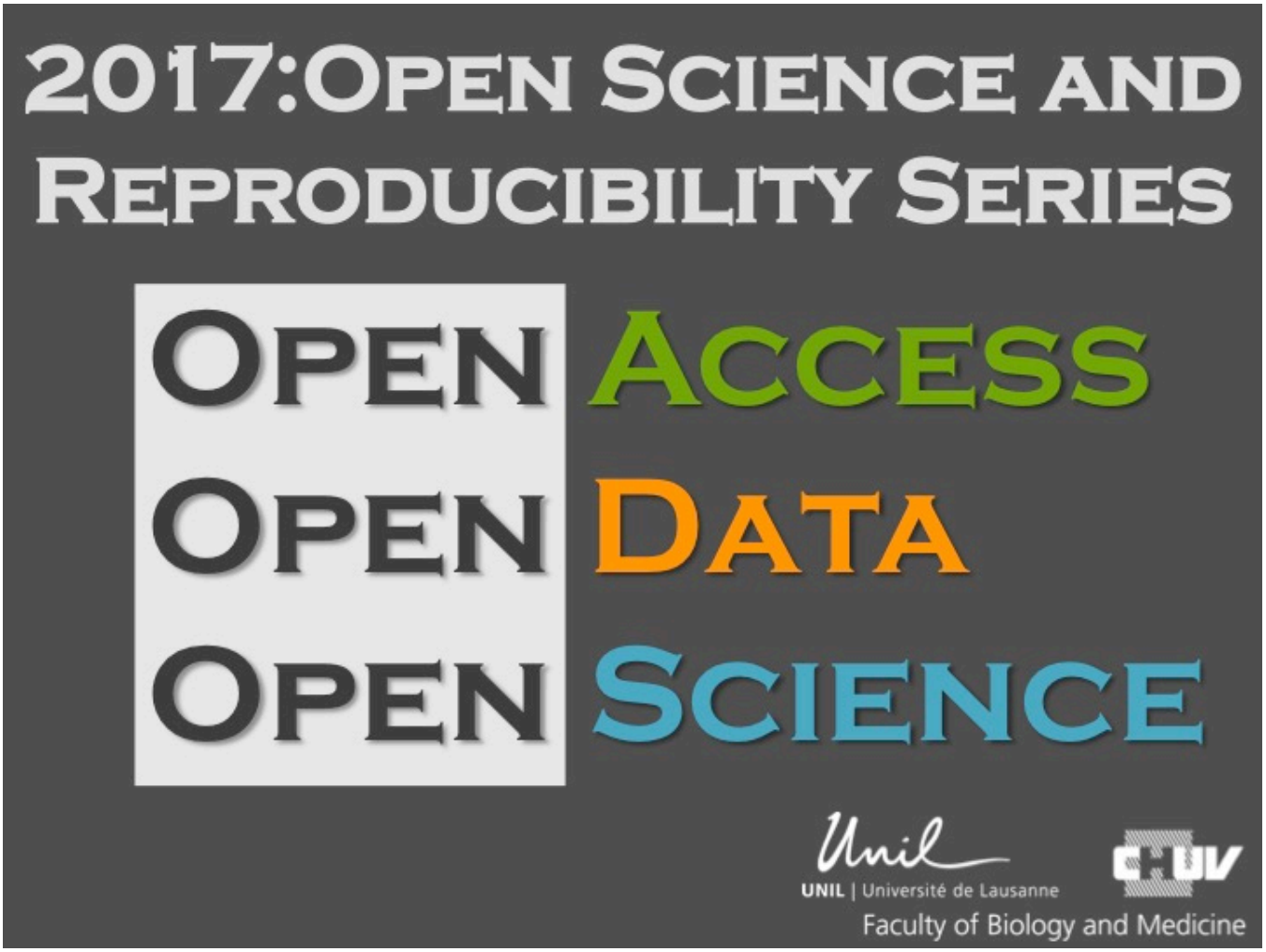

OPEN SCIENCE AND REPRODUCIBILITY SERIES

WORKSHOP III

"DATA MANAGEMENT & OPEN DATA"

LAUSANNE, MAY 22, 2017

Open Science and Reproducibility Series
Programme for Workshop III "Data Management & Open Data"
Monday May 22, 2017, Department of Fundamental Neurosciences, UNIL

Accredited by the RESAL as a one day training in animal research.

Workshop III on "Data management & Open Data" is part of the Lemanic **"Open Science & reproducibility" workshop series** that aims at sensitizing researchers from UNIL, CHUV, EPFL, UNIGE and HUG to the notion of Open Science in order to improve transparency and reproducibility of their research. This series of events, supported by the FBM Publication & Data Management Unit at CHUV Library, Lemanic Neuroscience Doctoral School, and Resal will be a unique opportunity for scientists to discuss and discover Open Science best practices and standards at all stages of the research process.

During the first part of this workshop, researchers will discover what are the SNFS funding agency new requirements **concerning data management** plan **and data sharing**. You will also discover the visions of **Prof. F. Bussy (Vice Rector for Research and International Relations, UNIL**) and **Prof. P. Vandergheynst (Vice President for education, EPFL)** on Open Science. You will be introduced to what is exactly a **Data Management Plan (DMP) and how it can make a difference throughout your career.** To help you, professionals involved in Big Data management at VitalIT/SIB as well as in Data Management Plan preparation at EPFL and DLCM will share with you **best practices to optimize** research data management (how to collect, describe, store, secure, share and archive research data).

The second part of the workshop will be dedicated to **Open Data**. This session will provide researchers with guidance on how to share their data to increase the visibility of their work. You will learn about the **journal guidelines** (PloS, Nature Publishing Group) **concerning data sharing**. You will learn about **data paper** (Scientific Data, a Nature Research Journal) as well as Zenodo and figshare, **two adapted data repositories to meet journal requirements** for publishing biomedical research data underlying their publication. We will present how making published work and their accompanying datasets freely accessible through Open Access can benefit both researchers and the scientific community.

This workshop will provide researchers with **tools to generate robust and excellent quality studies that are reproducible and reusable**. Importantly, it will provide you with effective support to produce high quality publications complying with the guidelines established by journal publishers and funding agencies.

The workshop is **free** and **open to researchers in the biomedical field from UNIL, CHUV, EPFL, UNIGE and HUG at all career stages** as well as **to scientific information specialists.**

Location: Department of Fundamental Neurosciences (DNF), Rue du Bugnon 9, Metro M2 stop "OURS", Lausanne.

For registration use the link below:
Workshop III – "Data management & Open Data" (May 22, 2017)

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Open Science and Reproducibility Series
Programme for Workshop III "Data Management & Open Data"
Monday May 22, 2017, Department of Fundamental Neurosciences, UNIL

Time	
08.45 - 09.15	**Arrival Coffee**
09.15 - 09.30	Welcome venue from Prof. Nicolas Fasel, Vice-Dean of Research and Innovation, Faculty of Biology and Medicine, UNIL/CHUV.
09.30- 09.45	"The Open Science Policy of UNIL", Prof. François Bussy, Vice Rector for Research and International Relations, UNIL.
09.45 – 10.30	"Future implementation of SNSF policies on Data Management Plan and Open Research Data", Dr. Sarah Gerster & Dr. Lionel Perini, SNSF.
10.30 – 11.15	"Data Life Cycle Management and Data Management Plan: an Introduction", Dr. Aude Dieudé, EPFL & DLCM.
11.15 – 12.00	"Data life cycle management in large scale projects", Dr. Mark Ibberson & Dr. Robin Liechti, VitalIT, SIB.
12.00 – 13.15	**Lunch "Mauro Buffet"**
13.15 – 14.00	"Data sharing, credit and re-use: who's accountable?", Dr. Catriona MacCallum, PLOS Advocacy Director.
14.00 – 14.15	"Open Science at EPFL", Prof. Pierre Vandergheynst, Vice President for education, EPFL.
14.15 – 15.00	"Sharing your data and software on Zenodo", L Holm Nielsen, Manager of the H2020 Data repository Zenodo.
15.00 – 15.30	**Coffee Break**
15.30 – 16.15	"The State of Open and FAIR Data", Dr. Mark Hahnel, founder of the data repository figshare.
16.15 – 17.00	"Beyond supplementary material: sharing data effectively through repositories and data journals", Andrew L. Hufton, Managing Editor, *Scientific Data*, a Nature Research Journal.
17.00 – 17.45	"Data driven, executable articles in Authorea", Nathan Jenkins, founder of the platform Authorea.
17.45 – 18.00	Concluding remarks from Dr. Cécile Lebrand, CHUV Library, FBM, UNIL &CHUV.

✓ Welcome venue

Prof. Nicolas Fasel, Vice-Dean of Research and Innovation, Faculty of Biology and Medicine, UNIL/CHUV.

https://www.unil.ch/ib/en/home/menuinst/research/fasel--nicolas.html

Nicolas Fasel is full professor at the Faculty of Biology and Medicine at the University of Lausanne. After studying biology at the University of Fribourg (Switzerland) and obtaining a doctoral degree at the Swiss Institute for Experimental Cancer Research working on mouse mammary tumor virus, he took up a post-doctoral position at the University of California Los Angeles working on immunoglobulin gene regulation. On his return to Switzerland, he studied post-translational modifications of cell surface antigens. As an independant researcher of the Dr. Max Cloëtta Research Foundation, he had the opportunity to establish his own group investigating the molecular and cellular biology of protozoan parasites. His more recent research focuses on the host-pathogen interaction and the impact of co-infections. From 2003 to 2016, he directed the Department of Biochemistry of the Faculty and since August 2015 he is the Vice-Dean for Research and Innovation of the Faculty of Biology and Medicine.

✓ The Open Science Policy of UNIL

Prof. François Bussy, Vice Rector for Research and International Relations, UNIL.

https://www.unil.ch/central/en/home/menuinst/organisation/direction/f-bussy.html
http://www.unil.ch/iste/en/home/menuinst/recherche.html

François Bussy, 56, has left his position as Dean of the Faculty of Geosciences and Environment. He is now taking over as Vice Rector for research.
"I am deeply attached to this institution. After all the opportunities it has offered me, first as a student and then in research, it's time for me to devote my energy to it in return."
Does he have a particular aim in mind? "I don't have a specific agenda. But my primary role as Vice Rector for research is to motivate people and encourage them to develop their ideas. As my predecessor Philippe Moreillon used to say, we are facilitators."
Among other things, François Bussy will be able to draw on his experience as someone who has worked in the field to prompt as many people as possible to submit funding requests to organisations such as the SNSF. "It's within the reach of every researcher. And projects are evaluated by independent scientists from outside the university, which is essential."

✓ Future implementation of SNSF policies on Data Management Plan and Open Research Data.

Dr. Sarah Gerster & Dr. Lionel Perini, Swiss National Science Foundation

Dr. Sarah Gerster
Scientific Officer
Math, Natural and Engineering Sciences division
Swiss National Science Foundation

Dr. Lionel Perini
Scientific Officer
Humanities and Social Sciences division
Swiss National Science Foundation

Open Research Data is a fundamental contribution to the impact, transparency and reproducibility of scientific research. A number of initiatives around the globe are striving to make science and, in particular, research data accessible to all. The Swiss National Science Foundation (SNSF) agrees with the underlying principles of these initiatives.

The SNSF is introducing a new requirement in project funding: as of October 2017, researchers will have to submit a Data Management Plan as an integral part of their research proposal. Data management plans are an important step towards Open Research Data, and encourage researchers to think about the lifecycle of their data before starting on their project. In our presentation, we will discuss SNSF's policy on Open Research Data and the upcoming SNSF Data Management Plans."

✓ Data Life Cycle Management and Data Management Plan: an Introduction

Dr. Aude Dieudé, EPFL & DLCM.
http://library.epfl.ch/research-data/en
https://www.dlcm.ch/

What is exactly a Data Management Plan (DMP)? To what extent would a DMP be useful for you and what are the resources available for you to create one? This session will provide an introduction to demystify what is a DMP exactly, why does it matter and how it is not rocket science for you to prepare one.

✓ Data life cycle management in large scale projects

Dr. Mark Ibberson & Dr. Robin Liechti, VitalIT, SIB.
Vital-IT Group, Department of Systems Medecine et Biology, SIB Swiss Institute of Bioinformatics
https://www.vital-it.ch/

Dr. Mark Ibberson
Senior Scientist

Dr. Robin Liechti
Senior Scientist

Most modern research projects generate much more data than they can analyze during the lifetime of the project. It is thus important to ensure the reusability of such a resource by the scientific community, so that the value and return on investment of the project is maximized. Such well maintained data resources also provide a wealth of results and knowledge that can provide a fuel for new discovery efforts, rather than sitting on a laptop hard disk or worse, being deleted once the scientist has left the lab, or the project has finished.

In this presentation we will discuss the different stages of data life cycle management applied by us. These are: (i) capturing diverse data from different locations onto a centralized platform, (ii) standardization and curation of different data types, (iii) sharing and exploration of data, (iv) sustainability and interoperability of a data resource. We will use as an example a large European project on type 2 diabetes where clinical, phenotypic, genomics and metabolomics data were generated by a number of groups across Europe over a five-year period. The data were managed centrally on a dedicated platform where solutions for data integration, systems biology analysis and long-term sustainability were developed. Experience from this project is now being replicated in other projects within the Vital-IT competence center, which provides the expertise for a complete data life cycle.

✓ Data sharing, credit and re-use: who's accountable?

Dr. Catriona MacCallum, PhD, PLOS Advocacy Director
Member of the Boards, OASPA & OpenAire
http://orcid.org/0000-0001-9623-2225
https://www.plos.org/

PLOS introduced an updated data sharing policy in 2014 that requires the data underpinning the conclusions of an article be released at the time of publication. Our experience has shown that data sharing on a large scale is possible, while still a work in progress. For data to be shared effectively, however, it must be collected, analysed, transformed and stored with the intention to share. Data sharing must therefore be an integral part of the research cycle, not an afterthought at publication. Although challenging, there are emerging services, standards and best practices that in time can overcome the technical and infrastructure barriers to sharing data. But the lack of academic credit for data sharing remains the major impedim https://www.plos.org/ent. While the vast majority of PLOS authors voluntarily comply with our policy, in some communities, concerns have been expressed about the potential impact on future publications. As

long as the only form of academic credit is a first-author article publication, the pressure to publish will continue to undermine efforts to facilitate data sharing. The opportunity is to transform the existing culture into one in which data sharing, transparent reporting and good data stewardship are given at least as much prominence and status as journal publications. We argue that providing academic credit for data producers and curators is the best way to counter the perverse pressure on researchers and are implementing ways to do this. Without tangible benefits to individual researchers, through funding and career opportunities, there also remains no incentive for institutions, publishers or journals to change their current practice. To make this happen, all stakeholders need to work together to ensure that data-sharing and stewardship policies are aligned, regardless of business model or national and international jurisdiction.

✓ Open Science at EPFL

Prof. Pierre Vandergheynst, Vice President for education, EPFL.

http: //direction.epfl.ch/VPE
https://lts2.epfl.ch/

As of 1 January 2017, **Pierre Vandergheynst** will head EPFL's Vice Presidency for Education (VPE). Pierre Vandergheynst holds a PhD in mathematical physics from the Université Catholique de Louvain. His career at EPFL began with a post-doc in the Signal Processing Laboratory, which was funded by a research partnership with Logitech. He currently runs the Signal Processing Laboratory 2 (LTS2). He was appointed Vice Provost for Education in 2015, a position in which he set in motion a number of reforms, including the introduction of the Review Course. Throughout his time at EPFL, Professor Vandergheynst has pursued his commitment to education as director of the electrical engineering doctoral program and director of the electrical engineering section, as well as through his involvement in setting up a first-year course on global issues in conjunction with the College of Humanities.

✓ Sharing your data and software on Zenodo

L Holm Nielsen, Manager of the H2020 Data repository Zenodo.

http://about.zenodo.org/

To fully understand and reproduce research performed by others, it is necessary to have all the details. In the digital age, that means all the digital artefacts, which are all welcomed in Zenodo.

To be an effective catch-all, that eliminates barriers to adopting data sharing practices, Zenodo does not impose any requirements on format, size, access restrictions or licence. Quite literally we wish there to be no reason for researchers not to share!

Data, software and other artefacts in support of publications may be the core, but equally welcome are the materials associated with the conferences, projects or the institutions themselves, all of which are necessary to understand the scholarly process.

✓ The State of Open and FAIR Data

Dr. Mark Hahnel, founder of the data repository figshare.

https://orcid.org/0000-0003-4741-0309
https://figshare.com/about

There has been much talk around FAIR repositories - making content in a repository Findable, Accessible, Interoperable, and Discoverable - to help create efficiencies throughout the research workflow and allowing researchers to build on data and research that came before them. Figshare works with researchers, librarians and publisher to help bridge this gap and connect the valuable underlying data to the article, the institution and the researcher themselves, allowing for more credit for non-traditional outputs of research to spur scientific discovery and incentivize data sharing.

FAIR data relies on open APIs. This provides a great framework for research data management to be embedded into the fabric of national, regional and institutional plans to improve information pipelines and reduce administrative burden on academics. Good data management and infrastructure is at the foundation of reproducible research. By encouraging publishing of figures, data, code, and more rather than being limited to the traditional entire 'paper', knowledge can be shared more quickly and effectively in a transparent, reproducible fashion. Our ultimate focus to aid in the reproducibility, replication, and reuse of research data. Where possible, our SaaS layer looks to integrate with local and national infrastructure. This presentation will show how, by providing valuable infrastructure and bringing non-traditional research outputs to the forefront, discoverability and data reuse can raise institutional profiles as improve reproducibility and trust.

✓ Beyond supplementary material: sharing data effectively through repositories and data journals

Andrew L. Hufton, Managing Editor, *Scientific Data*, a Nature Research Journal.
http://www.nature.com/sdata/about/editorial-board
http://www.nature.com/sdata /

The Nature Research journals understand that effective data sharing supports reproducibility and can increase the impact of published works. Indeed, our policies have long recognized that data sharing is a fundamental part of research publication. The increasing complexity and size of research datasets, however, poses challenges for scientists who wish to share their data in a reusable and transparent manner. Based on my experience at *Scientific Data*, an open-access data-focused journal from Nature Research, I will provide tips on how researchers can share their data in an effective manner that promotes reuse, supports the credibility of their research, and ensures they get proper credit. This will include advice on writing better data-rich papers, the basics of presenting datasets in a useful manner, and tips on how to find the right repository for your data. I will also explain Scientific Data's editorial policies and share some of our experiences peer-reviewing and publishing data so far.

✓ Data driven, executable articles in Authorea

Nathan Jenkins, co-founder of the platform Authorea.

https://www.authorea.com/
https://www.authorea.com/product

Authorea is where more than 65,000 researchers in fields from Astronomy to Zoology write their documents online.
Behind the editor and the beautiful interface, Authorea is an incredibly powerful platform. Here are some the features that make it the most advanced scientific writing and collaborating platform.
In addition to static tables and figures, researchers can add interactive figures, live data, equations, LaTeX, and dynamic charts.
Everybody has their own flow when it comes to managing citations and references. You can drag and drop from your preferred library, or you can use Authorea's instant search to locate a reference by author, keyword, or DOI.
Write with as many co-authors as you need on the same document. The document reflects changes in real-time and enables commenting and version control. Write together across different formats
Browse a document's history to view the latest contributions and revert any unwanted changes.
Documents are faster to write and easier to understand when data is hosted underneath figures. Each Authorea document has a file repository to link data to specific tables and figures.