

Enroute to a sustainable PhD project?

Data management during the PhD phase

José Alejandro Cruces

How much data can it possibly be?

Two Docotral Researchers as examples:

Neuroscientist



800 days of stroke
patient screening

110 patient datasets
(written consent)

Publication on 48
datasets

- No consent
- Early release from hospital
- Exclusion criteria

- Early release from hospital
- Exclusion criteria
- Missing test score e.g. due to inability to read

Information
from 62
incomplete
datasets

787,06 GB data
581.674 files

How much data can it possibly be?

Two Doctoral Researchers as examples:

Materials Scientist



Synthesis of 2 different advanced alloys

Characterisation of Microstructure and Composition

Mechanical Deformation and Analysis of results

- Samples excluded which are not single-crystalline
- Problems with composition and sample size

- Samples for mechanical characterisation are small
- Sample need to be oriented appropriately

Information from 62 incomplete datasets

787,06 GB data
581.674 files

Who asks for sustainable data management?

Sustainable data management regulated by erc, DFG



European Research Council
Established by the European Commission



- Open Research Data and Data Management Plans
- Version 3.1 (July 2019)
- FAIR data principles (Wilkinson et al.)
- erc grants from 2017 onwards underlie the open research data pilot following 5 principles

- Principles for the Handling of Research Data (June 2010)
- Agreed upon by alliance of German research organisations
- Includes 6 central principles
- DFG asks applicants to openly address data handling and conservation in applications

How can data be handle sustainably?

Small and big tools for data management and storage

Data
lad



GIN
(G-Node Infrastructure)

SciNote



labfolder



GitHub

How realistic is sustainable data management?

Soft factors of sustainable data management

Employment

- Project Length
- Payment during Project
- Wrap-up Phase

Education

- Universities
- Graduate Schools
- Supervisors, PIs, Group and Institute heads

Infrastructure

- Common Guidelines
- Financial Support
- HGF Data management infrastructure
- Data Journals

Questions for Discussion

- How realistic is the publication of „failed experiments“?
- Sustainable data management requires standards – Who formulates these?
- When will the HGF have a generalised and centralised data management system and portfolio?

Who asks for sustainable data management?

The principles of sustainable data management



European Research Council
Established by the European Commission



- Data set description
- Standards and meta Data
- Name and persistent identifier for all the data sets
- Curation and preservation methodology
- Data sharing methodology

- Preservation and accessibility
- Differences between the scientific disciplines
- Scientific recognition
- Teaching and qualification
- Use of standards
- Development of infrastructures