







GSI Helmholtz Centre for Heavy-ion Research



Heavy-ion accelerator laboratory in Darmstadt, Germany with 1350+ employees & 700 external users

Who we are?

- Founded in 1969
- Elements H to U can be accelerated

Past Achievements

- Discovering six new elements, many exotic nuclei
- Developing a new cancer treatment
- **Our Future**
- First stage injector for the FAIR facility
- **Experiments continuing** (FAIR Phase-0)







- **Research Interests**
- Accelerator Physics
- Detector Development
- Atomic Physics
- Nuclear and particle Physics
- Plasma Physics
- Biophysics
- Materials Research
- High-Performance Computing
- Theoretical Physics

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry - Research Data Management at GSI: An overview and an outlook

FAIR (Facility for Antiproton and Ion Research in Europe)

- New accelerator facility
- Top priority for European Nuclear Physics Community
- International: 50 countries, 3000 researchers
- Diverse community from atomic to particle physics
- 'FAIR goes F.A.I.R.': commitment to open science
- Towards the next generation "data challenge" Volume, Velocity, Veracity, Variety, and Complexity
- ~TB/s data rates, online processing, ~5x10⁵ cores
- Data stored to disk 40+ PB/year
- Distributed computing with a large user community
- Data preservation and accessibility key to success





FAIR ESS it



Research Data and Software at GSI: Rich and varied



GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry – Research Data Management at GSI: An overview and an outlook

Research Data at GSI: Metadata Complexities

Large metadata: materials research

Raw material

- Polymer
- Size
- Finish
- Manufacturer
- Purchase/Fabrication date
- Irradiation
 - Beamline
 - Ion species
 - Charge state
 - Energy
 - Flux
 - Fluence
 - Date

Pore etching

- Bath
- F. Koch Research Data Management at GSI/FAIR (2022)
- Duration

GSI Helmholtzzentrum für Schwerionenforschung GmbH

- Sputter Coating
 - Material
 - Parameters
 - Date
- Back Electrode Deposition
 - Material
 - Electroplating Bath
 - Time
 - Current/Voltage
 - Date
- Wire Deposition
 - Material
 - Electroplating Bath
 - Time
 - Current/Voltage
 - Date

- XRD Analysis
 - Scan Type
 - Angular Range
 - Tube Parameters
 - Recorded Intensity
 - Date
- Dissolution of Template
 - etchant
 - duration
 - drying method
 - Date
- Offline SEM Analysis
 - Scan mode
 - Detector Type
 - Acceleration Voltage
 - Magnification
 - Date
 - Wire length
 - Wire diameter

Online SEM Analysis

- Scan mode
- Acceleration Voltage
- Magnification
- Online Electrical Analysis
 - Manipulator Type
 - Image of contact
 - Wire length/Diameter
 - I/V-curve
- Irradiation
 - Ion species
 - Charge state
 - Flux
 - Fluence
 - Date



Andrew Mistry - Research Data Management at GSI: An overview and an outlook



What is Open Science?



- Science lives on the <u>Open exchange of Knowledge</u>; This open culture of scientific endeavor is captured by the term "<u>Open Science</u>"
- In essence: Make research results (+ infrastructure) openly available to science, industry, and society for reuse with as <u>few barriers as possible</u> in sustainable infrastructures
- Defined by: <u>Openness, Transparency, Sustainability, Transfer,</u> <u>Collaboration and Sharing</u>
- How to <u>define and shape Open Science practices</u>, tools and dissemination in a way that **maximises the rewards and benefits**?
- Many challenges are faced:
 - How open do we want to go?
 - What resources do we require?
 - How do we showcase the benefits of Open Science to the communities?



Why is Open Science important?





Accelerates knowledge transfer by breaking down access barriers to research outputs. -> Synergies with AI



Fosters collaboration within and across disciplines, leading to quality improvements and new solutions.



Open Science promotes transparency, building trust among researchers and the public.



Attracting future researchers: Open Science signals inclusivity and appeals to diverse talent.



Sustainability improves as resource sharing reduces repetitiveness



Offers a new metric for research assessment to remove the outdated dependence on e.g. h-index



Strengthens technology transfer with industry partners



On the agenda of many governments and funding bodies (Helmholtz POF IV+ requires it) Ultimately, the researcher should benefit from making research outputs open

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry – Research Data Management at GSI: An overview and an outlook

Findable Centrally orchestrated storage and access of data essential to enable the data/software to be findable.

https://www.go-fair.org/fair-principles/

- Usage of Persistent IDentifiers (PID), Digital Object Identifiers (DOI) -> Guarantee access to Digital Research Objects.
- · Generation of 'data record' in discipline specific repositories

F.A.I.R. Principles outlines (briefly for info)

Accessible

- Data and software produced/dedicated for F.A.I.R communities and publications centrally stored
- Common & "user-friendly" interface to store and retrieve data

Interoperable

- Common metadata formats
- F.A.I.R-produced data operable with other datasets

Reusable

- Ensure (as reasonably as possible) data stored long term
- Metadata should be retained indefinitely



Image credit: ARDC licensed under a Creative Commons Attribution 4.0 International License

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Findable Accessible Interoperable Reusable

Pn **60**

RGSI

Open Science @ GSI/FAIR: What do we want to achieve?

Open Access Publications : Mandatory publication of Open Access articles









Considerations:

- > The steps and processes to achieve this are complex... Start smaller and work up
- > Do not make it too 'general', needs finer granularity and use-cases
- Aim to address all researchers who use GSI/FAIR: Students, Postdocs, PI's, Group leaders...

Andrew Mistry - Research Data Management at GSI: An overview and an outlook



GSI/FAIR Open Science Working Group: Testing and implementing Open Science practices



GSI/FAIR Open Science related Policies/Guidelines





Info and Material: GSI/FAIR Open Science Website



Email address: open-science@gsi.de

GSI/FAIR Open Science Webpage: https://www.gsi.de/open-science



What is Open Science?

Open Science is the practice of making scientific research output openly available in the form of data, software, publications, hardware and infrastructure. This promotes transparency, collaboration, and reproducibility in research, as well as wider access to knowledge for the public and to researchers.

GSI and FAIR are committed to Open Science practices and provide tools, support, and information to internal and external researchers involved in GSI/FAIR projects. Organisations such as <u>cr</u> UNESCO, the <u>cr</u> <u>DFG</u>, the <u>cr</u> <u>BMBF</u> and <u>cr</u> <u>Helmholtz</u> among many others have recognized the benefits of Open Science, and have issued recommendations to support the movement.

The GSIFAIR Open Science Working Group hosts monthly meetings to promote and advance Open Science at within the facilities. Membership of this group comprises researchers from a variety of disciplines, as well as members from the accelerator division, Grant Office, Technology Transfer, IT, and Library and Documentation.

Adopting Open Science principles aligns with good scientific practice, and more information on this can be found on the g^o GSI/FAIR Ethics and Rules webpage

The GSI policy on Research Data Management can be found in here

The GSI guidelines on Software licences can be found in here (Internal only).

Open Access of publications	
Open Data	
Data Management Planning	
Data Publication	
Open Software	
Experiment logging and Notebooks	
Links to Open Science Projects GSI/FAIR Involvement	
Additional Material and Training	

GSI/FAIR Involvement in External OS initiatives



ESCAPE	European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures: Open Source Software Repository (OSSR) developer and maintainer
Nu PE(C	Nuclear Physics European Collaboration Committee: Participation and writing Open Science section of the LRP 2024
neosc	European Open Science Cloud: GSI/FAIR both observer members, contribution and suggestions for EOSC Future
	EuroLabs: Work Package on Open, diverse and inclusive Science
PUNCH 4 N F D I	Particles, Universe, NuClei and Hadrons for the NFDI: Two Task areas; Developments on data portal, AAI, data lake and other infrastructure from GSI IT department and Research division
MTM	Matter and Technology, Data Management and Analysis: IT contributions
HELMHOLTZ Open Science	HGF Open Science: Members of the OS, software and POF IV indicators working groups
CHMC> HELMHOLTZ Metadata Collaboration	Helmholtz Metadata Collaboration: Participation in HMC funded projects, links and connections to Matter division
	Exploring the Universe from Microscopic to Macroscopic Scales: Supporting Open Science area of the project (as well as other direct research areas)

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry - Research Data Management at GSI: An overview and an outlook

GSI

6

4

2

Dissemination: RDM/Open Science workshops @ GSI

2022 RDM Slides and information: https://indico.gsi.de/event/14680/ 2023 Open Science Slides and information: https://indico.gsi.de/event/17498/

100+ *participants* per meeting in two hybrid events,

Variety of themes dicussed:

- Inform and make researchers aware of RDM/Open Science, plans and progress
- Talks from: individual research groups, external projects, Tech Transfer, Helmholtz Open Science, Library, Grant office, IT department
- **RDM guestionnaire** distributed after workshop->identify areas to address





13

Goals

- to ensure good RDM practices at GSI/FAIR;
- promote and assist researchers in 'publishing' data;
- to aim (as best as reasonably possible) that data is published according to the Findable Accessible Interoperable and **Reusable** principles;
- develop the tools and infrastructure needed to do this (repositories, Electronic logbooks ...)

FAIR Data is not an end goal

continual process of improving practices and adapting research. resources with technology innovations

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry – Research Data Management at GSI: An overview and an outlook

Open Research Data and Management

Research Data Management

encompasses all aspects of handling research data, from planning, its generation and processing to publication, longterm archiving, and eventual deletion, while adhering to the principles of good scientific practice.







Current Tools for Open Science at GSI





Large volume data storage + archiving (Lustre/LTSM/FSQ) - I-U-S-t-r-e-



Compute Cluster Virgo



Data Management Planning Software (RDMO): New





Electronic Logbooks (ELOG)



Collaboration Tools (Wiki, Indico, Mattermost, Seafile)





Internal publication repository (JOIN2)



Code management system (GitLab)



 $(J)_{n^2}$



LxFRS -> new FRS tool!

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Tools for Open Science at GSI: Coming soon examples



<MyCoRe/> New GSI publications repository: Testing 2024 -> Go Live 2025

Ecosystems/External projects PUNCH4NFDI, EuroLabs: 2023 - 2027

Instrument/Infrastructure PIDs: Conceptual phase

Metadata schema (NAPMIX): 2024/25/26



16



HELMHOLTZ

EUR@+LABS





RDM in practice: Data Management Plans

schema...)



Living document • Prepared at the start and follows the research data (RD) lifecycle through

Juseful for IT/Resource users and encourage preparation Funding he leasy for users mates V Goal Make it easy for jents Data management **Data Management** Lifecvcle Contents incl. RD information (Data type, size, scope, and generation method...) Overhead (Data protection, costs, project planning...) Achieving and reuse (repository selection, data to be published, metadata

RDM in practice: Data Management Plans

✓ Goal : Make RDM easy for users and encourage preparation



GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry – Research Data Management at GSI: An overview and an outlook

RDM in practice: Data Management Plans



- Data Management with easy to use tool: RDMO (Research Data Management Organiser)
- Currently running production version at GSI/FAIR
- GSI/FAIR tailored catalogue available
- https://www.rdmo.gsi.de
- Highly Recommended to prepare one for your project!
- For 2024 Experimental Proposal Call:

Research Data Management Plan (RDMP):

Currently, a RDMP is not obligatory but recommended for accepted experiment proposals.





- What should I publish? -> Raw; Pre-processed; Result...
- What do I need to replicate my result? -> Software; Workflows...
- Where should I publish? -> Zenodo; HEPData...
- How do I 'publish' data? -> Uploading, linking
- How much time will it take? -> 1 day, 3 months...
- What Licence should I choose? -> CC-BY?
- How do I accurately describe my data? -> Metadata; Written Explanations...
- How do I ensure that the data is findable? -> Linking to article/software?...

RDM in practice: What should I publish?



[What do I need to replicate the result(s)?]

Raw -> Ideal FAIR data case? Often too large; difficult to process.

Pre-processed -> Feasible option; needs some processing and analysis. Allows further examination of the data

Result -> 'Easiest' case. Can be the data used to make plots etc.



Software -> analysis scripts;

Workflows -> Containers, Descriptions of processing procedures

RDM in practice: Where/How do I publish?



Interim strategy



GSI Helmholtzzentrum für Schwerionenforschung GmbH

RDM in practice: Where/How do I publish? [Example]



[What do I need to replicate the result(s)?]

Starting out and keeping it simple: Result data + analysis scripts



GSI Helmholtzzentrum für Schwerionenforschung GmbH

RDM in practice: Where/How do I publish? [Example]





RDM in practice: Where/How do I publish? [Example]



[What do I need to replicate the result(s)?]

Another step forward: semi-derived data, software codes, result data and accompanying metadata





Article -> Dataset -> Software linking

Zenodo: Dataset Record

Publications Repository: Journal Article Record

GSI



Datasets/software records in GSI publications repository count towards POF IV

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry – Research Data Management at GSI: An overview and an outlook

Metadata for nuclear physics experiments



How do I accurately describe my data?

Metadata essential in Research Data Management and Open Science to enable FAIR data (and code) More info: <u>Guide to HMC Better Metadata Booklet</u>

When publishing data, also publish machine readable metadata

- Allows datasets to be searched for and found
- Enables interoperability between datasets
- Enables reprocessing of data: transparency and integrity
- Efficient use of resources

However: No common schema yet between nuclear physics experiments - >This is needed

Caution: Don't reinvent the wheel! Look for synergies with NeXUS, HELPMI, OpenAIRE et al.



By Randall Munroe: <u>https://xkcd.com/927</u>

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Metadata for nuclear physics experiments



Project specific metadata Project name, Discipline publication date, collaborations, specific researchers, institutions, 2 metadata journal references **Experiment** Projectile data, target specific data, infrastructure, 3 metadata detector type, datasets, software repositories Reaction data, gammaray energies, time of flight parameters,

GSI Helmholtzzentrum für Schwerionenforschung GmbH

instrument records

Metadata for nuclear physics experiments:NAPMIX

https://www.oscars-project.eu/projects/napmix-nuclear-astro-and-particle-metadata-integration-experiments

The NAPMIX project emerged to address a significant gap in nuclear physics: the lack of a unified metadata schema necessary for achieving FAIR datasets. This challenge extends to the fields of particle and astro-particle physics, highlighting the need for a collaborative European effort to create a common metadata schema with user-friendly infrastructure. By integrating expertise across these domains, NAPMIX aims to enhance data management practices.



GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry – Research Data Management at GSI: An overview and an outlook

GSİ

Metadata for nuclear physics experiments: NAPMIX

- A front-end generator for user input, producing outputs in both machine and human-readable formats.
- A multi-layered 'nodal' structure to incorporate use cases based on real experimental datasets.
- Connections to existing ontologies such as DataCite, OpenPMD and NeXus.
- Interfaces to current infrastructures as a pathway to an Open Science Ecosystem.
- Testing and deployment of the schema with multiple use cases to enhance data FAIRness.
- Training for researchers on the importance and application of metadata.

> 24 Months (01/12/24 - 30/11/26)

Concept	Design	Implementatio n	Schema
Image: Second	Image: state		<pre>very set of the s</pre>

GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry – Research Data Management at GSI: An overview and an outlook



Instrument/Infrastructure Persistent Identifiers (PIDs)

For more information see e.g.:

https://datascience.codata.org/articles/10.5334/dsj-2020-018



- Goal: Have a persistent identifier for instruments/infrastructure at GSI
- > Why?
 - Informing users: act as a reference point giving insights into technical specifications and capabilities. This also aid discoverability of the device.
 - **Providing traceability:** ensuring that the devices can be traceable and distinguishable for effective monitoring.
 - **Direct citation:** not replacing any technical articles on instruments, but acting as a reference point that can be updated as needed.
 - Version control: keep updates and modifications reported.
 - Long-term accessibility and knowledge retention: ensure the long-term accessibility of information related to the instrument.
- What is it? A short (max 2 page) document/web document outlining the technical description of the device including figures. Include versioning and crossovers with other infrastructure
- How? Create the record and upload to the GSI publications repository under 'physical object'. A digital object identifier (DOI) will be assigned.

Instrument/Infrastructure Persistent Identifiers (PIDs)



FAIR 🖬 📻 🏦

GSI Helmholtzzentrum für Schwerionenforschung GmbH



GSI Helmholtzzentrum für Schwerionenforschung GmbH



GSI Helmholtzzentrum für Schwerionenforschung GmbH

Andrew Mistry - Research Data Management at GSI: An overview and an outlook

Open Science/RDM @ GSI/FAIR

- **Open Science** cultural change under development at GSI/FAIR, a step-wise iterative approach is underway to best suit the research areas
- We are working to develop **more complex infrastructures** for data/software/workflow F.A.I.R.'ness (with FAIR in mind....): RED/Library/IT department with PUNCH4NFDI, EURO-LABS, ESCAPE etc.

What can/should be done now?

- Data Management Plans
- **Open Access Publications**
- Concepts for Simple Data/Software publication
 - Engagement with collaborative tools

We can support: Data Management Plan preparation, Data and Software publication, Metadata development, and other Open Science ideas

Open Science related questions or Comments: <u>open-science@gsi.de</u> Website Open Science @ GSI/FAIR: https://www.gsi.de/open-science

OPEN DATA



RESOURCES

