException:
64 64 + //Element <button class="btn btn-primary btn-sm" type="submit"> is not clickable
65 65 + //at point (632,4916534423828,972) because another element
66 66 + //<nav class="navbar navbar-info bg-dark fixed-bottom navbar-expand-md"> obscures it
67 67 +
68 68 + $browser_type = 'chrome';
69
70 70 // setup and initialization of the webdriver
71 71 - $capabilities = array(WebDriverCapabilityType::BROWSER_NAME => $browser_type);
72 72 + $capabilities = DesiredCapabilities::chrome();
73 73 + $capabilities->setCapability("overlappingCheckDisabled", true);
74
75 75 // In case a test cannot obtain a webdriver object due to a 30 s timeout,
76 76 // we terminate gracefully without throwing an error but still display
77 77 // a warning message. Needed for ChromeDriver.
... .. @@ -100,6 +109,7 @@ class RaxmlTesting extends PHPUnit\Framework\TestCase
100 109
101 110 public function atest_empty_job()
102 111 {
112 112 + //DO WE NEED THIS AS NO CHECK IS DONE BROWSER SIDE?
113 113 // transfer the function name to local variable to be used in screenshot method
114 114 $this->errorTest = __FUNCTION__;
115 115 $this->driver->get($this->initial_url);
... .. @@ -113,7 +123,7 @@ class RaxmlTesting extends PHPUnit\Framework\TestCase
113 123 }
114 124
115 125
```

Version control of your code

For the command line haters out there...



Version control – dynamic report generation

R Markdown Notebook

(Demo)

- Open Rstudio
- Create a new project
- Create a new Markdown file – MS Word
- Use the given template and 'Knitr'

Version control – dynamic report generation

Jupyter

R Markdown Notebook

Knitr

MATLAB Live Editor

Version control

General advice for reproducible research:

- *Don't do things by hand*
- *Don't point and click*
- *Document software environment*

Version control with GUI programs

ExPASy SIB Bioinformatics Resource Portal **LALIGN** Home | Contact

Find multiple matching subsegments in two sequences

This is William Pearson's *lalign* program. A manual page for this program is available [here](#). The *lalign* program implements the algorithm of Huang and Miller, published in *Adv. Appl. Math.* (1991) 12:337-357.

This program is part of the FASTA package of sequence analysis program.

Usage: Paste your two sequences in one of the supported **formats** into the sequence fields below and press the "Run lalign" button.
Make sure that both format buttons (next to the sequence fields) shows the correct formats

Choose the alignment method: local (default) global global without end-gap penalty

Number of reported sub-alignments: E-value threshold: (default 10.0)

Scoring matrix:

Opening gap penalty: (default -12) Extending gap penalty: (default -2)

First sequence title (optional):

Input sequence format:

1st Query sequence: or ID or AC or GI (see above for valid formats)

Second sequence title (optional):

Input sequence format:

2nd Query sequence: or ID or AC or GI (see above for valid formats)

SIB Swiss Institute of Bioinformatics | Contact △ Back to the Top

ExPASy SIB Bioinformatics Resource Portal **LALIGN**

Readseq version 2.1.30 (12-May-2010) Readseq version 2.1.30 (12-May-2010)

lalign output for THIO_RAT vs. THIO_HUMAN

[EMBnet-Server] Date: Tue Jan 15 9:16:30 2019

```
# bin/lalign36 -E 10.0 -f -12 -g -2 19596.1.seq 19596.2.seq -J -K 3
LALIGN finds non-overlapping local alignments
version 36.3.5e Nov, 2012(preload8)
Please cite:
X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

Query: 19596.1.seq
1>>>sp|P11232|THIO_RAT (Txn)RecName: Full=Thioredoxin; Short=Trx;[Ra
Library: 19596.2.seq
105 residues in 1 sequences

Statistics: (shuffled [500]) MLE statistics: Lambda= 0.1799; K=0.0581
statistics sampled from 1 (1) to 500 sequences
Threshold: E() < 10 score: 23
Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010)
Parameters: BL50 matrix (15;-5), open/ext: -12/-2
Scan time: 0.000

>>sp|P10599|THIO_HUMAN (TXN)RecName: Full=Thioredoxin; S (105 aa)
Waterman-Eggert score: 641; 170.5 bits; E(1) < 5.3e-48
89.4% identity (96.2% similar) in 104 aa overlap (1-104:1-104)

      10      20      30      40      50      60
sp|P11 MVKLIESKEAFQEALAAAGDKLVVDFSATWCGPCKMIKPFHSLCDKYSNVVFLVDVD
::: :::: :::::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
sp|P10 MVKQIESKTAFAQEALDAAGDKLVVDFSATWCGPCKMIKPFHSLSEKYSNVIFLEVDVD
      10      20      30      40      50      60

      70      80      90     100
sp|P11 DCQDVAADCEVKCMPTFQFYKKGQKVGFEFGANKEKLEATITEF
::::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
sp|P10 DCQDVAADCEVKCMPTFQFYKKGQKVGFEFGANKEKLEATITEF
```

Version control with GUI programs

Find multiple matching subsegments in two sequences

This is William Pearson's *lalign* program. A manual page for this program is available [here](#). The *lalign* program implements the algorithm of Huang and Miller, published in *Adv. Appl. Math.* (1991) 12:337-357.

This program is part of the FASTA package of sequence analysis program.

Usage: Paste your two sequences in one of the supported **formats** into the sequence fields below and press the "Run lalign" button.
Make sure that both format buttons (next to the sequence fields) shows the correct formats

Choose the alignment method: local (default) global global without end-gap penalty

Number of reported sub-alignments: E-value threshold: (default 10.0)

Scoring matrix: Opening gap penalty: (default -12) Extending gap penalty: (default -2)

First sequence title (optional): Input sequence format:

1st Query sequence: or ID or AC or GI (see above for valid formats)

Second sequence title (optional): Input sequence format:

2nd Query sequence: or ID or AC or GI (see above for valid formats)

SIB Swiss Institute of Bioinformatics | Contact Home | Contact

LALIGN

Readseq version 2.1.30 (12-May-2010) Readseq version 2.1.30 (12-May-2010)

lalign output for THIO_RAT vs. THIO_HUMAN

[EMBnet-Server] Date: Tue Jan 15 9:16:30 2019

```
# bin/lalign36 -E 10.0 -f -12 -g -2 19596.1.seq 19596.2.seq -J -K 3
LALIGN finds non-overlapping local alignments
version 36.3.5e Nov, 2012(preload8)
Please cite:
X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

Query: 19596.1.seq
1>>>sp|P11232|THIO_RAT (Txn)RecName: Full=Thioredoxin; Short=Trx;[Ra
Library: 19596.2.seq
105 residues in 1 sequences

Statistics: (shuffled [500]) MLE statistics: Lambda= 0.1799; K=0.0581
statistics sampled from 1 (1) to 500 sequences
Threshold: E() < 10 score: 23
Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010)
Parameters: BL50 matrix (15;-5), open/ext: -12/-2
Scan time: 0.000

>>>sp|P10599|THIO_HUMAN (TXN)RecName: Full=Thioredoxin; S (105 aa)
Waterman-Eggert score: 641; 170.5 bits; E(1) < 5.3e-48
89.4% identity (96.2% similar) in 104 aa overlap (1-104:1-104)

      10      20      30      40      50      60
sp|P11 MVKLVESKEAFQEALAAAGDKLVVDFSATWCGPCKMIKPFHSLCDKYSNVVFLVVDV
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
sp|P10 MVKQIESKTAFAQEALDAAGDKLVVDFSATWCGPCKMIKPFHSLSEKYSNVVFLVVDV
      10      20      30      40      50      60

      70      80      90     100
sp|P11 DCQDVAADCEVKCMPTFQFYKKGQKVGDFSGANKERLEATITEF
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
sp|P10 DCQVASEGCVKCMPTFQFYKKGQKVGDFSGANKERLEATITEF
```



3. ACTIVE PHASE

b. Data Documentation & Metadata

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?
- 2.3 How will you handle copyright and Intellectual Property Rights issues?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

1.3. What documentation and metadata will you provide with the data?

Questions you might want to consider

- What information is required for users (computer or human) to read and interpret the data in the future?
- How will you generate this documentation?
- What community standards (if any) will be used to annotate the (meta)data

SNSF says...

Describe all types of **documentation (README files, metadata, etc.)** you will provide to help secondary users to understand and reuse your data. **Metadata** should at least include basic details allowing other users (computer or human) to find the data. This includes at least a **name** and a **persistent identifier** for each file, the **name of the person who collected or contributed to the data**, the **date of collection** and the **conditions to access the data**. Furthermore, the documentation may include details on the **methodology used**, information about the **performed processing and analytical steps**, **variable definitions**, references to vocabularies used, as well as units of measurement. Wherever possible, the documentation should follow existing community standards and guidelines. Explain how you will prepare and share this information. (This relates to the *FAIR Data Principles* I1, I2, I3, R1, R1.2 & R1.3)

Documentation for Research Data at various levels

Project level

what the study set out to do, how it contributes new knowledge to the field, what the research questions/hypotheses were, what methodologies were used, what sampling frames were used, what instruments and measures were used, etc.

File or database level

how all the files (or tables in a database) that make up the dataset relate to each other; what format they are in; whether they supercede or are superceded by previous files. A readme.txt file is the classic way of accounting for all the files and folders in a project.

Variable or item level

the key to understanding research results is knowing exactly how an object of analysis came about. Not just, for example, a variable name at the top of a spreadsheet file, but the full label explaining the meaning of that variable in terms of how it was operationalised.

Documentation to provide with the data

Examples

- laboratory notebooks & experimental protocols
- questionnaires, codebooks, data dictionaries
- software syntax and output files
- information about equipment settings & instrument calibration
- database schema
- methodology reports
- provenance information about sources of derived or digitised data

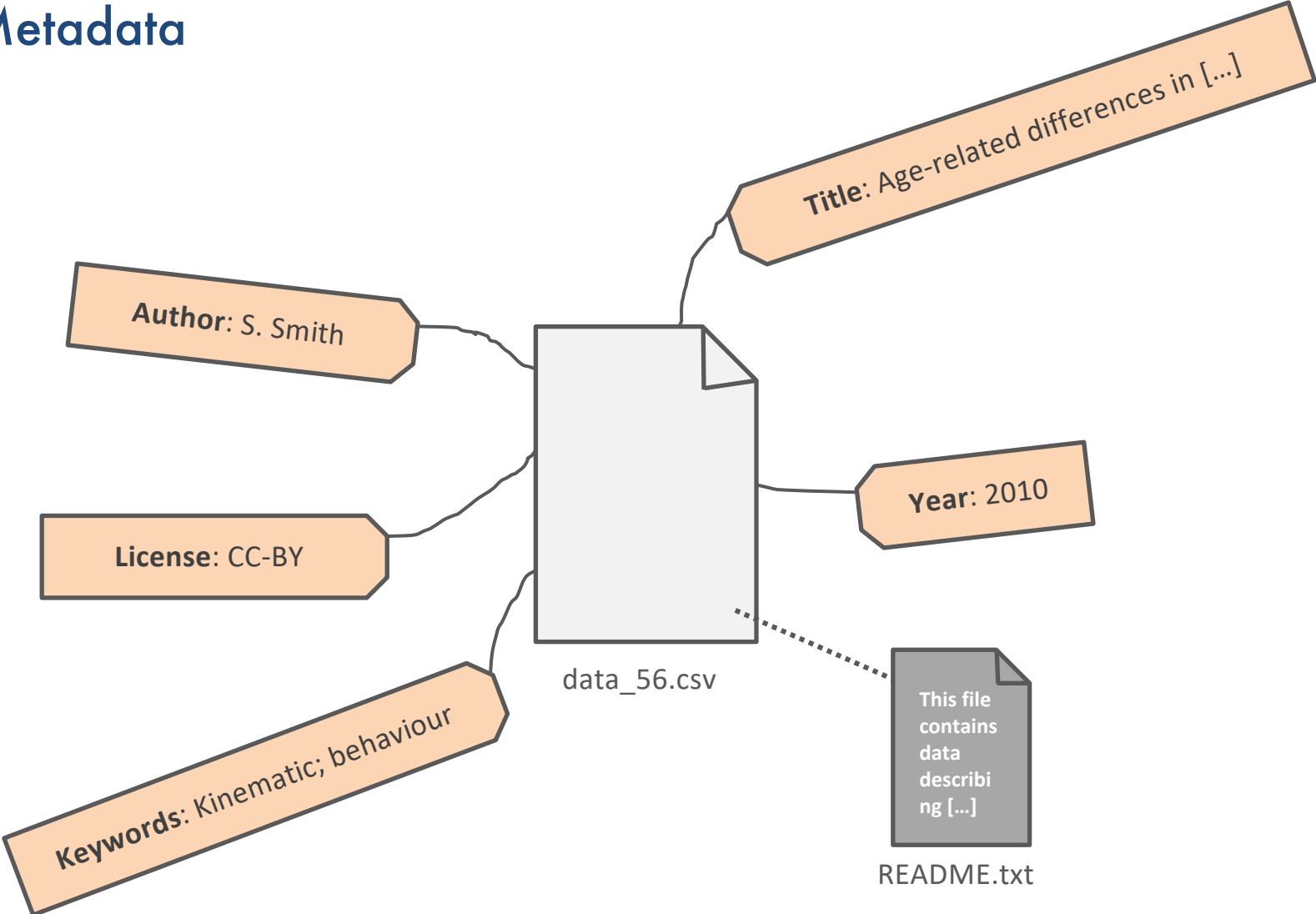
What does metadata mean?

- The term metadata is commonly defined as "data about data", **information that describes or contextualises** the data.
- The difference between documentation and metadata is that the first is meant to be read by humans and the second **implies computer-processing** (though metadata may also be human-readable).
- Documentation is sometimes considered a form of metadata, because it is information about data, and when it is very structured it can be. The importance of metadata lies in the **potential for machine-to-machine interoperability**, providing the user with added functionality, or 'actionable' information.

Sample DataCite XML (HTML tags)

```
<resource xmlns="http://datacite.org/schema/kernel-2.1" xmlns:xsi="http://www.w3.org/2001/meta/ker
<identifier identifierType="DOI">10.1594/WDCC/CCSRNIES_SRES_B2</identifier>
<creators>
<creator><creatorName>Toru, Nozawa</creatorName></creator>
<creator><creatorName>Utor, Awazon</creatorName>
      <nameIdentifier nameIdentifierScheme="ISNI">1422 4586 3573 0476</nameIdentifier></creator
</creators>
<titles>
<title>National Institute for Environmental Studies and Center for Climate System Research Japan</title
<title titleType="Subtitle">A survey</title>
</titles>
<publisher>World Data Center for Climate (WDCC)</publisher>
<publicationYear>2004</publicationYear>
<subjects>
<subject>Earth sciences and geology</subject>
<subject subjectScheme="DDC">551 Geology, hydrology, meteorology</subject>
</subjects>
<contributors>
<contributor contributorType="DataManager">
<contributorName>PANGAEA</contributorName>
</contributor>
<contributor contributorType="ContactPerson >>
```

Metadata



Poll 3: Metadata

What metadata would you need to describe a dataset in your field?

Select the 3 answers that best correspond to your situation

- Identifier (DOI), Creator, Title
- Publisher, Publication Year
- Resource Type
- Subject, description
- Contributor
- Related Identifier
- Geolocation
- Funding Reference
- Data size, Format, Version
- Rights (licenses)

Join at:
vevox.app

ID:
193-360-850



[nature](#) > [world view](#) > [article](#)

WORLD VIEW | 05 September 2022

Without appropriate metadata, data-sharing mandates are pointless



Funders and investigators must demand appropriate metadata standards to take data from foul to FAIR.

[Mark A. Musen](#) 



Know the relevant standards in your community

- Many communities have developed specific guidelines for reporting certain kinds of data
- Check journal guidelines to see what is required.
- Can help you format your data in a useful manner

 <https://biosharing.org>

- Browse information on over 600 reporting standards
- Find standards that are relevant to your type of data

About ▾

News ▾

Events ▾

Services ▾

Guidance ▲

Briefing Papers

How-to Guides

Case Studies

Policy Analysis ▾

Metadata ▲

Disciplinary
Metadata

Curation Lifecycle
Model

Data Management
Plans

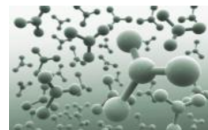
Research ▾

Publications ▾

Search by Discipline



Social Science & Humanities



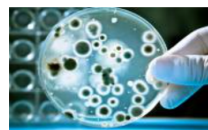
Physical Science



General Research Data



Earth Science



Biology

Search by Resource Type

Metadata Standards

Specifications for the minimum information that should be collected about research data in order for it to be re-used.

Profiles and Extensions

Standards that have been adapted for use in particular types of repositories, or for particular types of data.

Use cases

Institutional repositories and data portals using standards to determine which metadata should be collected upon data deposit.

Tools

Software that has been developed to capture or store metadata conforming to a specific standard.

Data Documentation Initiative (DDI)

International standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences.



[Learn](#) ▾ [Products](#) ▾ [Membership](#) ▾ [Events](#) ▾ [Publications](#) ▾ [About](#) ▾

Document, Discover and Interoperate

The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding, interpretation, and use -- by people, software systems, and computer networks.

Why Use DDI?

- ✓ Generate interactive codebooks
- ✓ Implement data catalogs
- ✓ Build question banks

- ✓ Create concordance mappings
- ✓ Harmonize and compare data
- ✓ Manage longitudinal data sets

Featured DDI Adopters



[The Dataverse Project](#)

The Dataverse Project is a community and an open source web application to share, preserve, cite, explore and analyze research data. It facilitates making data available to others, and allows you to replicate others work. Researchers, data authors, publishers, data distributors, and... [read more](#)



DataCite Metada Schema

Allow any kind of data to be understood and reused by other members of the research group and add contextual value to the datasets for future publishing and data sharing.

| Mandatory | Recommended | Optional |
|----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| <p>Identifier Creator Title Publisher Publication Year Resource Type</p> | <p>Subject Contributor Date Related Identifier Description Geolocation FundingReference</p> | <p>Language Alt. Identifier Size Format Version Rights</p> |

Readme file

Help ensure that your data can be correctly interpreted and reanalyzed

Two ways to include a README :

- Provide a separate README for each individual data file
- Submit one README for the data package as a whole

Dryad recommend that a README be a plain text file containing the following:

- for each filename, a short description of **what data it includes**, optionally describing the **relationship** to the tables, figures, or sections within the accompanying publication
- for **tabular data**: definitions of column headings and row labels; data codes (including missing data); and measurement units
- any data **processing steps**, especially if not described in the publication, that may affect interpretation of results
- a description of what associated datasets are stored **elsewhere**, if applicable
- whom to contact with questions
- If text formatting is important for your README, PDF format acceptable.

Readme file template

RESEARCH DATA MANAGEMENT SERVICE GROUP

Comprehensive Data Management Planning & Services

Guide to writing "readme" style metadata

A readme file provides information about a data file and is intended to help ensure that the data can be correctly interpreted, by yourself at a later date or by others when sharing or publishing data. **Standards-based metadata** is generally preferable, but where no appropriate standard exists, for internal use, writing "readme" style metadata is an appropriate strategy.

 Want a template? **Download one** and adapt* it for your own data!

- [Best practices](#)
- [Recommended content](#)
 - [General information](#)
 - [Data and file overview](#)
 - [Sharing and access information](#)
 - [Methodological information](#)
 - [Data-specific information](#)
- [References](#)
- [Related information](#)

Readme XML or text file tool: DataCite Metadata Generator

XML file automatically generated using the [DataCite Metadata Generator](#) after filing the form requesting intrinsic metadata.

Readme XML file ensures compatibility with international standards and is human as well as machine-readable.

DataCite Metadata Generator - Kernel 4.0

Mandatory Elements

DOI:
[DOI]

Title(s): +
DATASET of Two specific populations of GABAergic neurons originating from the medial and the caudal ganglionic ei
[titleType] ▼

Creator(s): +
Cécile Lebrand
[GIVEN NAME] (optional) [FAMILY NAME] (optional)
[NAME IDENTIFIER] [NAME ID SCHEME]
[IDENTIFIER SCHEME URI] +
University of Lausanne +

Publisher:
[PUBLISHER]

Publication Year:
[YYYY]

Resource Type:
[RESOURCE TYPE] [resourceTypeGeneral] ▼

- Recommended Elements

Subject(s): +
axonal guidance [SUBJECT SCHEME]
[SUBJECT SCHEME URI]
[SUBJECT VALUE URI]

Data Repositories

Open Science Framework

A scholarly commons to connect the entire research cycle



search on figshare

Browse

Upload

Sign up

Log in



About

For researchers

For organizations

Contact us

Log in



DryadLab is a collection of hands-on, educational models for scientific inquiry using real

Submit data now

[How and why?](#)

zenodo

Research. Shared.

Search

Communities

Browse

Upload

Get started

Sign In

Sign Up

Browse for data

Recently published

Popular

Recently published data

deVries MS Data from: The role of feeding diet breadth of sympatric stomatopods
<http://dx.doi.org/10.5061/dryad.5d78>

Botterill-James T, Slince J, Uller T, Cha from: Experimental manipulation suggests within-pair aggression in the social Sociobiology
<http://dx.doi.org/10.5061/dryad.5d78>

Martin SH, Van Belleghem S Data from: genome using topology weighting. G

Faculty of Biology and Medicine at University of Lausanne & Lausanne University Hospital

Recent Uploads

02 October 2009 Dataset Open access

Transient neuronal populations are required to guide callosal axons: a role for semaphorin 3C.

Niquille M ; Garel S ; Mann F ; Hornung JP; et al.

Abstract The corpus callosum (CC) is the main pathway responsible for interhemispheric communication. CC agenesis is associated with numerous human pathologies, suggesting that a range of developmental defects can result in abnormalities in this ...

View

30 June 2014 Journal article Open access

Mutations in Em1 lead to ectopic progenitors and neuronal heterotopia in mouse and human.

Michel Kleiar ; Françoise Phan Dinh Tuy ; Sara Bizzotto ; Cécile Lebrand; et al.

Neuronal migration disorders such as lissencephaly and subcortical band heterotopia are associated with epilepsy and intellectual disability. DCX,PAFAH1B1 and TUBA1A are mutated in these disorders; however, corresponding mouse mutants do not show ...

View

2

Community collection

UNIL | Université de Lausanne

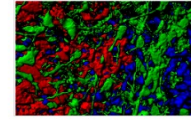
Faculté de biologie et de médecine

Faculty of Biology and Medicine at University of Lausanne & Lausanne University Hospital

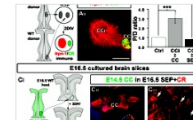
The scientific production in biomedical research of UNIL & CHUV openly available : i) author accepted manuscripts and ii) supplementary files and key datasets accompanying publications.

[Read more](#)

Title:
Faculty of Biology and Medicine at University of Lausanne & Lausanne University Hospital
Curated by:
CLEbrand



This illustration of a coronal corpus callosum section of a mouse embryo...
Rachel Jones 27/10/2009



CC region enriched in glutamatergic CR-positive neurons attracts callo...
Mathieu Niquille 27/10/2009

[figshare](#)
[Zenodo](#)
[OSF](#)
[Dryad](#)

Metadata associated with the dataset for data sharing

Files with Internal Metadata: e.g. microscope images contain a range of metadata (magnification, lens, zoom, gain, etc...).

Metadata for publishing datasets on nonprofit unstructured data repositories:

- XML metadata (machine readability / interoperability) generated after filing the repository submission form. Metadata comprise a persistent identifier, a publication date and conditions of access to (type of license) the dataset.
- Readme file(s) with the dataset in text with more detailed information.

Field specific metadata standards:

[Digital Curation Center.](#)

<https://fairsharing.org/standards/?q>

Metadata for Research Data

Take home message

- Metadata should be as complete as possible, using the standards and conventions of a discipline, and should be machine readable.
- Metadata should always accompany a dataset, no matter where it is stored.
- Many academic disciplines have formalized specific metadata standards.



3. ACTIVE PHASE

c. Data Access Management & Storage

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 2.3 How will you handle copyright and Intellectual Property Rights issues?
- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

2. 2. How will data access and security be managed?

Questions you might want to consider

- What are the main concerns regarding data security, what are the levels of risk and what measures are in place to handle security risks?
- How will you regulate data access rights/permissions to ensure the security of the data?
- How will personal or other sensitive data be handled to ensure safe data storage and transfer?

SNSF says...

If you work with personal or other sensitive data you should outline the **security measures in order to protect the data**. Please **list formal standards** which will be adopted in your study. An example is ISO 27001-Information security management. Furthermore, describe the **main processes or facilities for storage and processing of personal or other sensitive data**. (This relates to the *FAIR Data Principle A1*)

“who has access and with what restrictions”

Security issues

Data loss and/or leak

Measures

- Efficient backup
- Cloud-based storage
- Encryption of sensitive data



Credit: Dave Hill, CC-BY-NC-SA 2.0 Generic. <https://www.flickr.com/photos/dmh650/4031607067/in/gallery-wk170-72157633022909105/>

Security issues

Uncontrolled/Unwanted data access

- Global Malware Threats
- Data Theft or Corruption (virus)

Measures

- Firewall / VPN
- Institutional antivirus (automatic updates)
- OS and applications: applying security patches (regular updates)
- Caution with phishing
- Do not open attachments contained in unsolicited emails

Security issues

Uncontrolled/Unwanted data access

- Global Malware Threats
- Data Theft or Corruption (virus)
- Access rights

Measures

- Reduce permission rights of your files and folders to a strict minimum (in compliance with the group)
- Never change status to rwxrwxrwx
- If files not frequently used or archives: prevent unwanted deletion or modifications.

Example: by removing corresponding access rights: r - - r - - - - -

Personal and sensitive data

Personal data

“any piece of information that someone can use to identify, with some degree of accuracy, a living person”

- A name and surname
- A home address
- An email address
- An identification card number
- Location data
- An Internet Protocol (IP) address

Sensitive data

“a specific set of “special categories” that must be treated with extra security”

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Genetic data
- Biometric data (where processed to uniquely identify someone).

Personal and sensitive data

Data protection at the institution level

- The IT service should provide appropriate protection of hosted data on its technical infrastructure, including the confidentiality and integrity of the data. For information security, the **Institution is (or plan to be) compliant with the following international standards: ISO 27001 - International information security standard.**

Data protection at the group/individual level

- Sensitive data **must be encrypted** by your lab members for storage on NAS, or on USB/external disks or on laptops in case you have to use these supports.

Recommended: (internal) **data sharing** via SWITCHdrive and **data transfer** via SWITCHfilesender.

Collaborative data sharing: SWITCHdrive

SWITCHdrive

Store, synchronise, share and edit files quickly and reliably - all in the secure SWITCH cloud.

Facts at a glance

- Standard 100 GB online storage per user
- Easy collaboration and protected access with SWITCHaai or [SWITCH edu-ID Login](#)
- Synchronise files & folders across multiple devices
- Access via desktop client, browser or mobile application

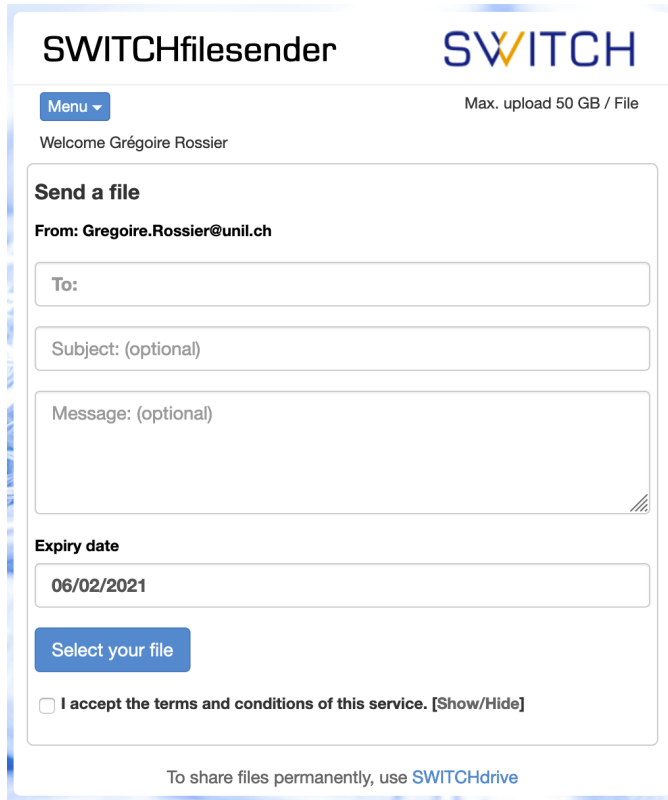


Your data at SWITCH

- All data is stored on SWITCH servers in Switzerland.
- Full compliance with Swiss data protection regulations
- No data and metadata exchange with other Office companies - not even for diagnoses or future extensions.

No backup!

Data transfer: SWITCHfilesender



The screenshot shows the SWITCHfilesender web interface. At the top left is the logo 'SWITCHfilesender' and at the top right is the 'SWITCH' logo. Below the logos, there is a 'Menu' button and a status indicator 'Max. upload 50 GB / File'. A welcome message 'Welcome Grégoire Rossier' is displayed. The main section is titled 'Send a file' and shows the sender's email 'From: Gregoire.Rossier@unil.ch'. There are three input fields: 'To:', 'Subject: (optional)', and 'Message: (optional)'. Below these is an 'Expiry date' field with the value '06/02/2021'. A blue button labeled 'Select your file' is positioned below the expiry date. At the bottom of the form, there is a checkbox and the text 'I accept the terms and conditions of this service. [Show/Hide]'. A footer note at the bottom of the interface states 'To share files permanently, use SWITCHdrive'.

- Send or receive large files by mail notification
- Max 300 GB
- Web based, nothing to install
- Set expiration date
- Invite guests
- More information:
<https://www.switch.ch/services/filesender/>

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 2.3 How will you handle copyright and Intellectual Property Rights issues?
- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

3.1. How will your data be stored and backed-up during the research?

Questions you might want to consider:

- What is your storage capacity and where will the data be stored?
- What are the back-up procedures?

SNSF says...

Please mention what the needs are in terms of **data storage and where the data will be stored**. Please consider that data storage on laptops or hard drives, for example, is risky. **Storage through IT teams is safer**. If external services are asked for, it is important that this does not conflict with the policy of each entity involved in the project, especially concerning the issue of sensitive data. Please specify your **back-up procedure** (frequency of updates, responsibilities, automatic/manual process, security measures, etc.)

Researchers are responsible for the safety,
security and integrity of the data they generate.



Available storage solutions

- *Computer hard drive or external hard drive*
- Network Attached Storage (NAS)
- Remote servers (secured for sensitive data)
- Cloud service solutions (Switch, OneDrive, **Dropbox, GoogleDrive**)

Don't pull all your eggs in one basket!



Credit: By Bryan Tong Minh / CC-BY-2.5 (http://commons.wikimedia.org/wiki/File:Brand_bouwkunde_-_TU_Delft_-_13_Mei_2008.jpg)

Institutional storage

Network Attached Storage NAS

- Access to files from anywhere in the world (with VPN)
- High upload/download speeds
- Very high resistance to hardware and software failures
- Security measures (protection against viruses, loss of data, etc)

Data back-up and safeguarding (UNIL example)

- Incremental backup (snapshot)
- Full daily backup of the data (2 times per day, 90 days retention)
- Replication of the data on a second NAS in another building (every 4 hours)
- Third copy

Prices vary and depend on needs – talk to your institution

Other Swiss storage options: SWITCHengines

SWITCHengines

Dynamic compute and storage services for research and education

Welcome to SWITCHengines!

SWITCHengines provides compute and storage services in the form of virtual machines to researchers, lecturers and IT-services of Swiss universities and related institutions.

Custom-tailored for academic use cases

SWITCHengines was created to expand the boundaries of local infrastructures for research, education and IT-services under maximal considerations of your needs. Check out the following three flavors.

Research

Education

Enterprise

Contact



Konrad Jaggi

Outreach Manager


+41 44 268 15 17

konrad.jaggi@switch.ch

Other Swiss storage options: data@SWITCH

Data@SWITCH

Data@SWITCH
Storage services that meet your needs – all in the secure SWITCH cloud



Short-term and long-term storage for large volumes of data



SWITCHfilesender

«I would like to send large volumes of data easily and securely.»



SWITCHdrive

«I would like to store files, share them with others and work on them together.»



SWITCHengines storage

«I would like to store large volumes of data, e.g. for off-site backups or sharing of datasets.»



SWITCHcoldstor

«I would like to store data in an archive for a longer period of time e.g. for regulatory reasons, or data management purposes.»



SWITCHtube

«I would like to make videos available to my students securely.»



4. CONCLUDING PHASE

a. Data Preservation

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?
- 2.3 How will you handle copyright and Intellectual Property Rights issues?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

3.2. What is your data preservation plan?

Questions you might want to consider

- What procedures would be used to select data to be preserved?
- What file formats will be used for preservation?

SNSF says...

Please specify **which data will be retained, shared and archived after the completion of the project** and the **corresponding data selection procedure** (e.g. long-term value, potential value for re-use, obligations to destroy some data, etc.). Please outline a long-term preservation plan for the datasets beyond the lifetime of the project. In particular, comment on the choice of file formats and the use of community standards. (This relates to the *FAIR Data Principles F2 & R1.3*)

Scientists losing data at a rapid rate

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.

*(image under © 2019 Macmillan Publishers Limited, part of Springer Nature. All Rights Reserved.
Follow the link below to access the complete article.)*

Probability of
finding the data
associated with
a paper
declined by
17% every year

Long term data archiving

Archive collected data on long term storage space for at least 10 years (e.g. UNIL [Directive 4.2, art. 2.4](#)).

Appropriate data archiving is under the responsibility of the Principal investigator Prof. X.

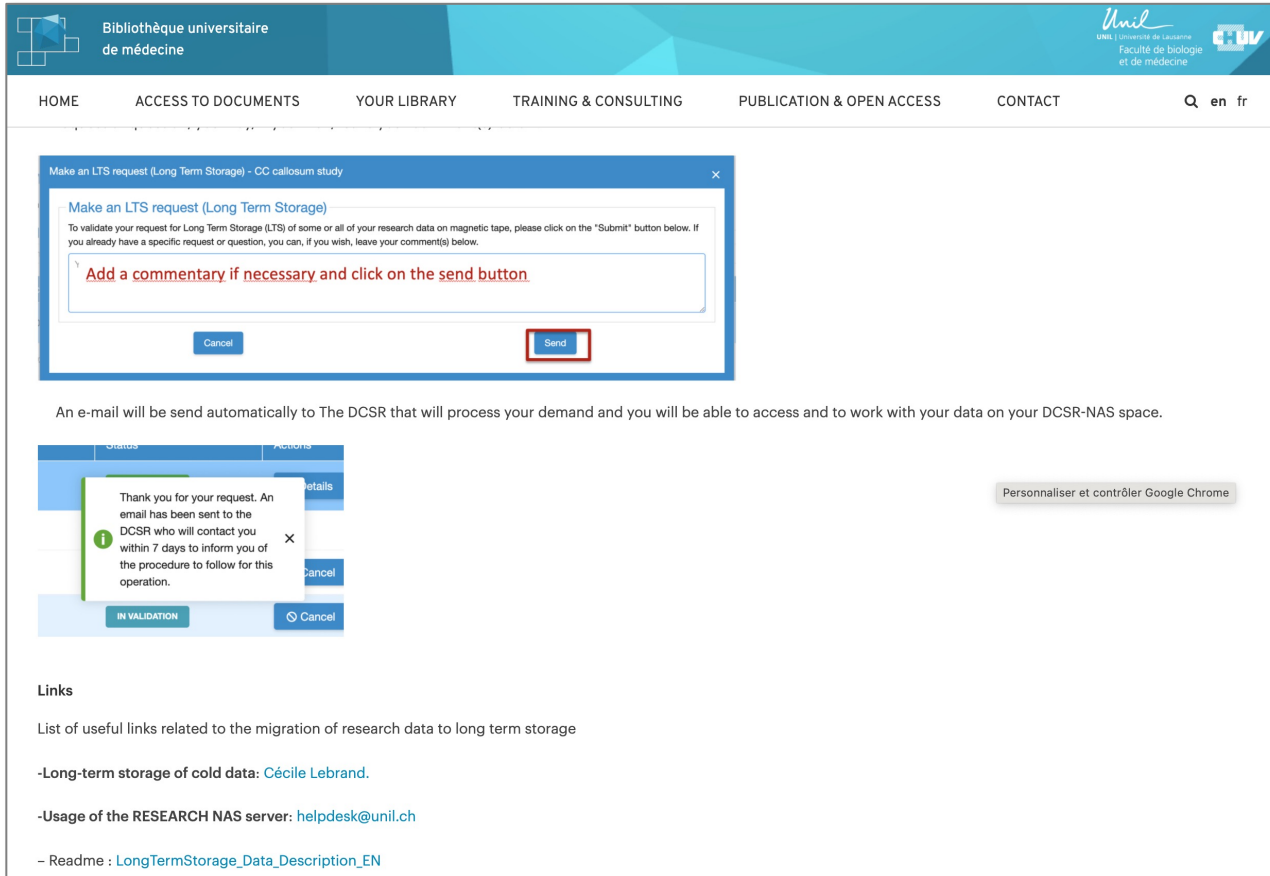
Published data

- Deposition and **sharing data via** repositories will ensure longevity.
- **Long Term Storage** of research data supporting published articles for at least 10 years after publication.

Unpublished data

- Save unpublished high-quality final data for **re-use**.
- **Bad quality data should be permanently discarded.**

Long Term Storage for « cold data » on a cost effective magnetic tape solution



The screenshot displays the website of the Bibliothèque universitaire de médecine at the University of Lausanne. The navigation menu includes HOME, ACCESS TO DOCUMENTS, YOUR LIBRARY, TRAINING & CONSULTING, PUBLICATION & OPEN ACCESS, and CONTACT. A search bar is located in the top right corner.

The main content area features a modal window titled "Make an LTS request (Long Term Storage) - CC callosum study". The modal contains the following text:

Make an LTS request (Long Term Storage)

To validate your request for Long Term Storage (LTS) of some or all of your research data on magnetic tape, please click on the "Submit" button below. If you already have a specific request or question, you can, if you wish, leave your comment(s) below.

Add a commentary if necessary and click on the send button.

At the bottom of the modal, there are "Cancel" and "Send" buttons. The "Send" button is highlighted with a red box.

Below the modal, a message states: "An e-mail will be send automatically to The DCSR that will process your demand and you will be able to access and to work with your data on your DCSR-NAS space."

A confirmation message is displayed in a green box:

Thank you for your request. An email has been sent to the DCSR who will contact you within 7 days to inform you of the procedure to follow for this operation.

Buttons for "Cancel" and "IN VALIDATION" are visible at the bottom of the confirmation message.

At the bottom of the page, there is a "Links" section with the following text:

List of useful links related to the migration of research data to long term storage

- Long-term storage of cold data: [Cécile Lebrand](#).
- Usage of the RESEARCH NAS server: helpdesk@unil.ch
- Readme : [LongTermStorage_Data_Description_EN](#)



4. CONCLUDING PHASE

b. FAIR Principles In Depth

The FAIR data principles



https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

- [FAIR Data Principles](#)
- [SNSF Explanation of the FAIR Data Principles \(PDF\)](#)
- [Wilkinson et al. \(2016\), The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3, doi:10.1038/sdata.2016.18 \(PDF\)](#)

The FAIR data principles

The FAIR Guiding Principles for scientific data management and stewardship.

Wilkinson et al.

Sci Data. 2016 Mar 15;3:160018. [doi: 10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).

to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

A set of guiding principles to enhance

- the value of all digital resources and
- its findability and use by machines and
- its reuse by individuals

Access & Citations

143k

Article Accesses

1401

Web of Science

1752

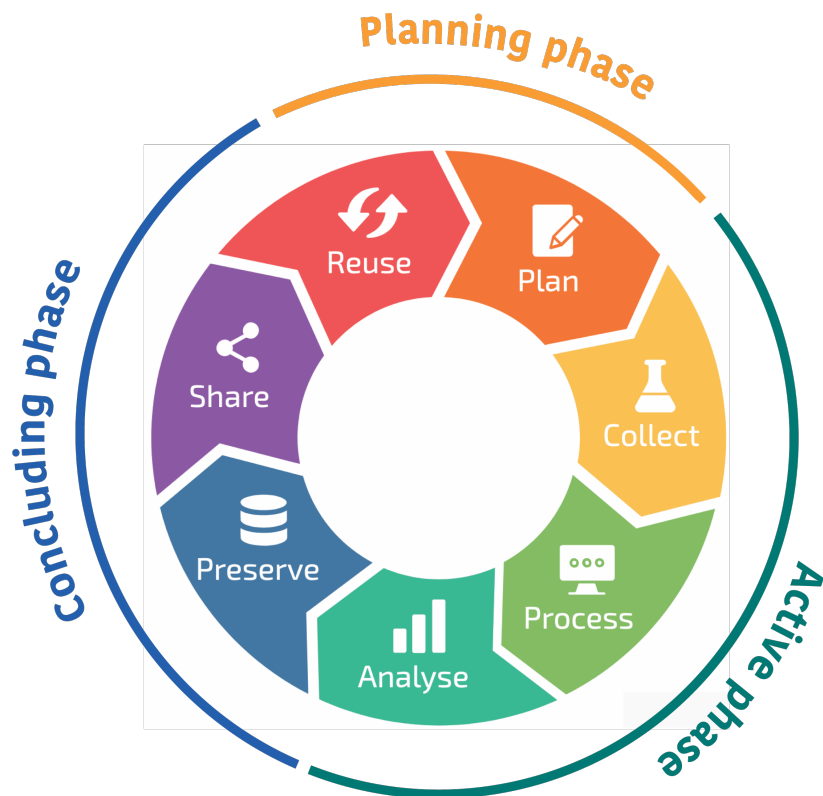
CrossRef



Attention Score

Support FAIR by design

Ideally, we should support FAIR data **by-design**, from the first steps of data management planning to the final steps of depositing data in public archives



The FAIR principles

inspired

by



Swim icons created by Freepik - Flaticon

The FAIR principles

F_{indable}



Can I find the data I want (unambiguously!) ?



Unambiguous identifiers supported by searchable metadata

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

=> someone else could not reuse/reassign the same identifier without referring to your data

=> identifiers (links) should not become invalid over time

Examples

<http://www.uniprot.org/uniprot/P98161>

(=> <https://www.uniprot.org/uniprotkb/P98161/entry>)

<https://doi.org/10.6084/m9.figshare.2056767.v2>

Globally unique and persistent identifiers are often provided by the repository or from a registry service.

The FAIR principles

F_{indable}



Can I find the data I want (unambiguously!) ?



Unambiguous identifiers supported by searchable metadata

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

=> “rich metadata” in F is for discovery, they describe what your dataset is about

Examples

Title, description, last version update, ...

(Rich metadata in R is for reusability)

The FAIR principles

F_{indable}



Can I find the data I want (unambiguously!) ?



Unambiguous identifiers supported by searchable metadata

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

=> the data identifier should be included in the metadata ...

Why?

Sometimes the metadata becomes separated from the data;

Thus depending on how someone finds your data, it may be impossible to find a link back to the data or to the metadata ...

The FAIR principles – F3

<https://www.metadata2go.com/>



Metadata Info Of Your File

The following table contains all the exif data and metadata

file_name
dneu22075sup0001suppfig1.tif

file_size
398 kB

file_type
TIFF

file_type_extension
tif

mime_type
image/tiff

exif_byte_order
Big-endian (Motorola, MM)

subfile_type
Full-resolution image

image_width
1854

image_height
1694

bits_per_sample
8 8 8

software
Adobe Photoshop CS5 Macintosh

modify_date
2013:02:10 22:29:01

predictor
Horizontal differencing

xmp_toolkit
Adobe XMP Core 5.0-c060 61.134777, 2010/02/12-17:32:00

format
image/tiff

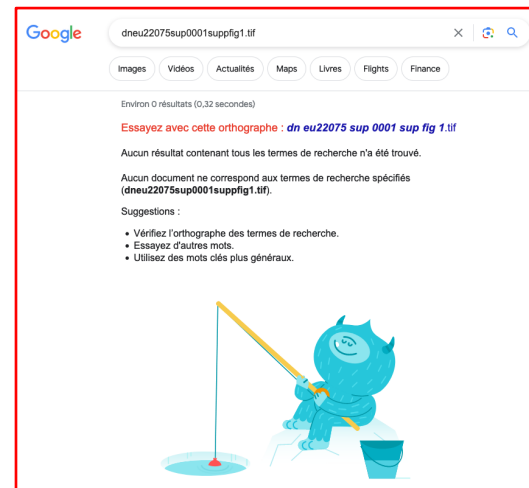
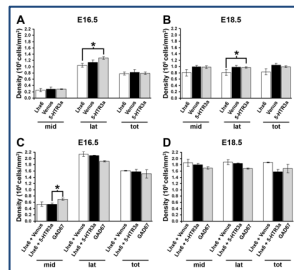
creator_tool
Adobe Photoshop CS4 Windows

create_date
2010:11:20 12:31:35+01:00

metadata_date
2013:02:10 22:29:01+01:00

document_id
uuid:F18178BB99F4DF119153B33BCB6278CB

dneu22075sup0001suppfig1.tif



The FAIR principles

F_{indable}



Can I find the data I want (unambiguously!) ?



Unambiguous identifiers supported by searchable metadata

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource**

- ⇒ automated indexation by Google, Bing, ...
- ⇒ specific metadata indexes: data catalog (DCAT), ...

Examples

<https://www.fairdatapoint.org/>

IMI Data Catalog - <https://datacatalog.elixir-luxembourg.org/>

The FAIR principles



Is there a way to access the data? Is it technically possible?

➡ *Clearly-defined access protocol, preferably machine-actionable*

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

⇒ focuses on how data and metadata can be retrieved (not “resolved”; A1 does not require web!)

⇒ defines what the protocol is to request access to the data (to be compliant)

The FAIR principles



Is there a way to access the data? CAN I re-use the data? Is it technically possible?

➡ *Clearly-defined access protocol, preferably machine-actionable*

A1. (Meta)data are **retrievable** by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

⇒ focuses on how data and metadata can be retrieved (not “resolved”; A1 does not require web!)

⇒ defines what the protocol is to request access to the data (to be compliant)

⇒ whatever the protocol, one should not have to buy a licence to use the tools required to access the data

Examples

https, sFTP ...

Telephone! (yes: the protocol may involve human intervention)

The FAIR principles



Is there a way to access the data? CAN I re-use the data? Is it technically possible?

➡ *Clearly-defined access protocol, preferably machine-actionable*

A1. (Meta)data are **retrievable** by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

⇒ focuses on how data and metadata can be retrieved (not “resolved”; A1 does not require web!)

⇒ one needs to define what the protocol is to request access to the data to be compliant

⇒ whatever the protocol, one should not have to buy a licence to use the tools required to access the data

Examples

https, sFTP ...

Telephone! (yes: the protocol may involve human intervention)

⇒ open protocol, not open data!

⇒ authorization via telephone, email, web, ...

FAIR ≠ Open Data

The FAIR principles



Is there a way to access the data? CAN I re-use the data? Is it technically possible?

➡ *Clearly-defined access protocol, preferably machine-actionable*

A1. (Meta)data are **retrievable** by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. **Metadata are accessible, even when the data are no longer available**

⇒ It is therefore recommended that data and metadata are separated!

1 database for the metadata and 1 database for the data;

do not embed all the metadata in the images;

Why?

⇒ metadata are more discoverable by metadata aggregators, such as OpenAIR <https://www.openaire.eu/>

⇒ metadata are valuable even if the data is lost or not available anymore

The FAIR principles

Interoperable



Can I understand the data? or better, can my machine understand the data?



Use shared vocabularies/ontologies in machine-accessible format

11. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
12. (Meta)data use vocabularies that follow FAIR principles
13. (Meta)data include qualified references to other (meta)data

⇒ most widely used for FAIRfication projects: Resource Description Framework (RDF)

Why?

Standard specifically designed to describe any type of information in a structured and machine-readable format.

Alternatives:

- JSON-LD - JavaScript Object Notation for Linking Data - <https://json-ld.org/>
- Microdata - [https://en.wikipedia.org/wiki/Microdata_\(HTML\)](https://en.wikipedia.org/wiki/Microdata_(HTML))
- ...

The FAIR principles

Interoperable



Can I understand the data? or better, can my machine understand the data?



Use shared vocabularies/ontologies in machine-accessible format

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

⇒ the vocabularies describing the (meta)data must themselves be Findable, Accessible, Interoperable and Reusable

The FAIR principles

Interoperable



Can I understand the data? or better, can my machine understand the data?




Use shared vocabularies/ontologies in machine-accessible format

11. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
12. (Meta)data use vocabularies that follow FAIR principles
13. (Meta)data include qualified references to other (meta)data

qualified references

⇒ cross-references that explains its intent: “... is a regulator of ...”

P04278 · SHBG_HUMAN

| | |
|-----------------------|-------------------------------------------------------------------------------------------------------------------|
| Protein ⁱ | Sex hormone-binding globulin |
| Gene ⁱ | SHBG |
| Status ⁱ |  UniProtKB reviewed (Swiss-Prot) |
| Organism ⁱ | Homo sapiens (Human) |

```
SELECT ?protein ?organism
WHERE
{
  ?protein a up:Protein .
  ?protein up:organism taxon:9606 .
}
```

Are these good qualified references?

For humans, yes. But not for machines ...

Proteins

derived from

Homo sapiens

The FAIR principles



Am I allowed to use the data? SHOULD I re-use the data? Is it useful for me?



Contextual information, allowing proper interpretation

Rich provenance information facilitating accurate citation

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

⇒ focuses on the ability of a user (machine or human) to decide if the data is actually USEFUL

⇒ metadata richly describes the context under which the data was generated

plurality : do as much as you can; be generous;

Examples

experimental protocol

the manufacturer and brand of the machine or sensor that created the data

the species used

the drug regime

the particularities, the limitations

...

(Rich metadata in F is for discovery)

The FAIR principles



Am I allowed to use the data? SHOULD I re-use the data? Is it useful for me?



Contextual information, allowing proper interpretation

Rich provenance information facilitating accurate citation

R1. (Meta)data are richly described with a **plurality** of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

⇒ data usage license: bad wording

(license or data access policy would have been better)

The conditions under which the data can be used should be clear to machines and humans.

The FAIR principles



Am I allowed to use the data? SHOULD I re-use the data? Is it useful for me?



Contextual information, allowing proper interpretation

Rich provenance information facilitating accurate citation

R1. (Meta)data are richly described with a **plurality** of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

⇒ they should allow the user to decide whether or not they **trust** the data

what did you do?

what organism was used?

what protocols were followed?

what was the question?

which workflows, software and versions of, filters?

who and how to cite?

etc.

The FAIR principles



Am I allowed to use the data? SHOULD I re-use the data? Is it useful for me?



Contextual information, allowing proper interpretation

Rich provenance information facilitating accurate citation

R1. (Meta)data are richly described with a **plurality** of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

⇒ "catch-all"

⇒ in addition to all the previous principles, you need to adhere to community standards

you should use your community standard ...

... even if it's not FAIR! And then you should also use a FAIR standard

The FAIR principles



Can I find the data I want (unambiguously!) ?

➡ Unambiguous identifiers supported by searchable metadata



Is there a way to access the data? CAN I re-use the data? Is it technically possible?

➡ Clearly-defined access protocol, preferably machine-actionable



Can I understand the data? or better, can my machine understand the data?

➡ Use shared vocabularies/ontologies in machine-accessible format



Am I allowed to use the data? SHOULD I re-use the data? Is it useful to my problem?

➡ Contextual information, allowing proper interpretation

➡ Rich provenance information facilitating accurate citation

The FAIR principles

To be Findable

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. metadata include the data identifier it describes.
- F4. (meta)data are registered or indexed in a searchable resource.

To be Accessible

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

To be Interoperable

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

To be Re-usable

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

How FAIR is your data?

Screening of ~5500 FDA-approved drugs and clinical candidates for anti-SARS-CoV-2 activity

Van Damme, Ellen

This report describes the most relevant results of screening a drug repurposing library consisting of ~5500 FDA-approved drugs and clinical candidates that have passed phase I studies for activity against SARS-CoV-2.

<https://zenodo.org/record/4774709>

Screening of ~5500 FDA-approved drugs and clinical candidates for anti-SARS-CoV-2 activity.xlsx

(Demo)

Pre-fairification

The disadvantage of depositing in Zenodo is that Zenodo is not the primary archive for bioactivity data discovery. There are only 75 views and 23 downloads of the CARE dataset from May 2021 to Sept. 2021.

Select the appropriate repository for your data

Also with limited metadata packed in a spreadsheet, it is difficult to find the dataset.

Metadata and ontologies are crucial

<https://fairplus-project.eu/>



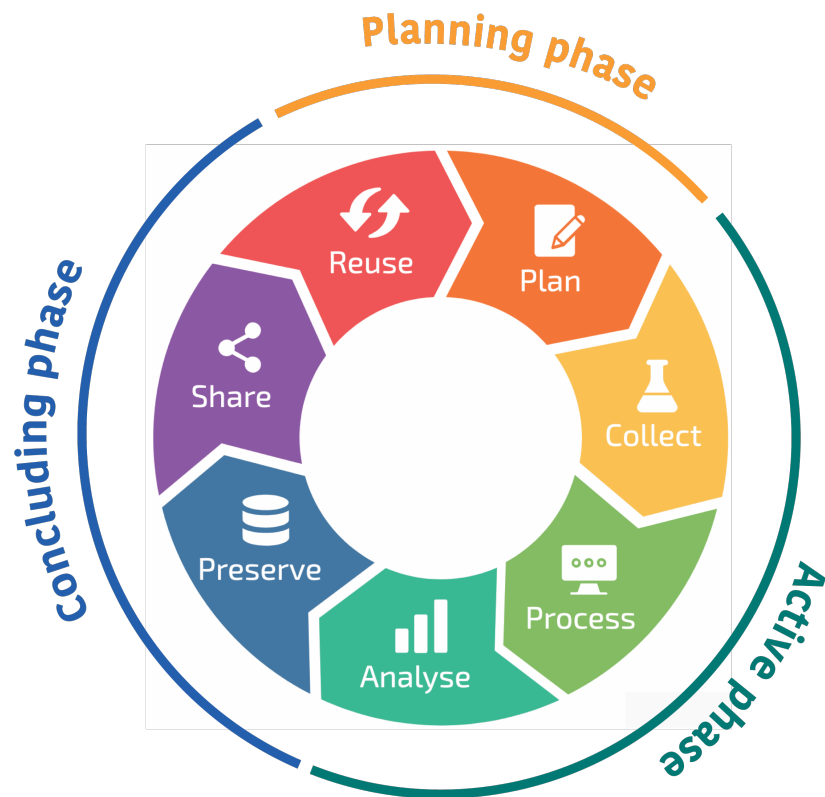
FAIRplus

<https://www.imi.europa.eu/projects-results/project-factsheets/fairplus>

FAIRplus FAIRification Process

1. Define FAIRification Goal – “competency question”
 1. Findability
 2. Reusability
2. Data Examination
 1. Identify Data requirements - **what you have**
 2. Identify Data FAIRification Capabilities – **what you can do**
 3. Identify Data FAIRification Resources – **who does what**
3. Design Decisions - **including cost-benefit evaluation**
4. Implementation Phase
 1. Generate InChI and SMILE identifiers
 2. Map data to the ChEMBL submissions model
 3. BioAssay Ontology (BAO), CHEMINF vocabularies for compound annotation
 4. Cellosaurus ID for cell line
 5. ChEMBL license used for data sharing, CC-SA 3.0

FAIR hands-on course autumn 2023!





4. CONCLUDING PHASE

c. Data Sharing & Reuse

mySNF Data Management Plan (DMP)

1. Data collection and documentation

- 1.1 What data will you collect, observe, generate or reuse?
- 1.2 How will the data be collected, observed or generated?
- 1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

- 2.1 How will ethical issues be addressed and handled?
- 2.2 How will data access and security be managed?
- 2.3 How will you handle copyright and Intellectual Property Rights issues?

3. Data storage and preservation

- 3.1 How will your data be stored and backed-up during the research?
- 3.2 What is your data preservation plan?

4. Data sharing and reuse

- 4.1 How and where will the data be shared?
- 4.2 Are there any necessary limitations to protect sensitive data?
- 4.3 All digital repositories I will choose are conform to the FAIR Data Principles.
- 4.4 I will choose digital repositories maintained by a non-profit organisation.

4.1. How and where will the data be shared?

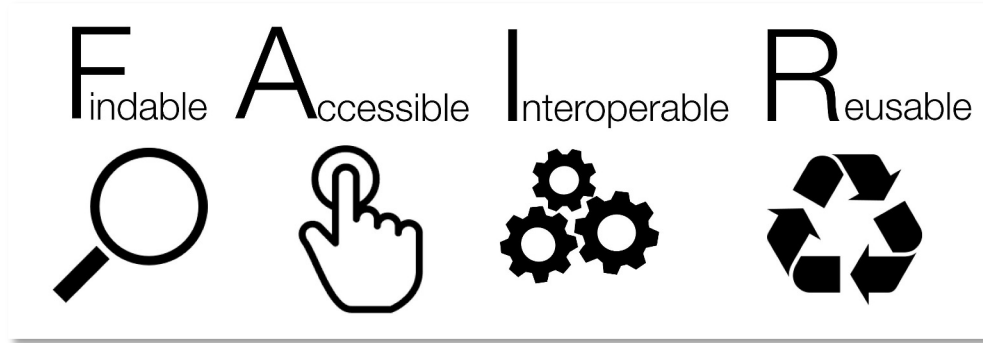
Questions you might want to consider

- On which repository do you plan to share your data?
- How will potential users find out about your data?

SNSF says...

Consider **how and on which repository the data will be made available**. The methods applied to data sharing will depend on several factors such as the type, size, complexity and sensitivity of data. Please also consider **how the reuse of your data will be valued and acknowledged by other researchers**. (This relates to the *FAIR Data Principles* F1, F3, F4, A1, A1.1, A1.2 & A2)

FAIR data repositories: SNSF info



Finding the "perfect" repository providing all necessary features to host FAIR data is challenging. To make the transition towards FAIR research data easier, the **SNSF decided to fix a set of minimal criteria that repositories have to fulfil to conform with the FAIR data principles.**

- [Guidelines](#)
- [Examples of data repositories \(PDF\)](#)
- [Dunning et al. \(2017\), Evaluation of data repositories based on the FAIR Principles for IDCC 2017 practice paper, TU Delft, Data Set, doi:10.4121/uuid:5146dd06-98e4-426c-9ae5-dc8fa65c549f](#)

Registry of FAIR Data Repositories

The screenshot displays the FAIRsharing.org website interface. At the top, there is a search bar with the text 'search through all content' and a 'SEARCH' button. Below the search bar are navigation buttons for 'STANDARDS', 'DATABASES', 'POLICIES', 'COLLECTIONS', 'ORGANISATIONS', 'ADD CONTENT', and 'STATS'. The 'DATABASES' button is highlighted in red. The main header area features a dark blue background with a network diagram and the text 'Databases' and 'A registry of knowledgebases and repositories of data and other digital assets.' Below this, there is a search bar for 'current results' and a 'SEARCH' button. The search results are filtered by 'Registry: Database' and 'Query string: Zenodo'. The results show one entry for 'Zenodo', which is a generalist research data repository built and developed by OpenAIRE and CERN. The entry includes a description, a 'Zenodo' logo, and a list of related categories: 'Data Mana...', 'Subject Ag...', 'Data Stora...', and 'All'. Below the entry, there is a list of related databases and policies with their respective counts: 'Standards Implemented' (14), 'Related Databases' (16), and 'Endorsing Policies' (19).

FAIRsharing.org standards, databases, policies search through all content Q SEARCH LOGIN

STANDARDS DATABASES POLICIES COLLECTIONS ORGANISATIONS ADD CONTENT STATS

Databases

A registry of knowledgebases and repositories of data and other digital assets.

Search through current results. Q SEARCH

MATCH ALL TERMS MATCH ANY TERM

MAINTAINED NOT MAINTAINED

RECOMMENDED NOT RECOMMENDED

READY DEPRECATED UNCERTAIN IN DEV.

Registry

Search APPLY

Record Type

Search APPLY

Clear All Registry: Database Query string: Zenodo

1

Displaying 1 to 4 of 4.

Zenodo

Zenodo is a generalist research data repository built and developed by OpenAIRE and CERN. It was developed to aid Open Science and is built on open source code. Zenodo helps researchers receive credit by making the research results citable and through OpenAIRE integrates them into existing reporting lines to funding agenci...

Data Mana... Subject Ag... Data Stora... All

- Standards Implemented 14
- Related Databases 16
- Endorsing Policies 19

Data submission wizard to find the right repository at EMBL

Data submission

Use this data submission wizard to find the right archive for your data in a few simple steps.

1 What **type of data** do you have?

DNA/RNA sequence

Expression data

Protein data

Structures

Systems

Chemical biology

Ontologies

Images

Multi-omics or other cross-domain study

Other biological research data

Why submit data to an archive?

- Submission of primary data and derived information to public data repositories is an essential step in the scientific process.
- Through submission, the scientific community is fed the raw materials for the building and maintenance of the complete and up-to-date data sets that support searches and analysis on the latest sequences, structures and molecular profiles of living systems.
- Serving as a complement to the literature publication process and supporting early data sharing, the EMBL-EBI offers a number of submission services appropriate for different types and scales of data.

Need help?

🗨 If you need help with your data submission, please [contact support](#).

All EMBL-EBI data repositories

[Array Express](#): functional genomics data

[BioImage Archive](#): bioimaging data

[BioModels](#): computational models

[BioSamples](#): reference sample data

[BioStudies](#): biological research data

[ChEBI](#): chemical entities

[DGVa](#): structural genetic variation data

[EFO](#): experimental variables

[EGA](#): human data that requires controlled access

[EMPIAR](#): raw image data

[ENA](#): nucleotide sequence data

[EVA](#): genetic variation data

[GO](#): Gene ontology annotations

[GWAS Catalog](#): Genome-wide association study data

[IntAct](#): molecular interactions

[IntEnz](#): enzyme nomenclature

[MetaboLights](#): metabolomics data

[Metagenomics](#): raw sequence data & associated meta-data

[OneDep](#): electron microscopy, X-ray crystallography & NMR data

[PRIDE](#): protein & peptide identification data

[UniProt SPIN](#): protein sequences & annotations

[UniProt](#): updates or corrections

Registry of Data Repositories



<http://www.re3data.org/>

Find the right repository for your data



<http://www.nature.com/sdata/policies/repositories>

Browse our recommended data repository online.

- *We currently list more than 90 repositories, across the biological, physical and social sciences*
- *We advise authors on the best place to store their data*

Domain-specific public repositories

Sequencing

[dbSNP](#)

[dbVar](#)

[Database of Genomic Variants Archive \(DGVa\)](#)

[DNA DataBank of Japan \(DDBJ\)](#)

[EBI Metagenomics](#)

[EMBL Nucleotide Sequence Database \(ENA\)](#)

[European Variation Archive \(EVA\)](#)

[GenBank](#)

[miRBase](#)

[NCBI Sequence Read Archive \(SRA\)](#)

[NCBI Trace Archive](#)

[Uniprot](#)

Omics

[ArrayExpress](#)

[Biological General Repository for Interaction Datasets \(BioGRID\)](#)

[Database of Interacting Proteins \(DIP\)](#)

[dbGAP](#)

[The European Genome-phenome Archive \(EGA\)](#)

[IntAct Molecular Interaction Database](#)

[Gene Expression Omnibus \(GEO\)](#)

[GenomeRNAi](#)

[GPM DB](#)

[MetaboLights](#)

[NURSA](#)

[PeptideAtlas](#)

[Proteomics Identifications \(PRIDE\)](#)

[ProteomeXchange](#)

Structural Databases

[Biological Magnetic Resonance Data Bank \(BMRB\)](#)

[Crystallography Open Database \(COD\)](#)

[Coherent X-ray Imaging Data Bank \(CXIDB\)](#)

[Electron Microscopy Data Bank \(EMDB\)](#)

[FlowRepository](#)

[Protein Circular Dichroism Data Bank \(PCDDb\)](#)

[Worldwide Protein Data Bank \(wwPDB\)](#)

Neuroscience

[Functional Connectomes Project International Neuroimaging Data-Sharing Initiative](#)

[\(FCP/INDI\)](#)

[NeuroMorpho.org](#)

[OpenfMRI](#)

Model Organisms

[Eukaryotic Pathogen Database Resources \(EuPathDB\)](#)

[FlyBase](#)

[Mouse Genome Informatics \(MGI\)](#)

[Rat Genome Database \(RGD\)](#)

[SmedGD](#)

[The Arabidopsis Information Resource \(TAIR\)](#)

[VectorBase](#)

[WormBase](#)

[Xenbase](#)

[Zebrafish Model Organism Database \(ZFIN\)](#)

Taxonomic & Species Diversity

[Integrated Taxonomic Information System \(ITIS\)](#)

[Global Biodiversity Information Facility \(GBIF\)](#)

[NCBI Taxonomy](#)

[The Knowledge Network for Biocomplexity](#)

Biomedical Sciences

[Influenza Research Database](#)

[National Addiction & HIV Data Archive Program \(NAHDAP\)](#)

[National Database for Autism Research \(NDAR\)](#)

[The Cancer Imaging Archive \(TCIA\)](#)

[SICAS Medical Image Repository](#)

[PhysioNet](#)

Biochemistry

[caNanoLab](#)

[Kinetic Models of Biological Systems \(KiMoSys\)](#)

[The Mass spectrometry Interactive Virtual Environment \(MassIVE\)](#)

[PubChem](#)

Social sciences

[Inter-university Consortium for Political and Social Research \(ICPSR\)](#)

[Qualitative Data Repository](#)

[Swedish National Data Service](#)

[Data Archiving and Networking Services \(DANS\)](#)

<http://journals.plos.org/plosone/s/data-availability#loc-recommended-repositories>


https://figshare.com/articles/Scientific_Data_recommended_repositories_June_2015/1434640

Dataset versions, Data repositories & Open Data

Confidentiality of clinical data
& intellectual property



Communauté FBM sur Zenodo (dépôt H2020) pour dépôt de sets de données non structurées

Upload Communitiescecile.lebrand@chuv.ch

April 11, 2017 Dataset Open Access

Dataset Multifaceted intervention for patients admitted to an emergency unit for suicide attempt: an exploratory study

Brovelli Sebastien; Dorogi Yves; Feiner Adam-Scott; Golay Philippe; Stiefel Friedrich; Bonsack Charles; Michaud Laurent

This dataset is related to "Multifaceted intervention for patients admitted to an emergency unit for suicide attempt: an exploratory study" (Brovelli S., Dorogi Y., Feiner A.-S., Golay P., Stiefel F., Bonsack C. & Michaud L.)

[Preview](#)


| Age | Gender | Included | Reason of refusal | Sent by | Origin | Living | Marital Status | Children | Education | Occupation |
|-------|--------|----------|-------------------|---------|--------|--------|----------------|----------|-----------|------------|
| 52.00 | F | 0.00 | 0.00 | 2.00 | 1.00 | 2.00 | 4.00 | 1.00 | 0.00 | 5.00 |
| 31.00 | H | 0.00 | 3.00 | 3.00 | 1.00 | 3.00 | 1.00 | 2.00 | 0.00 | 5.00 |
| 33.00 | H | 0.00 | 0.00 | 3.00 | 2.00 | 3.00 | 1.00 | 2.00 | 0.00 | 5.00 |
| 34.00 | H | 0.00 | 3.00 | 4.00 | 2.00 | 5.00 | 2.00 | 1.00 | 0.00 | 5.00 |
| 51.00 | H | 0.00 | 0.00 | 3.00 | 2.00 | 1.00 | 5.00 | 1.00 | 0.00 | 2.00 |
| 38.00 | H | 0.00 | 0.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.00 | 6.00 | 2.00 |
| 58.00 | F | 0.00 | 0.00 | 3.00 | 2.00 | 2.00 | 1.00 | 1.00 | 0.00 | 7.00 |
| 52.00 | F | 0.00 | 0.00 | 3.00 | 1.00 | 2.00 | 3.00 | 2.00 | 0.00 | 5.00 |

Communities

Faculty of Biology and Medicine at University of Lausanne & Lausanne University Hospital Remove

69 views 39 downloads
[See more details...](#)


Indexed in



October 25, 2022

Software Open Access

connectomicslab/connectomemapper3: Connectome Mapper v3.1.0

 Tourbier, Sebastien;  Rue Queralt, Joan;  Glomb, Katharina;  Aleman-Gomez, Yasser;  Mullier, Emeline;  Griffa, Alessandra;  Schöttner, Mikkel;  Wirsich, Jonathan;  Tuncel, Anil;  Jancovic, Jakub;  Bach Cuadra, Meritxell;  Hagmann, Patric

Connectome Mapper 3 implements, in accordance to the [BIDS-App standard](#), full anatomical, diffusion, and resting/state functional MRI processing pipelines, from raw T1, DWI, BOLD, and preprocessed EEG data to multi-resolution brain parcellation with corresponding connection matrices, based on a new version of the Lausanne parcellation atlas ([Cammoun et al, 2012](#)), aka Lausanne2018.

This first minor version integrates the new pipeline dedicated to EEG modality and fixes a few bugs. 😊

What's Changed

Updates

- The conda environment files (`conda/environment.yml` and `conda/environment_macosx.yml`) for `cmpbidsappmanager` have been greatly modified (PR #212).
- In all conda environment `*.yml` and `requirements.txt` files, datalad and its container extension have been updated to version `0.17.2` and `1.1.6` respectively. (PR #209)

New features

- The new pipeline dedicated to EEG modality has been integrated into the BIDS App and `cmpbidsappmanager`. (PR #201 and PR #205)
- Option to apply or not band-pass filtering in fMRI pipeline. (PR #200)

Code refactoring

- Major refactoring of all the code related to the EEG pipeline. (PR #198)

Bug fix

- Problems to install and launch `cmpbidsappmanager` on Ubuntu. (PR #212)
- Fix `nibabel` to `3.2.2` as the imported functions of `nibabel.trackvis` has been moved since `4.0.0` and caused errors. (PR #211)
- Fix problem of traits not updated while making the diffusion pipeline config with ACT. (PR #200)

Documentation

- Update/add documentation for the EEG pipeline. (PR #208)

Software development life cycle

- Optimization of resources stored in the cache and in the workspace. (PR #201)

Communities


Faculty of Biology and
Medicine at University of
Lausanne & Lausanne
University Hospital

[✕ Remove](#)

1,291

 views

63

 downloads[See more details...](#)

Available in



Indexed in



Publication date:

October 25, 2022

DOI:

[DOI 10.5281/zenodo.7249263](https://doi.org/10.5281/zenodo.7249263)

Keyword(s):

[neuroimaging](#) [workflow](#) [pipeline](#) [parcellation](#)
[connectome](#) [MRI](#) [BIDS](#) [BIDS-App](#)

Digital Object Identifier

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}

Data Principles

Digital Object Identifier

Persistent

<http://doi.org/10.5281/zenodo.400920>



<https://zenodo.org/record/400920>

Digital Object Identifier

RESEARCH ARTICLE

Development of a duplex real-time PCR for the detection of *Rickettsia* spp. and typhus group rickettsia in clinical samples

Stefano Giulieri¹, Katia Jaton², Alain Cometta³, Laurence T. Trellu⁴, Gilbert Greub^{1,2}

¹Infectious Diseases Service, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland; ²Institute of Microbiology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland; ³Service of Internal Medicine, Yverdon Hospital, Yverdon, Switzerland; and ⁴Service of Dermatology, University Hospital, Geneva, Switzerland

Correspondence: Gilbert Greub, Institute of Microbiology, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland. Tel.: +41 21 314 49 79; fax: +41 21 314 40 60; e-mail: Gilbert.Greub@chuv.ch

Received 15 August 2011; revised 31 October 2011; accepted 11 November 2011. Final version published online 12 December 2011.

DOI: 10.1111/j.1574-695X.2011.00910.x

Keywords: rickettsia; polymerase chain reaction; spotted fever; typhus.

Abstract

Molecular diagnosis using real-time polymerase chain reaction (PCR) may allow earlier diagnosis of rickettsiosis. We developed a duplex real-time PCR that amplifies (1) DNA of any rickettsial species and (2) DNA of both typhus group rickettsia, that is, *Rickettsia prowazekii* and *Rickettsia typhi*. Primers and probes were selected to amplify a segment of the 16S rRNA gene of *Rickettsia* spp. for the pan-rickettsial PCR and the citrate synthase gene (*gltA*) for the typhus group rickettsia PCR. Analytical sensitivity was 10 copies of control plasmid DNA per reaction. No cross-amplification was observed when testing human DNA and 22 pathogens or skin commensals. Real-time PCR was applied to 17 clinical samples. Rickettsial DNA was detected in the skin biopsies of three patients. In one patient with severe murine typhus, the typhus group PCR was positive in a skin biopsy from a petechial lesion and seroconversion was later documented. The two other patients with negative typhus group PCR suffered from Mediterranean and African spotted fever, respectively; in both cases, skin biopsy was performed on the eschar. Our duplex real-time PCR showed a good analytical sensitivity and specificity, allowing early diagnosis of rickettsiosis among three patients, and recognition of typhus in one of them.

References

1. Walker DH (2007) Rickettsiae and Rickettsial Infections: The Current State of Knowledge. *Clin Infect Dis* 45: S39–S44.
2. Bavaro MF, Kelly DJ, Dasch DA, Hale BR, Olson P (2005) History of U. S. Military Contributions to the study of Rickettsial Diseases. *Military Medicine* 170:49–60. PMID: 15916283
3. Kelly DJ, Richards AL, Temenak J, Strickman D, Dasch GA (2002) The past and present threat of rickettsial diseases to military medicine and international public health. *Clinical Infect. Dis.* 34: S145–S169.
4. Rolain JM, Jensenius M, Raoult D (2004) Rickettsial infections—a threat to travellers? *Curr Opin Infect Dis* 17:433–7. PMID: 15353963
5. Kovacova E, Kazar J (2000) Rickettsial diseases and their serological diagnosis. *Clin Lab* 46:239–45. PMID: 10853230
6. La Scola B, Raoult D (1997) Laboratory diagnosis of rickettsioses: current approaches to diagnosis of old and new rickettsial diseases. *J Clin Microbiol* 35:2715–27. PMID: 9350721
7. Renvoisé A, Rolain JM, Socolovschi C, Raoult D (2012) Widespread use of real-time PCR for rickettsial diagnosis. *FEMS Immunol Med Microbiol* 64:126–9. doi: 10.1111/j.1574-695X.2011.00899.x PMID: 22092999
8. Giulieri S, Jaton K, Cometta A, Trellu LT, Greub G (2012) Development of a duplex real-time PCR for the detection of *Blastenia* spp. and typhus group rickettsia in clinical samples. *FEMS Immunol Med Microbiol* 64:42–7. doi: 10.1111/j.1574-695X.2011.00910.x PMID: 2098502
9. Prakash JA, Reller ML, Baran N, Savelle JG (2000) Assessment of a quantitative multiplex 5' nuclease real-time PCR for spotted fever and typhus group rickettsioses and *Orientia tsutsugamushi*. *Clin Microbiol Infect Suppl.* 2:292–3.
10. Jiang J, Chan TC, Temenak JJ, Dasch GA, Ching WM, et al. (2004) Development of a quantitative real-time polymerase chain reaction assay specific for *Orientia tsutsugamushi*. *Am J Trop Med Hyg* 70: 351–6. PMID: 15100446
11. Paris DH, Aukkanit N, Jenjaroen K, Blacksell SD, Day NP (2009) A highly sensitive quantitative real-time PCR assay based on the *groEL* gene of contemporary 339 Thai strains of *Orientia tsutsugamushi*. *Clin Microbiol Infect* 15:488–495. doi: 10.1111/j.1469-0691.2008.02671.x PMID: 19416296
12. Roux V, Rydkina E, Ereemeeva M, Raoult D (1997) Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the *rickettsiae*. *Int J Syst Bacteriol* 47: 252–261. PMID: 9103608

Digital Object Identifier

Metadata

```
<?xml version='1.0' encoding='utf-8'?>
<resource xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/sc
hema/kernel-3" xsi:schemaLocation="http://datacite.org/schema/kernel-3 http://schema.datacite
.org/meta/kernel-3/metadata.xsd">
  <identifier identifierType="DOI">10.5281/zenodo.400920</identifier>
  <creators>
    <creator>
      <creatorName>Carlos Ciller</creatorName>
    </creator>
    <creator>
      <creatorName>Sandro De Zanet</creatorName>
    </creator>
    <creator>
      <creatorName>Konstantinos Kamnitsas</creatorName>
    </creator>
    <creator>
      <creatorName>Philipp Meader</creatorName>
    </creator>
  </creators>
</resource>
```


How will you handle copyright and Intellectual Property Rights issues?











Open licenses for data sharing

Promote sharing and unlimited use of the data that you produced using explicit licences. For sharing your data, use a creative common CC0 license or a CC By license that is suitable for data sharing (check if there is a directive at your institution).

Create Commons (CC) licenses

Make your document freely accessible while protecting your copyright



| | |
|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  Attribution Others can copy, distribute, display, perform and remix your work if they credit your name as requested by you |
|  |  No Derivative Works Others can only copy, distribute, display or perform verbatim copies of your work |
|  |  Share Alike Others can distribute your work only under a license identical to the one you have chosen for your work |
|  |  Non-Commercial Others can copy, distribute, display, perform or remix your work but for non-commercial purposes only. |
|  | |
|  | |

Six different CC licences condition the legal copyright terms and allow the sharing and reuse of material.

Citation for a dataset

Data should be considered legitimate, citable products of research and be given the same importance in the scholarly record as citations of other research objects, such as publications

Minimum recommended format

Creator (PublicationYear): Title. Publisher. Identifier

Where *Publisher* is the data archive that holds the data and *Identifier* is displayed as linkable, permanent URLs.

Example

« Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo.

<http://dx.doi.org/10.1594/PANGAEA.726855> »

Data deposit on Zenodo: Demo

Zenodo

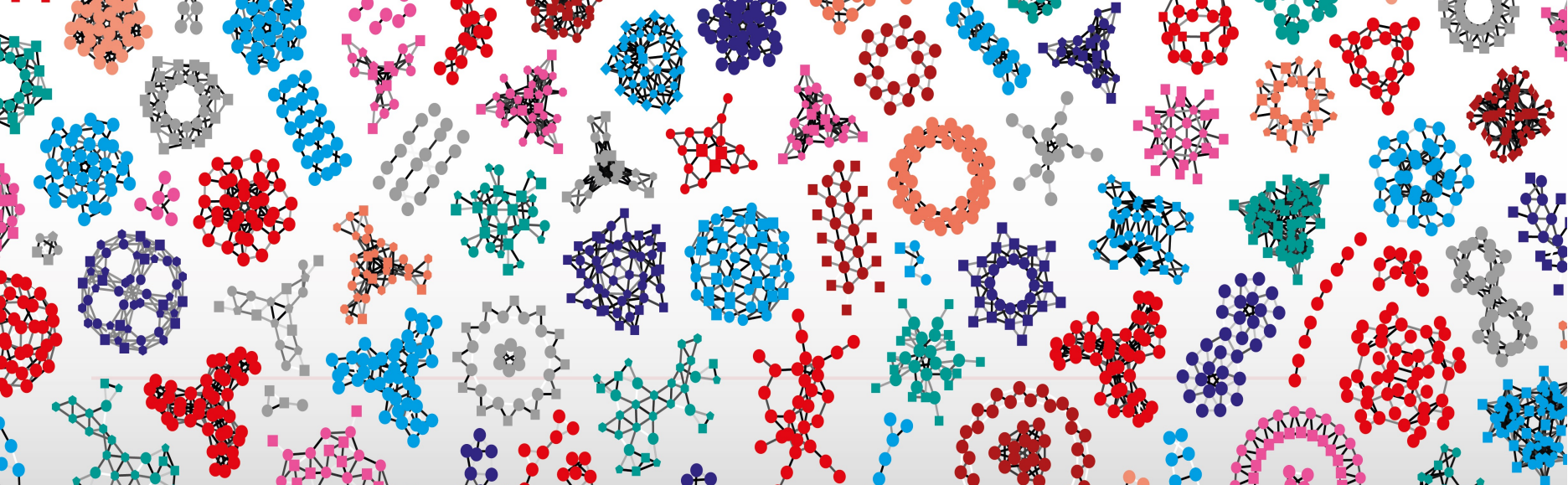
Upload

Describe

Publish

The screenshot shows the Zenodo website interface. At the top, there is a blue navigation bar with the Zenodo logo on the left, a search bar in the center, and 'Upload' and 'Communities' links on the right. Further right are 'Log in' and 'Sign up' buttons. Below the navigation bar, the main content area displays the name of the community: 'Faculty of Biology and Medicine at University of Lausanne & Lausanne University Hospital'. Underneath, there is a 'Recent uploads' section with a search bar and a 'View' button. A specific upload is listed with the date 'April 11, 2017', the type 'Dataset', and the status 'Restricted Access'. The title of the dataset is 'Dataset Multifaceted intervention for patients admitted to an emergency unit for suicide attempt: an exploratory study'. To the right of the upload list, there is a community profile card for 'Unil' (UNIL | Université de Lausanne) and 'CHUV' (Faculty of Biology and Medicine).

<https://zenodo.org/>



Thank you!

Data FAIRification course
coming soon!