Workshop: Strategies for Data Science and Data Management

Report of Contributions

https://indico.hiskp.uni-bonn.de/e/109

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

Welcome

Tuesday, January 17, 2023 9:15 AM (10 minutes)

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

Research Data Management at Bonn University

Tuesday, January 17, 2023 9:25 AM (30 minutes)

Research Data Management and the FAIR principles are becoming science policy mainstream. They are increasingly demanded on different levels by decision makers, funders and research institutions alike.
Yet, many scientific communities still lack established conventions, guidelines and infrastructures for handling research data properly. This makes it difficult for many researchers to implement a sound and practical RDM strategy that fulfils the FAIR principles.
The University of Bonn formed the Research Data Service Center (RDCS) in a move to provide the best possible support to researchers in addressing these challenges. The RDSC offers advice, training and basic IT services for handling research data.
This presentation gives an overview on the dynamically evolving RDM landscape in Europe and Germany and presents the current and planned support infrastructure for researchers at Bonn University.

Presenter: Dr BITTNER, Christian (Servicestelle Forschungsdaten SFD-ULB, Uni-Bonn)

Session Classification: Plenary Talks

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

New developments in artificial intelligence for data science

Tuesday, January 17, 2023 11:30 AM (30 minutes)

Presenter: Prof. BAUCKHAGE, Christian (IAI)

Session Classification: Plenary Talks

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

Open Science in Germany and Europe

Tuesday, January 17, 2023 9:55 AM (30 minutes)

Presenter: Prof. FÖRSTNER, Konrad (ZB MED)

Session Classification: Plenary Talks

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

High Performance Computing in Bonn

Wednesday, January 18, 2023 10:00 AM (30 minutes)

Infrastructures for high performance and high throughput computing.

Presenter: Dr BARBI, Dirk (HPC Team Leader at HRZ, University of Bonn)

Session Classification: Plenary Talks

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

Knowledge graphs for spatio-temporal data analytics

Tuesday, January 17, 2023 12:00 PM (30 minutes)

Presenter: Prof. DEMIDOVA, Elena (Institut für Informatik, Universität Bonn)

Session Classification: Plenary Talks

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

Digital Research Infrastructures at Bonn University

Tuesday, January 17, 2023 3:50 PM (30 minutes)

In this presentation, we will provide an overview of the technical infrastructure and services to support scientists at Bonn University throughout the research data life cycle. The overview includes services such as data storage infrastructure, research data repositories, and electronic lab notebooks.

Presenter: Dr ZERR, Sergej (Servicestelle Forschungsdaten SFD-HRZ, Uni-Bonn)

Session Classification: Plenary Talks

Contribution ID: **8** Type: **not specified**

NFDI4Health

Tuesday, January 17, 2023 1:30 PM (20 minutes)

Presenter: FLUCK, Juliane (uni-bonn)

Session Classification: Plenary Talks

Contribution ID: **9** Type: **not specified**

NFDI4Culture

Tuesday, January 17, 2023 1:50 PM (20 minutes)

Presenter: ZINGSHEIM, Domenic (Institute for Computer Science II (Uni Bonn))

Session Classification: Plenary Talks

Contribution ID: **10** Type: **not specified**

FairAgro

Tuesday, January 17, 2023 2:10 PM (20 minutes)

FAIRagro: FAIR Data Infrastructure for Agrosystems

Presenter: HAUNERT, Jan-Henrik (uni-bonn)

Session Classification: Plenary Talks

Contribution ID: **11** Type: **not specified**

PUNCH4NFDI

Tuesday, January 17, 2023 2:50 PM (20 minutes)

Presenter: BECHTLE, Philip (uni-bonn)

Session Classification: Plenary Talks

Contribution ID: **12** Type: **not specified**

NFDI4Earth

Session Classification: Plenary Talks

Contribution ID: **13** Type: **not specified**

NFDI4Microbiota

Session Classification: Plenary Talks

Contribution ID: **14** Type: **not specified**

NFDI4Objects

Tuesday, January 17, 2023 2:30 PM (20 minutes)

Presenter: LANG, Matthias (uni-bonn)

Session Classification: Plenary Talks

Contribution ID: **15** Type: **not specified**

Big data and life sciences

Wednesday, January 18, 2023 9:00 AM (30 minutes)

Presenters: SCHULTZE, Joachim L. (DZNE, Universität Bonn); SCHULTZE, Joachim (uni-bonn)

Session Classification: Plenary Talks

Contribution ID: **16** Type: **not specified**

Collaborative Software Development and Data Analysis Pipelines

Tuesday, January 17, 2023 11:00 AM (30 minutes)

Presenter: DELGROSSO, Nicholas

Session Classification: Plenary Talks

Contribution ID: **19** Type: **not specified**

Federated Digital Infrastructures for Astro and Particle Physics

Tuesday, January 17, 2023 3:30 PM (20 minutes)

Presenter: FREYERMUTH, Oliver (University of Bonn)

Session Classification: Plenary Talks

Contribution ID: **20** Type: **not specified**

Data-driven Research in the Life Sciences

Wednesday, January 18, 2023 9:30 AM (30 minutes)

Presenter: HASENAUER, Jan (Faculty of Mathematics and Natural Sciences, Rheinische Friedrich-Wilhelms-Universität Bonn, 53115 Bonn, Germany)

Session Classification: Plenary Talks

Contribution ID: **21** Type: **not specified**

Hands on session: Research documentation - Electronic Lab-Logbook RSpace

Wednesday, January 18, 2023 10:50 AM (2 hours)

Presenter: Dr RUDOLF, Daniel (Servicestelle Forschungsdaten SFD-ULB, Uni-Bonn)

Session Classification: Hands-On Tutorials

Contribution ID: **22** Type: **not specified**

How to document and share my data?

Wednesday, January 18, 2023 10:50 AM (2 hours)

Digitization is changing day-to-day research practices across all fields of academic inquiry. Research data of every kind are collected, processed, analyzed, published and archived in digital systems. The term Research Data Management (RDM) refers to a range of activities and topics relating to the handling of digital research data.
In this hands-on workshop the focus will be on data organization (folder structures, file naming, versioning), documentation (ways of data documentation, metadata: definition, examples and importance) and sharing (why data sharing is key, research data repositories, do's and don'ts about data sharing).

Presenter: Dr BRES, Ewa (Servicestelle Forschungsdaten SFD-ULB, Uni-Bonn)

Session Classification: Hands-On Tutorials

Contribution ID: **24** Type: **not specified**

Registration

Tuesday, January 17, 2023 8:30 AM (45 minutes)

Contribution ID: **25** Type: **Poster**

PhenoRob Data Management

Our poster presents the data management strategy of the cluster of excellence PhenoRob from the University of Bonn. PhenoRob is an interdisciplinary project that combines research from robotics and phenotyping and aims for sustainable crop production. Due to the interdisciplinarity, we deal with very heterogeneous data from different research fields, with multiple file formats, and varying sizes. To handle this data, we introduce a data concept that categorizes the data into three types: research data, metadata, and basis data. We consider research data to be data that is generated for a particular research task. Each research data set needs to be annotated with a metadata record. We want to emphasize that we developed the metadata scheme especially for the needs of PhenoRob, integrating existing metadata standards for geodata and plant phenotyping. The basis data comprises general information on the field sites that is of interest for many researchers. Our data management infrastructure can be accessed by the researchers via a web interface (www.phenoroam.phenorob.de).

Primary authors: VEDDER, Lucia (University of Bonn); BONERATH, Annika (University of Bonn)

Presenters: VEDDER, Lucia (University of Bonn); BONERATH, Annika (University of Bonn)

Session Classification: Posters

Contribution ID: **26** Type: **Poster**

Machine Learning based Similarity Models for Research Dataset Discovery Systems

This poster is focussing on improving the process of research dataset discovery by developing machine learning models that can estimate the similarity between two datasets. In addition to traditional keyword-based search methods, the relevance of a dataset can also be determined by its similarity to existing, relevant datasets. The proposed models incorporate metadata about the datasets as well as the scientific publications that cite and use them, known as the "context". The evaluation of these models shows that considering the context of a dataset leads to more accurate estimates of dataset similarity.

Primary authors: GEBREYOHANNES, Aberham (Bonn University); Dr ZERR, Sergej (Hochschulrechenzentrum, Bonn University)

Presenter: Dr ZERR, Sergej (Hochschulrechenzentrum, Bonn University)

Contribution ID: **27** Type: **Poster**

High Performance Computing at the University of Bonn

High Performance Computing (HPC) leverages the power of multiple compute nodes and architectures to solve complex problems, often with large data sets. Typical applications range from large scale simulations to machine learning and data analysis. The University of Bonn maintains several HPC clusters for specialized and general purposes, such as a massively parallel computing system with GPU accelerator partitions. This poster gives an overview of the HPC infrastructure and its future development, use cases as well as the HPC & Analytics Lab team, which is the main contact for scientific questions from HPC users of the University. We also briefly discuss the importance of research data management in the context of HPC.

Primary author: HIGH PERFORMANCE COMPUTING & ANALYTICS LAB, The (University of Bonn)

Presenter: HIGH PERFORMANCE COMPUTING & ANALYTICS LAB, The (University of Bonn)

Session Classification: Posters

Contribution ID: **28** Type: **Poster**

OpenMuseum - a data- and society-driven platform for Open Science

The OpenMuseum is a digital platform that provides access to the collections of the University of Bonn. The platform is a combined tool for research, cataloguing, teaching and outreach, thus providing an access to cultural heritage collections curated for diverse social groups and user communities. The OpenMuseum allows these collections to be used for their respective interests (FAIR and CARE principles) and also provides an infrastructure for analogue and virtual exhibition and mediation formats that encourages active engagement with the collections. Current challenges include establishing a common infrastructure and a workflow able to meet the needs of the different collections, their requested data types and expectations. This includes the optimisation for databases at varying levels of digitisation and a data model flexible enough to accommodate the different scientific fields and categories relevant to each collection. An additional virtual exhibition with novel features is planned for the latter half of the project.

Primary authors: Mrs STAUSS, Elizabeth (uni-bonn); GRIGOWSKI, Edouard (uni-bonn); Dr PALLAN, Carlos (uni-bonn); HANNIG, Alma (uni-bonn)

Presenter: Mrs STAUSS, Elizabeth (uni-bonn)

Session Classification: Posters

Contribution ID: **29** Type: **Poster**

Making neuroscientific analyses replicable: An open science fMRI preprocessing pipeline

Abstract

Open science principles (such as sharing ideas, data, and results; Merton, 1973) have not yet been fully adopted by the neuroscientific community impeding replications of research results (e.g., Poldrack et al., 2017). Apart from the actual data analysis, analytical flexibility regarding preprocessing of the MRI data can lead to varying study results (Botvinik-Nezer, et al., 2020). To enable reproducibility of neuroimaging analyses from our group, we have implemented an open science MRI preprocessing pipeline, which connects several open-source neuroimaging software and scripts: 1) conversion to a standard data format with HeuDiConv (Halchenko et al., 2017), 2) basic preprocessing steps with fMRIPrep (motion correction, field unwarping, normalisation, bias field correction, and brain extraction; Esteban et al., 2019), and 3) quality control with MRIQC (Esteban et al., 2017). Besides ensuring a standard quality of our data and saving resources by automation of preprocessing steps, our openly available preprocessing pipeline can be used by other researchers to reproduce our results, and for their own neuroimaging data. As such, our pipeline exemplifies how open science can contribute to more robust and standardized research, which is especially relevant when translating complex neuroimaging analyses to clinical research.

References

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., . . . Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. Nature, 582 (7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PLOS ONE, 12(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. Nature Methods, 16 (1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., . . . Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific Data, 3 (1), 60044. https://doi.org/10.1038/sdata.2016.44

Halchenko, Y., Goncalves, M., Velasco, P., Di Oleggio Castello, M. V., Ghosh, S., Salo, T., Hanke, M., Wodder, J. T., Michael, Dae, Kent, J., Brett, M., Amlien, I., Gorgolewski, C., Lukas, D. C., Markiewicz, C., Tilley, S., Stadler, J., Kahn, A., . . . Meyer, K. (2022). Nipy/heudiconv : V0.11.6 (Version v0.11.6). Zenodo. https://doi.org/10.5281/ZENODO.7278515

Merton, R. K. (1973). The sociology of science: Theoretical and empirical investigations. University of Chicago press.

Shrout, P. E., Rodgers, J. L., et al. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. Annual review of psychology, 69 (1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Primary authors: GEYSEN, Steven (uni-bonn); Dr KOBELEVA, Xenia (Department of Neurology, University of Bonn, Bonn, Germany; German Center for Neurodegenerative Diseases (DZNE) Bonn, Bonn, Germany.); LEONE, Riccardo (uni-bonn)

Presenter: GEYSEN, Steven (uni-bonn)

Session Classification: Posters

Contribution ID: **30** Type: **Poster**

Data Science and Management in Virtual Product Development

Numerical simulations of car crashes are a key component of the virtual product research and development process in the automotive industry. In recent decades, virtual crash tests of vehicles on the computer, using commercial simulation software, supplemented the costly physical testing-only option. Nowadays, virtual crash tests outnumber their physical counterparts during the development of new cars by orders of magnitude. OEMs typically perform more than 10,000 car crash simulations per week.
We illustrate strategies to incorporate modern data analysis procedures into the virtual product development process of the automotive industry. Research on suitable machine learning approaches for this task takes place at the Institute for Numerical Simulaton and Fraunhofer SCAI.

Primary author: GARCKE, Jochen (uni-bonn)

Presenter: GARCKE, Jochen (uni-bonn)

Session Classification: Posters

Contribution ID: **31** Type: **Poster**

Streamlining of a metadata-based computational and statistical analysis pipeline

Efficient, extendible knowledge representation and reasoning enabling data mining and data visualization is becoming increasingly important in systems biology due to the constant growth, accumulation and availability of vast multi-variate and multi-modal data sets rendered possible through former and recent advantages in e.g., high-throughput experimental measuring techniques, i.e., omics and related technologies. To foster and to make the transfer from raw data to knowledge representation, thus giving the data meaning, ultimately through a contextual knowledge management, additional information describing the data, metadata is required. A prerequisite to allow for an efficient and extendible knowledge representation are a controlled vocabulary and ontology. By following established industry-standard modelling approaches, i.e., by defining a conceptual data model and in turn a logical data model, defining a physical data model resulting in data integration into a graph-based data base for knowledge representation we strive to streamline the computational workflow of metadata management and integration into GUI-based reactive, user-friendly custom statistical analysis tools (R Shiny). Making use of proper metadata representation to extract knowledge from the data sets, derived analysis and visualization tools can be developed in addition. To facilitate the collection of metadata we emphasize our decision for a low-entry barrier solution to input metadata and link data for users through an Excel spreadsheet (but not limited to) integrating readily into previously mentioned analysis tools and demonstrate use-cases of our web-based analysis pipeline for a repeatable reproducible omics analysis based on metadata.

Primary author: SEEP, Lea (Faculty of Mathematics and Natural Sciences, Rheinische Friedrich-Wilhelms-Universität Bonn)

Co-authors: Dr GREIN, Stephan (Faculty of Mathematics and Natural Sciences, Rheinische Friedrich-Wilhelms-Universität Bonn, 53115 Bonn, Germany); Prof. HASENAUER, Jan (Faculty of Mathematics and Natural Sciences, Rheinische Friedrich-Wilhelms-Universität Bonn, 53115 Bonn, Germany)

Presenter: SEEP, Lea (Faculty of Mathematics and Natural Sciences, Rheinische Friedrich-Wilhelms-Universität Bonn)

Session Classification: Posters

Contribution ID: **32** Type: **Poster**

Central Supercomputing Support for CRC 1502 DETECT (Project Z04)

The Collaborative Research Centre DETECT (CRC 1502) relies heavily on numerical modeling, data processing, and analysis. In this context a Central Supercomputing Support project has been established, which is responsible for: (i) Providing basic support for a multitude of HPC-related issues, (ii) helping with model development (iii) model simulation support (iv) training of early-career scientists at the interface of geoscience, HPC, and big data science. Thereby this project will help the CRC to efficiently conduct its demanding numerical model and simulation experiments as well as processing and analysis tasks, in a big data context. A few examples from ongoing work are presented.

Primary authors: Dr CAVIEDES-VOULLIEME, Daniel (FZJ); Dr GÖRGEN, Klaus (FZJ); MUTZEL, Petra (uni-bonn); Dr KELLER, Johannes (FZJ); Dr POLL, Stefan (FZJ)

Presenter: MUTZEL, Petra (uni-bonn)

Session Classification: Posters

Contribution ID: **33** Type: **Poster**

AHRD_Snakemake - Automatically Update Function Annotation on a Multi-Proteome-Scale

Genome-scale protein annotation can be performed by the transfer of functions from known proteins matched via sequence similarity. Errors can propagate when annotations falsely generated in this manner make their way into public databases and are used as basis for subsequent function transfers. Our program ''Automatic assignment of Human Readable Descriptions" (AHRD) can overcome these pitfalls by avoiding previously transferred annotations. It emulates the decision process of a human curator to select a description and GO terms from similarity search results. Through integration into a Snakemake workflow AHRD can now be easily applied to many proteomes at the same time. New annotations submitted to public databases can increase the annotation completeness and quality of stale proteomes released by past research projects. So the re-annotation quantity AHRD_Snakemake is capable of providing has a quality all of its own because it is able to keep a data base of genome-scale protein collections up to date with the current public knowledge.

Primary authors: BOECKER, Florian (uni-bonn); Prof. SCHOOF, Heiko (uni-bonn)

Presenter: BOECKER, Florian (uni-bonn)

Session Classification: Posters

Contribution ID: **34** Type: **Poster**

Data Management Plan as a strategy for an EU-funded project

Sustainable development projects often rely on a multitude of data sets from multiple sources given their trans-disciplinary profile. The EU-funded CLEVER (Creating Leverage to Enhance Biodiversity Outcome of Global Biomass) research project crosscuts several disciplines by aiming to quantifying biodiversity and other impacts of trade in major raw and processed non-food biomass value chains. The project involves 12 partners, which are divided in 8 working-packages, where they must work together, sharing data and results. Because of that, the funders requires a Data Management Plan (DMP), with clear guidelines for implementing the FAIR principles on both re-used and newly generated data via empirical field research. The objective of this poster is to provide an empirical example of DMP usage in a project with multiple partners. The DMP defines the handling of research data along the research data life cycle, documenting the context in which research data is collected, created and processed within the project. Some key topics addressed by the plan are utility, re-usage, storage, ethics, and access protocol to the data. In addition, data analyses and the publication of research data are planned ahead to ensure the interpretability and reproducibility of the research results generated during the project period and beyond. As examples of applicability of the DMP, the project commits to use Zenodo as repository, to publish open-access and to assign a globally unique persistent identifier (PI), i.e., a digital object identifier (DOI), for all metadata and generated and reused data. In this context, the long-term availability of the data and the archiving of the research data are also specified. Finally, all measures to ensure and maximize the reusability of the research data of the project are also defined. The EU requires an update on the DMP once an year, which means two times before the end of the project in three years.

Primary author: SILVA MARTINELLI, Fernanda (uni-bonn)

Presenter: SILVA MARTINELLI, Fernanda (uni-bonn)

Session Classification: Posters

Contribution ID: **35** Type: **Poster**

The data challenge in CRC 1502: Large-scale data handling and computing for modeling the atmospheric and terrestrial water and energy cycles

The Collaborative Research Centre 1502, DETECT, deals with the various anthropogenic changes affecting energy and water redistribution in the atmosphere and subsurface. For this, a considerable amount of data is being used. Experts from DETECT, including hydrologists, meteorologists, land use modelers, geodesists, remote sensing experts, agricultural economists, and social scientists, will use and generate diverse data in a variety of formats.
Service Project Z03 (DATA INFRASTRUCTURE AND SERVICES) is a sub-project of DETECT with the goal of providing and managing sustainable open research data infrastructures for DETECT. Z03 focuses on implementing the principles of FAIR (findability, accessibility, interoperability, and reusability) to facilitate efficient data flow paths, data processing, and modeling chains. Gridded forcing data, static fields, simulation results, restart files, model settings, regional climate and hydrologic model results, validation data based on in situ observations, satellite data products, and geographic baseline data are some of the data sets that will be managed in Z03 to be efficiently reused in an active research area. This issue requires a large-scale data and computing infrastructure and a collection of data integration and aggregation services, including HPC data workflows and management. These services will enable studies on data sets of various scales and heterogeneous quality. The technology will be complemented by a web-based visualization component to assist users in data exploration.

Primary authors: Prof. HAUNERT, Jan-Henrik (Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany); MOHSENI, Farzane (Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany)

Presenter: Prof. HAUNERT, Jan-Henrik (Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany)

Session Classification: Posters

Contribution ID: **36** Type: **Poster**

Research data management for DETECT: Storage and data processing in the HPC environment of JSC

Within the Collaborative Research Centre 1502, DETECT, large amounts of research data from various sources are being produced and shared between the CRC partners and to the outside world. These sources comprise model input and output data, observational data from satellites and large networks as well as economical and statistical information affecting land use and land cover developments. Reanalysis and ensemble simulation results from the regional climate model TerrSysMP will reach a data volume in the order of Petabytes, which calls for HPC oriented storage and workflow strategies to enable effective data analysis and sharing.
Jülich Supercomputing Center (JSC) at the Research Center Jülich hosts one of Europe's largest supercomputing systems, JUWELS, providing high-performance computing power as well as high-performance and high-capacity storage resources from its versatile storage infrastructure JUST. The Service Project Z03 of DETECT set up a data project at JSC, which can hold the data amounts to be produced within the CRC and which provides the necessary infrastructure for HPC related workflows and long-term archiving. Special care, already in the planning phase, needs to be taken of data formats, metadata, and general structure of the repository in order to allow for FAIR data handling and effective data processing. netCDF is favored as data format for geoscientific data, as it already comes with established metadata standards and elaborated tools for modification, analysis, and visualization. In addition, the use of data cubes for faster extraction of sub-datasets will be explored. Alternatively to classical HPC access with ssh, JSC now also offers datalad, a distributed data management system. Datalad opens up an easy way to access JUST data from your remote Computer, allowing to handle management of large datasets and to select user-defined files and datasets for up- and download.

Primary authors: Dr STEIN, Olaf (Forschungszentrum Jülich); NIKFAL, Amirhossein (Forschungszentrum Jülich)

Presenter: Dr STEIN, Olaf (Forschungszentrum Jülich)

Session Classification: Posters

Contribution ID: **37** Type: **Poster**

High Performance Computing at the University of Bonn

High Performance Computing (HPC) leverages the power of multiple compute nodes and architectures to solve complex problems, often with large data sets. Typical applications range from large scale simulations to machine learning and data analysis. The University of Bonn maintains several HPC clusters for specialized and general purposes, such as a massively parallel computing system with GPU accelerator partitions. This poster gives an overview of the HPC infrastructure and its future development, use cases as well as the HPC & Analytics Lab team, which is the main contact for scientific questions from HPC users of the University. We also briefly discuss the importance of research data management in the context of HPC.

Primary author: PETERSEN, Malte (uni-bonn)

Presenter: PETERSEN, Malte (uni-bonn)

Session Classification: Posters

Contribution ID: **38** Type: **Poster**

The Research Data Service Center

Digitization is changing day-to-day research practices across all fields of academic inquiry. Research data of every kind are collected, processed, analyzed, published and archived in digital systems. The term Research Data Management (RDM) refers to a range of activities and topics relating to the handling of digital research data, including technical, methodological, organizational and legal factors.
At the beginning of 2019, the University of Bonn formed the Research Data Service Center (ger. Servicestelle Forschungsdaten) in a move to provide the best possible support to researchers in addressing these challenges.
As an RDM competence center, we cover the entire research process from project planning to the final publication and archiving of research data. Our services include advice, conducting training and development of basic IT services for handling of research data.
Consultation Service: We offer advice to all researchers at the University - from doctoral candidates to working group leaders and from individual projects to collaborative research.
Need to write a project application and would like to clarify what the funder requirements are with regard to research data?
Are you looking for suitable backup routines or appropriate storage solutions tailored to your needs?
Want to publish your research data in a repository but need advice on what platform to choose and what metadata to assign?
No worries! We support you in the planning, application, implementation and completion phases of your project. simply send your request to our address: **researchdata@uni-bonn.de**
Training: We regularly offer training courses covering various open science and research data management aspects in German and English. On request, we are happy to hold workshops tailored to the needs of individual institutions such as institutes, collaborative projects or graduate schools.
If you have any questions or would like to receive more information about our training courses, please contact us at **researchdata@uni-bonn.de**
Data Management Plan (DMP) Service: We advise you individually on the creation of a DMP for your research project. You can also send us your draft version which we will revise and comment. We offer a Guide to the 'Handling of Research Data'in DFG project proposals (available in English and German) covering all required aspects.
Research Data Repositories: We administer the data repositories RADAR (current system) and bonndata (scheduled to go live in March 2023), that facilitate the professional publication and archiving of research data according to the FAIR principles. Such data typically derives from completed academic/scientific studies and projects. Publication of your research data in a repository like ours ensures the traceability, reproducibility and transparency of research results while heightening the visibility of research data through independent publication. Additionally, the published data can be used to address new research questions.
Electronic laboratory notebook (ELN): An ELN is a software designed to replace paper laboratory notebooks that are commonly used in the natural and life sciences to document and analyse research data. In comparison to standard laboratory notebooks, ELNs provide the user with, for example, enhanced search functionality, simplify copying the data and creating their backups, allow collaborative usage and enable access control.
The ELN RSpace is centrally hosted at the University Computer Center (HRZ). RSpace offers individually adjustable read and write permissions for a wide range of options for collaborative work. All entries are assigned to an individual user and are subject to complete versioning. RSpace complies with various standards and requirements (for example GLP, FDA Title 21 CFR Part 11). We also offer basic introduction courses.
Storage Services: Deciding on which data storage system and which data format to use will depend primarily on your project's requirements and the resources that you have available. We

generally advise you not to rely on local devices, external storage media or commercial cloud services (cf. Information on storage and backup strategies). Faculty and graduate students at the University of Bonn have access to a range of data storage services provided by University IT: the cloud storage service Sciebo, the high-performance network storage system FDI Research Data Infrastructure supplemented by the availability of virtual machines and an option for transferring very large files. Research staff at the Faculty of Medicine also have access to the storage services at the University Hospital Bonn.

Further Services that are soon to come: gitlab, Atlassian, Jupyter

Primary author: Dr BRES, Ewa (Servicestelle Forschungsdaten uni-bonn)

Presenter: Dr BRES, Ewa (Servicestelle Forschungsdaten uni-bonn)

Session Classification: Posters

Contribution ID: **40** Type: **Poster**

AI-guided long-term monitoring of epileptic mice

We present a pipeline for analyzing long-term video monitoring data from a mouse model of epilepsy. Mice were constantly monitored for 5 consecutive days, generating a large amount of high-resolution video data. Subsequently, we apply markerless pose estimation of user-defined body parts using deep neural networks (DNNs) to extract exact 2D motion data. To automatically identify behavioral episodes indicative of epileptic seizures, we utilize A-SOID (Tillmann, Hsu, et al. 2022), allowing data-efficient supervised behavior classification. Note that the workflow we present here is not limited to the analysis of epileptic seizures but holds the potential to study various behavioral phenotypes. Thus, we introduce an efficient, unbiased, and reliable method for detailed behavioral analysis in small rodents across long observation periods (days of constant monitoring). In general, automated data analysis techniques - like the one we present here - are particularly beneficial when dealing with large quantities of data, as they allow an unbiased, efficient and accurate analysis. In the future, we aim to develop and utilize automated workflows to streamline data analysis processes.

Primary authors: Dr TILLMANN, Jens (University Bonn, Medical Center); Mr MITLASOCZKI, Bence; Dr WENZEL, Michael (University Bonn, Medical Center); Dr SCHWARZ, Martin K. (University Bonn, Medical Center)

Presenter: Dr TILLMANN, Jens (University Bonn, Medical Center)

Session Classification: Posters

Contribution ID: **41** Type: **Poster**

Data management and long term storage strategies for Structural Biology

At the Institute of Structural Biology we aim to unravel structure-function relationships in biological macromolecules. For this purpose we use state of the art methods such as cryo-EM and X-ray crystallography to determine experimental structures of macromolecules. We also provide various in silico labelling tools (”http://www.mtsslsuite.isb.ukbonn.de/”) for scientists in the EPR- and FRET communities. Here we present our current data handling and storage strategies and point out possible bottlenecks for future developments.

Primary authors: HAGELÜKEN, Gregor (Institute of Structural Biology, University of Bonn); Prof. GEYER, Matthias (Institute of Structural Biology)

Presenter: HAGELÜKEN, Gregor (Institute of Structural Biology, University of Bonn)

Session Classification: Posters

Contribution ID: **42** Type: **Poster**

First LHCb open data release and analysis preservation techniques

LHCb and other experiments at CERN have the commitment to make the collected data available to the public. In December of 2022, LHCb has released the first data set containing 200 Terabytes of data from proton-proton collisions collected in the years 2011 and 2012 (Run1 of the LHC). This data is now available through CERN Open Data Portal. The aim of the open data release is to encourage particle physics research by third parties and aid in building the legacy of detector experiments at CERN. The report on this initial release is provided in the poster. It has become clear that it is important to preserve not only the data or the final results of physics analyses but also the steps taken to obtain these results. These steps include everything from data selection to statistical methods used in the analysis. It is also very important to preserve the computational environment which was used for the analysis, so, at a later date analysis could be rerun from start to finish for cross-checks or when bigger data samples become available. One framework to help with the preservation of the analysis techniques is REANA reproducible analysis platform which will also be briefly introduced in the poster.

Primary author: SARPIS, Mindaugas (Helmholtz-Institut für Strahlen- und Kernphysik)

Presenter: SARPIS, Mindaugas (Helmholtz-Institut für Strahlen- und Kernphysik)

Session Classification: Posters

Contribution ID: **43** Type: **Poster**

The bound states of the strong interaction - Baryon spectroscopy @ ELSA

The bound states of the strong interaction of hadrons is one of the worst understood areas of the standard-model of particle physics. Main questions are how the strong interaction produces its massive bound states from almost massless quarks and which (exotic) hadrons do exist. In order to investigate this, experiments are carried-out at ELSA and other accelerators in the world.

This poster describes the CBELSA/TAPS experiment at ELSA which is designed to do light baryon spectroscopy using real high energetic photons. The current and near future experiment tasks are to study photoproduction of the neutron and multi-meson photoproduction. An estimate of the amount of data acquired is given in order to evaluate the current demands for data storage and computing. Plans for future upgrade of the experiment allow studies of strange baryons. This requires higher performance in computing and data storage. Estimated demands are shown.

Primary authors: SCHMIDT, Christoph (Helmholtz-Institut für Strahlen- und Kernphysik); Dr LANG, Michael (HISKP); HARTMANN, Jan (Helmholtz-Institut für Strahlen- und Kernphysik)

Presenter: Dr LANG, Michael (HISKP)

Session Classification: Posters

Contribution ID: **44** Type: **Poster**

RDM & Data Science in (Systematic) Theology

Theology has a long history of working with (primarily text-based) artifacts. Until recently, digitization was used for digitizing those artifacts rather than considering the "new" digital world itself as holding significance for theological research. Consequently, questions related to research data management (RDM) infrastructure and the application of data scientific methodologies are still in their infancy.
The rise in mass data analysis in particular presents theological research with new opportunities. However, due to the religious nature of the communication analyzed, the resulting datasets can hold uniquely sensitive data. The most concerning RDM issue is thus the discussion around data security. Approaches such as anonymization and aggregation hold great potential while still presenting special challenges for reusability and transferability of results.
The main challenges theology faces in the application of data science approaches are theoretical and practical in nature. On the theoretical side, epistemological questions about theological inquiry interfere with the need to operationalize concepts. On the practical side, a lack of technical training and expertise in data scientific research prevents those skills from being part of current curricula. The interdependent nature of these issues makes them challenging to address. Thankfully, small clusters for Digital Theology are developing, tackling the main issues necessary for this topic to receive broader attention. However, better access to no-code and low-code applications for data science is needed to support the process beyond those clusters and help theology's advancement as a modern science.

Primary author: FRÖH, Johannes (uni-bonn)

Presenter: FRÖH, Johannes (uni-bonn)

Session Classification: Posters

Contribution ID: **45** Type: **Poster**

Data Science and Data Management in Geophysics and Geosciences

In Geophysics, physical methods are applied to the earth's subsurface and related objects of investigation on both laboratory and field scale.
The data resulting from measurements and synthetic modeling are acquired mostly digitally or are at least managed in digital data structures in order to be analysed both quantitatively and qualitatively.
Therefore, the digital infrastructure for data science and data management in Geophysics and Geosciences needs to methodologically support the storage, handling and processing of multiple large data sets.

From the perspective of Data Science and Data Management in geophysical and geoscientific research, teaching and administrative work, there are multiple points that need to be considered: From the user's perspective as a data scientist or manager, the speed and straightforwardness of the access are as important as the security. Data and software should be available and accessible independent of platform and user location. In addition, data structure and metadata are obligatory factors of a working data science and management workflow. While it has always been important from the technical and administrative standpoint, the question of resources becomes also interesting for the end user.

The importance of the challenges arising from these points combined with the infrastructure of hard- and software that are needed in geophysical and geoscientific work are showcased. Besides few commercial (geo)physics specialised data analysis software mostly python-based self-developed and subject-specific software are used.
Based on these needs and challenges in geophysical and related geoscientific work, a few approaches that are pursued regarding metadata management and centralisation of hard- and software services will be presented, for example the *Helmholtz Metadata Collaboration* and the *FAIR Guiding Principles* and a few other strategies and projects, that will be partwise offered by the Hochschulrechenzentrum of the University in the future.

Primary author: Mr HEIDEMANN, Niklas (uni-bonn, Geophysics Section)

Presenter: Mr HEIDEMANN, Niklas (uni-bonn, Geophysics Section)

Session Classification: Posters

Contribution ID: **46** Type: **Poster**

Challenges of Capturing Temporal Volumetric Video Data for Relightable Object Reconstruction

Since its inception, computer graphics research is focused on modeling and creating computer-generated content that captures the appearance of the real world. For several decades, a lot of work went into reconstructing objects and scenes through photographs and other capturing methods. This does not only include geometry acquisition but also material and appearance reconstruction. In more recent years, the rapid developments in hardware and improvements of computing power led to the research of so-called light stages. Consisting of dozens of cameras and light sources, these dome setups can be used to capture a new form of research data called volumetric videos. This type of data allows for temporal consistent and relightable reconstructions of, for example, humans in arbitrary virtual environments. However, due to the necessity of high spatial and temporal resolution, even short videos can result in huge amounts of data. In this poster, we summarize some of the recent approaches in reflectance data capturing and outline problems of volumetric videos in the context of research data management.

Primary authors: ZINGSHEIM, Domenic (Institute for Computer Science II (Uni Bonn)); Prof. KLEIN, Reinhard (Institute for Computer Science II (Uni Bonn))

Presenter: ZINGSHEIM, Domenic (Institute for Computer Science II (Uni Bonn))

Session Classification: Posters

Contribution ID: **47** Type: **Poster**

Dealing with New Security Threats: Global Threats Index (GTI) and Chat GTP

Introduction and Punch line:
•The GTI based on Chat GPT + Excel aims to show the position of a country and eight global (sub)geographic regions in the context of an emerging constellation of transnational threats.
• Research question: To what extent have global threats pervaded the multiple facets of human security and how AI can be helpful to provide prompt and timely information?
• Objective: to show improvements/deterioration in planetary sustainability based on the use of AI
•Methodology: The GTI is assigned a score ('banded') on a scale of 1 to 5 and overall scores are produced for each country or territory. A score closer to 1 records that a country is less prone to facing global threats (or has more capacity to face these threats). A score closer to 5 shows the opposite.
Approach: The GTI is multidimensional and relative measure that aims to show the exposure to planetary instability by producing one simple and easy to interpret number.

Primary author: MADRUEÑO, Rogelio (uni-bonn)

Presenter: MADRUEÑO, Rogelio (uni-bonn)

Session Classification: Posters

Contribution ID: **48** Type: **Poster**

A lack of tools is not the problem in maize hybrid RNA-seq

Cross-pollinated F1-hybrids are more vigorous than their parents, produce more biomass, have a faster development and greater fertility. This phenotypic variation between the parental mean and the hybrid is often accompanied by transcriptomic variance. The transcriptome is an important link between the genome and the phenotype of organisms. Sequencing the messenger RNA (mRNA) captures the coding transcriptome. In our study, RNA-seq is used to analyse hybrid vigour. The raw data are made publicly available. But different research questions often require a new experimental setup. If and how this data could be reused remains an open question. Further, there are a variety of tools for processing RNA-seq data. Most are computationally intensive, run on a server, and require scripting knowledge. Technical implementation details, complete parameters and analysis scripts are stored locally or on Sciebo. Improving the sharing of this technical knowledge remains one of the challenges.

Primary author: PITZ, Marion (uni-bonn)

Co-author: Prof. HOCHHOLDINGER, Frank (uni-bonn)

Presenter: PITZ, Marion (uni-bonn)

Session Classification: Posters

Contribution ID: **49** Type: **Poster**

Reproducible Clinical Research Necessitates Metadata Harmonization and Standardization

Metadata is structured information that describes data objects. It gives the data user the necessary context to extract information from the data. In research, metadata is essential for data quality checks, interpreting findings, and reproducing experiments. Currently, each institute or group has its own catalog of metadata they require for experiments and subsequent analyses. This makes data integration from different research groups almost impossible and hinders the replication of experiments. With the publication of the FAIR (findable, accessible, interoperable, and reusable) data principles, FAIR data as well as data and metadata standards receive increasing recognition. However, many researchers are unaware of the possibilities for structuring their metadata and how to enrich metadata from data providers or other third parties. In addition, clinical data pertaining to human subjects comes with its own ethico-legal challenges regarding privacy and security. Therefore, we want to highlight community standards for minimal clinical metadata harmonization and standardization that are applicable across a wide range of biomedical research disciplines.

Applying our experience from developing the German Human Genome-Phenome Archive metadata model, we collected community standards that serve to enhance data and metadata quality. These cover minimal reporting standards, various ontologies, and other best practice guidelines from clinical research and sequencing applications. As many scientists do not have to deal with legal challenges arising from human-related data, we additionally want to shed light on possible issues and offer workarounds that are GDPR-compliant and still enable fair data collection and sharing.

Standardization and harmonization of data are key steps during all steps of data collection and processing. Educating researchers about data and metadata standards in clinical research fields counteracts the impending reusability crisis and increases overall data quality.

Primary authors: SCHULTZE, Joachim L. (DZNE, Universität Bonn); MAUER, Karoline (DZNE); ULAS, Thomas (Universität Bonn, DZNE)

Presenter: MAUER, Karoline (DZNE)

Session Classification: Posters

Contribution ID: **50** Type: **Poster**

VAMPIRA provenance generation

We have developed VAMPIRA, software capable of automatically generating provenance for data-intensive scientific workflows. Provenance generated by VAMPIRA describes the record of the data processing, metadata, infrastructure and user data involved within a workflow as well as the interactions between them. Armed with this extra information, scientists will be able to make more informed decisions on the trustworthiness of data products, pipelines, or pipeline components. Therefore, VAMPIRA can help to solve the so-called "black box problem"which is prevalent in modern artificial intelligence (AI) research due to the increasing intricacy and complexity of AI workflows.

Primary authors: JOHNSON, Michael (Max Planck Institute for Radio Astronomy); Dr LACKEOS, Kristen (Max Planck Institute for Radio Astronomy)

Co-authors: Dr KLOECKNER, Hans-Rainer (Max-Planck-Institut für Radioastronomie); Dr CHAMPION, David (Max-Planck-Institut für Radioastronomie); Dr SCHINDLER, Sirko (DLR-Institut für Datenwissenschaften); Dr DEMBSKA, Marta (DLR-Institut für Datenwissenschaften); Dr PARADIES, Marcus (DLR-Institut für Datenwissenschaften)

Presenters: JOHNSON, Michael (Max Planck Institute for Radio Astronomy); Dr LACKEOS, Kristen (Max Planck Institute for Radio Astronomy)

Session Classification: Posters

Contribution ID: **51** Type: **Poster**

Research Data in Lattice Quantum Field Theory

Research in Lattice Quantum Field Theory (LQFT) is performed by international collaborations in both overlapping and disjoint research projects studying a wide range of non-perturbative physical problems. The generated data, which can roughly be classified into three tiers, span the whole spectrum of storage, metadata and lifetime requirements.

LQFT simulations are some of the most expensive in computational science and there are efforts underway to make some of the research data accessible in a FAIR way. The challenges to be overcome to get there are substantial.

We present some of these challenges and examples of putative solutions to provide a basis for a discussion of the transformations required at the level of policy and infrastructure in order to encourage FAIR principles to be adopted as widely as possible in the community.

Primary authors: KOSTRZEWA, Bartosz (Univ. of Bonn, High Performance Computing & Analytics Lab); URBACH, Carsten (Helmholtz-Institut für Strahlen- und Kernphysik)

Presenter: KOSTRZEWA, Bartosz (Univ. of Bonn, High Performance Computing & Analytics Lab)

Session Classification: Posters

Contribution ID: **52** Type: **Poster**

Parallel, Distributed, Adaptive Simulation Data Management

A key to defining the organizational structure of simulation data is the computational mesh. It encodes where in space, and possibly in time, simulation data points are located. This is essential for the inner functioning of the simulation on the one hand and for in-situ processing, storing, and post-processing the data on the other. The primary danger is losing the parallel, highly optimized access to the data when moving it from simulation memory to intermediate exchange formats and/or mass storage.

We illustrate perspectives that reach beyond pure simulation at the example of p4est. The p4est software library implements fast algorithms for large-scale distributed adaptive mesh refinement and data location and serves as the parallel mesh backbone for various third-party simulation codes. It employs a distributed forest-of-octrees data structure and has been demonstrated to scale to 3e6 MPI processes and .5e12 mesh elements. p4est lends itself as a generic tool for partition-independent management of spatial data. This means that the specific division pattern of data among the many parallel processing units shall influence neither the results of the computation nor the definition of any data exchange format. This capacity is a precondition for flexible, reproducible re-processing.

On this poster we present several aspects. First of all, we illustrate the concept and practical role of parallel partitioning in general. Second, we present a partition-independent I/O mechanism for simulation data. It employs the MPI I/O standard and at the same time guarantees write- and read-equivalent files over an arbitrary partition and even without MPI support. The stored data continues to be amenable to our highly scalable algorithms used for spatial search and simulation. Third, we provide an example of a p4est-based partition-independent simulation. The octree-based design enables efficient remote search functionalities, for example to locate physical measurement points (such as floating tide gauges) and to integrate over rays or curves (such as lines of sight in satellite imaging).

Primary authors: Mr GRIESBACH, Tim (University of Bonn Institute for Numerical Simulation); Mr BRANDT, Hannes (University of Bonn Institute for Numerical Simulation); Prof. BURSTEDDE, Carsten (University of Bonn Institute for Numerical Simulation)

Presenter: Mr GRIESBACH, Tim (University of Bonn Institute for Numerical Simulation)

Session Classification: Posters

Contribution ID: **53** Type: **Poster**

Data exchange in the transregional research network TRR237

The SFB/Transregio 237 explores the mechanisms and functional consequences of Nucleic Acid Immunity. On the one hand, we focus on gaining fundamental insights into the specific molecular mechanisms that control the defense against pathogen-derived foreign nucleic acids. On the other hand, we address the functional role of this system in health and disease at the systemic level of the whole organism.
In our efforts, a variety of types of data will be created by imaging, FACS analysis, ELISA, arrays, genomics, proteomics, data of human patients, mice, cell lines, bacteria, and many more. An additional challenge will be the integration of these data in a multi-location research consortium.

The TRR237 will address these challenges from multiple angles in the central infrastructure project (INF):
- A team of bioinformaticians, researchers, and administrators will provide and monitor the structure of the INF project
- Data will be stored and archived locally (University's Infrastructure)
- Data will be exchanged via a TRR237 Server running Nextcloud (TRR237 Infrastructure)
- A Bioinformatician and a Data Steward provide administration of the TRR237 Server.
- Scientists are empowered in good scientific practice, the FAIR principles, data analysis, and management (Data Steward, Workshop, Data Policy)

By these and further measures, we aim to improve the quality of our data and cooperation for better research in our Transregio and research in general.

Primary author: GÖRGEN, Simon (uni-bonn)

Presenter: GÖRGEN, Simon (uni-bonn)

Session Classification: Posters

Contribution ID: **54** Type: **Poster**

Data Management for Collaborative Research Projects: Common Challenges and Solutions

Background
Scientific research often involves collection, storage, analysis and reporting large amounts of multi-disciplinary and multi-site data. Data should be findable, accessible, interoperable, and reusable. We present common obstacles and propose solutions.
Challenges
- Infrastructure for real-time collection and sharing across networks
- Maintaining privacy of open access data among primary and secondary users
- Need for training in inter-disciplinary competences e.g., Spatial data management
Solutions
- Develop and keep updated project data management plan
- Consent should disclose, short- and long-term and potential secondary users of data
- Avail inter-disciplinary trainings in spatial data management
Reference
- Wilkinson, M. et al., 2016, Sci Data 3, https://doi.org/10.1038/sdata.2016.18

Primary author: Ms NABATANZI, Maureen (uni-bonn)

Co-author: Prof. BIBER-FREUDENBERGER, Lisa (University of Bonn)

Presenter: Ms NABATANZI, Maureen (uni-bonn)

Session Classification: Posters

Contribution ID: **55** Type: **Poster**

Data Management and Data Science in Time-Resolved Fluorescence Microscopy of Individual Bacterial Cells

Modern advancements in the field of fluorescence microscopy generate large amounts of data in the form of image files. However, it becomes more problematic when it comes to modify, evaluate and analyze image files of different types. And if this was not enough, corresponding raw data needs to be generated not only in two dimensions, but sometimes in three dimensions. Additionally, acquisition in multiple channels along with time-resolved experiments just adds to the complexity of these data sets.
The project presented here deals exactly with this problem. It entails acquiring and analyzing time-resolved experiments of fluorescently labeled bacteria treated with antibiotics. The generated data is deconvoluted, images reduced from 3 dimensions to two, aligned, segmented and measured. All these steps generate new forms of data, which have to be stored and eventually evaluated.
Open questions that need addressing:
-At which points of the workflow does data have to be stored?
-Can we reduce the amount of data that needs to be stored?; and
- Further possibilities for data mining and if data evaluation can generate further information.

Primary authors: BRAJTENBACH, Dominik (uni-bonn); Prof. KUBITSCHECK, Ulrich (Clausius-Institute)

Presenter: BRAJTENBACH, Dominik (uni-bonn)

Session Classification: Posters

Contribution ID: **56** Type: **Poster**

Rhineland Study Data Management

The Rhineland Study is an ongoing prospective population-based cohort study in Bonn. Its two study centers are located in Beuel and Duisdorf, where all residents ≥30 years are invited to participate. The participants undergo a broad range of examinations including, e.g., assessment of cardiovascular measures, brain imaging, cognitive testing and neurologic functioning with the aim to the find the key to a healthy life into old age. The data management team of the Rhineland Study is involved in all steps of information processing ranging from the data collection in the centers, the sustainment of data quality and security to the generation of research variables. This poster presents our data flow and software solutions for the Rhineland Study including current challenges and development directions.

Primary author: KLITZ, Margrit (DZNE)

Presenter: KLITZ, Margrit (DZNE)

Session Classification: Posters

Contribution ID: **57** Type: **Poster**

Data management for the Vocal Control and Vocal Well-Being Lab

School teachers are disproportionately more likely to experience voice problems compared to the general population. One understudied component of this elevated risk is the role of stress on motor control for voice and speech production. The goal of this study is to determine the functional and clinical relevance of stress on the motor control for voice and speech production in the brain and will determine neural (functional MRI), psychobiological (cortisol, personality), and vocal function (surface electromyography, acoustic) signatures of stress responders and nonresponders in early career teachers with vocal fatigue and control participants. The research plan includes a variety of modalities spanning multiple acquisition methods and time spans for a hundred participants. The primary challenges of the data management for this project are the organization, storage, and sharing and analyzing of the large quantity of multimodal data. Furthermore, working within a clinical environment that requires strict participant privacy and its own data security presents additional difficulties such as needing different storage systems than the acquisition systems. Mitigation of these challenges includes the successful implementation of a research data management plan which is constructed from best practices. Important aspects of this plan include personnel hierarchy for accountability and management of the necessary data sharing and archival systems and proper documentation for the storage location, naming conventions, and type of the variety of data. Attending workshops and seeking advice from colleagues and experts such as the data privacy officer also facilitate data management. Successful implementation of a research data management plan for this study will facilitate efficient data analysis and continuity when there is turnover in lab membership.

Primary authors: VON DER HEYDE, Juliane (uni-bonn); BERARDI, Mark (uni-bonn); DIETRICH, Maria (uni-bonn)

Presenters: VON DER HEYDE, Juliane (uni-bonn); BERARDI, Mark (uni-bonn)

Session Classification: Posters