Enabling reproducibility in data science - learn why it matters and how you can do it

Report of Contributions

Contribution ID: **1** Type: **not specified**

Practical steps towards reproducible research

Thursday 9 June 2022 10:30 (2 hours)

In this workshop we will look at four key steps to get started with reproducible research:

(1) Get organized
(2) Use Open Source Software
(3) Use Version Control
(4) Make your work available online.

Although reproducible research practices can first seem intimidating, we will find a way to get started that works for you.

Presenter: SEIBOLD, Heidi

Session Classification: Workshop session

Contribution ID: **2** Type: **not specified**

Scientific metadata: Fundamentals of structured and standardized research data annotation

Thursday 9 June 2022 10:30 (2 hours)

This course is limited to 25 participants

Did you ever feel lost in incomprehensible research data documentation? This session introduces the basics of machine-readable research data annotation with domain-specific metadata schemas and standards. Learn, why accurate and harmonized description of research data is key for scientific exchange and how to find a suitable metadata framework in your research domain.

Presenters: STRUPP, Annika; GERLICH, Silke (HMC)

Session Classification: Workshop session

Contribution ID: **3** Type: **not specified**

Foundations of research software publication

Thursday 9 June 2022 10:30 (2 hours)

We will provide you with actionable advice about how to prepare your research code before publishing it or submitting it alongside a research publication.

This talk will cover the following topics:
• Code repository structuring
• Minimum coding practices
• Documentation
• Open source licensing
• Minimum software release practices
• Software citation

We will discuss these topics at the example of a data analysis script and will focus on minimum practices for every topic.

Presenter: SCHLAUCH, Tobias (DLR)

Session Classification: Workshop session

Contribution ID: **4** Type: **not specified**

Impulse lecture: Reproducibility in data science and machine learning

Thursday 9 June 2022 10:00 (20 minutes)

Machine Learning is becoming ubiquitous in many scientific domains. However, practitioners struggle to apply every new addition to the Machine Learning market on their data with comparable effects than published. In this talk, I'd like to present recent observations on reproducibility of Machine Learning results and how the community strives to tackle related challenges.

Presenter: STEINBACH, Peter (HZDR)

Contribution ID: **5** Type: **not specified**

Impulse lecture: Open science & reproducibility

Thursday 9 June 2022 09:40 (20 minutes)

Research is reproducible when it is possible to (independently) recreate the same results from the same data and same code/analysis as used by the original researcher or team of researchers. Reproducibility enhances collaboration and transparency in science and supports reusability of scientific products. This closely links with the open science endeavour towards the cultural change in science and science communication. Open science aims for a more effective and open exchange of information within science and the promotion of methods and the transfer of scientific results to society, the economy, and politics.

Presenter: HELMHOLTZ OPEN SCIENCE OFFICE

Contribution ID: **6** Type: **not specified**

Welcome

Thursday 9 June 2022 09:30 (10 minutes)

Presenters: HIDA; HELMHOLTZ OPEN SCIENCE OFFICE

Contribution ID: **7** Type: **not specified**

Wrap-up

Thursday 9 June 2022 12:30 (30 minutes)

Primary authors: STRUPP, Annika; SEIBOLD, Heidi; GERLICH, Silke (HMC); SCHLAUCH, Tobias (DLR)

Presenters: STRUPP, Annika; SEIBOLD, Heidi; GERLICH, Silke (HMC); SCHLAUCH, Tobias (DLR)