Research Data Management at the Institute for Nuclear Physics:

The Example of the Research Cluster ELEMENTS

Johann Isaak
Technische Universität Darmstadt

Spokespersons:
Prof. Dr. Luciano Rezzolla (GU Frankfurt)
Prof. Dr. Tetyana Galatyuk (TU Darmstadt, GSI)
Prof. Dr. Dr. h.c. mult. Norbert Pietralla (TU Darmstadt)
Research in ELEMENTS

... addresses the physics of gravity, hadrons, nuclei, and atoms with numerical simulations and accelerator-based experiments.

From microscopic dynamics to the equation of state (EOS) of dense nuclear matter

From collisions of heavy ions to collisions of neutron stars

Nucleosynthesis of heavy elements

Electromagnetic signals from compact stars
Research in ELEMENTS

Macrophysics:
- Neutron-star mergers and supernovae
- Lightcurves and nucleosynthesis

Microphysics:
- Matter under extreme conditions
- Nuclear structure
- Nuclear and atomic reactions

Infrastructure:
- Particle accelerators
- GSI/FAIR & S-DALINAC
Research data in ELEMENTS

Expected generated data very diverse

- measurement of raw data / signals (experiments with stable & unstable nuclei, astronomical observations, ...)

- calculation of observables with different theoretical approaches (chiral EFT, EDF, lattice QCD, ...)

- analysis software (Python, C++, ROOT, ...)

- visualization of data (nuclear spectra, EOS, ...)

diverse projects with huge differences in data formats, processing, data sizes and storage requirements
Data management plan (DMP)

- The tool to support the planning and implementation of research data management at TU Darmstadt.

https://tudmo.ulb.tu-darmstadt.de/

- systematically deal with your research data from the very beginning!
- important to make your data interpretable and reusable for later time; also for third parties
- similar platform: Goethe-RDMO (https://rdmo.server.uni-frankfurt.de/)
Data management plan (DMP)

- Systematically deal with your research data from the very beginning!
- Important to make your data interpretable and reusable for later time; also for third parties
- Similar platform: Goethe-RDMO (https://rdmo.server.uni-frankfurt.de/)
- Each project: collaboratively create and maintain a DMP in the beginning
- Update DMP regularly (every six months)
- Templates being continuously further developed towards our field-specific needs

https://tudmo.ulb.tu-darmstadt.de/
RDM workflow

too large to be stored, irreversibility of data loss

raw data ~ PB

aggregated data ~ GB-TB

local storage, TUDatalib, GUDe

HEPdata, TUDatalib, GUDe

arXiv and/or Open Access

triggers + from detector signals to tracks + calibration + data QA

analysis

data for plots kB-MB

publication

select + pre-analysis

simulation output GB-TB

aggregated data ~ GB-TB

HEPdata, TUDatalib, GUDe

arXiv and/or Open Access

not necessarily stored, sustainable workflow, software is published

analysis

data for plots kB-MB

publication
Data classification: difference in scales of data sizes

- raw data \( \sim \) PB
- aggregated data \( \sim \) GB-TB
- data for plots kB-MB

- simulation output GB-TB
- aggregated data \( \sim \) GB-TB
- data for plots kB-MB
Data classification: difference in scales of data sizes

local storage: within projects

- raw data ~ PB
- aggregated data ~ GB-TB
- data for plots kB-MB
- simulation output GB-TB
- aggregated data ~ GB-TB
- data for plots kB-MB
Data classification: difference in scales of data sizes

access: upon request

local storage: within projects

- raw data ~ PB
- aggregated data ~ GB-TB
- data for plots kB-MB

- simulation output GB-TB
- aggregated data ~ GB-TB
- data for plots kB-MB
Data classification: difference in scales of data sizes

- Raw data: ~ PB
- Aggregated data: ~ GB-TB
- Data for plots: kB-MB
- Simulation output: GB-TB
- Aggregated data: ~ GB-TB
- Data for plots: kB-MB

Access: preferably open

Institutional long-term archiv: IKPStore, TUdataLib or GUDe
Data classification: difference in scales of data sizes

- **Domain-specific repository:** HEPdata
- **Institutional/generic repository:** TUdatalib, Zenodo, GitHub
- **Access:** open access & published
Data classification: difference in scales of data sizes

- **local storage:** within projects

- **access:** upon request

- **domain-specific repository:** HEPdata

- **institutional/generic repository:** TUdatalib, Zenodo, GitHub

- **access:** preferably open

- **institutional long-term archiv:** IKPStore, TUdatalib or GUDe

- **access:** open access & published
Realization of a multi-turn energy recovery accelerator

Conventional electron linear accelerators are essential research tools but limited in providing high beam currents. Energy recovery technology
Accompanying data publication

- datasets stored locally at IKP / accelerator group
Accompanying data publication

- datasets stored locally at IKP / accelerator group
- manuscript submitted
- in parallel: preparation of data publication on TUdatalib inkl. analysis code
Accompanying data publication

- datasets stored locally at IKP / accelerator group
- manuscript submitted
- in parallel: preparation of data publication on TUdatalib inkl. analysis code

Data availability

The experimental data presented in this manuscript are available from a TUdatalib repository. Source data are provided with this paper.

Code availability

The codes that support the findings of this study are available from a TUdatalib repository.

published scientific article explicitly linked to data publication via DOI's
Peer-reviewed software publication

- code publicly available on GitHub
- peer-review of code in journal
- published via Zenodo with DOI
Documentation: from paper to digital logbooks

Marie Curie’s experimental notebook
https://twitter.com/wellcometrust/status/496323565239955456
Documentation: from paper to digital logbooks

Marie Curie’s experimental notebook
https://twitter.com/wellcometrust/status/496323565239955456
Marie Curie’s experimental notebook
https://twitter.com/wellcometrust/status/496323565239955456
Training & education in RDM

• introductory to RDM policy & common RDM tools
• regular training to raise awareness for RDM
• recent event: ECR-Day on October 19, 2023
Training & education in RDM

- introductory to RDM policy & common RDM tools
- regular training to raise awareness for RDM
- recent event: ECR-Day on October 19, 2023

- "good RDM from the very beginning..."
- train students in RDM at early stage of their studies
- implement RDM in the curriculum
- theoretical basics and practical application of methods and tools
- make sustainable RDM common practice!

development of (compact) training material
presentation / slides & hands-on exercises

Screenshot from a scene of "A serious man", 2009.
Research-oriented teaching: advanced lab courses

- Preparation
- Setup
- Measurement
- Analysis
- Visualization
- Report
Research-oriented teaching: advanced lab courses

- **TU demo**
- **ELOG**
- **TU datalib**

**Setup**
- Preparation
- Measurement
- Analysis
- Visualization

**Report**
- **LateX**
- **GitLab**
- **DOI**
Networking with other RDM initiatives in progress

**National research data infrastructures (NFDI)**

- DFG initiative to establish (inter)national coordinated access to valuable science and research data in a sustainable and qualitative manner

NFDI consortium of particle, astro-, astroparticle, hadron and nuclear physics
Networking with other RDM initiatives in progress

**National research data infrastructures (NFDI)**
- DFG initiative to establish (inter)national coordinated access to valuable science and research data in a sustainable and qualitative manner

**Helmholtz Metadata Collaboration (HMC)**
- "From the Past To the Future: Legacy Data in Small and Medium-Scale PUNCH Experiments - a Blueprint for PUNCH and Other Disciplines"

NFDI consortium of particle, astro-, astroparticle, hadron and nuclear physics

Many activities on FAIR data taking in the future

**But what about already existing datasets?**
- “FAIRification” of datasets on IKPStore
Networking with other RDM initiatives in progress

**National research data infrastructures (NFDI)**
- DFG initiative to establish (inter)national coordinated access to valuable science and research data in a sustainable and qualitative manner

**Helmholtz Metadata Collaboration (HMC)**
- "From the Past To the Future: Legacy Data in Small and Medium-Scale PUNCH Experiments - a Blueprint for PUNCH and Other Disciplines"

**Metadata for Nuclear Physics experiments in EURO-LABS partners**
- Joined effort to define common metadata schema

NFDI consortium of particle, astro-, astroparticle, hadron and nuclear physics

Many activities on FAIR data taking in the future

But what about already existing datasets?
- „FAIRification“ of datasets on IKPStore

Research communities of nuclear physics, accelerator and detector technologies for high energy physics