

CRC-TR 211 Data Policy

(Dated: December 17, 2019)

This document outlines the policy of the CRC-TR 211 with regards to research-data management and data sharing of all scientific projects.

I. INTRODUCTION

There are many benefits to managing and sharing your research data. Sustainable research data need to fulfill the FAIR principles, i.e., the data need to be Findable, Accessible, Interoperable, and Re-usable. According to the DFG guidelines [1], research data need to be preserved for at least ten years. This document serves to specify this requirement for our purposes considering the feasibility for the different applications. A minimal set of data and metadata that shall be stored will be defined below. In general, "data" refers here to any result from research that cannot easily be reproduced, that includes software, notes on analytical calculations, scripts as well as actual raw and condensed data displayed in plots in publications. The PIs are responsible for ensuring the implementation of the data policy. The Z02 project provides help with supporting infrastructure (e.g. a Wiki page [2]). Awareness and sensibility to good scientific practice and proper data management are a responsibility of all members of the CRC and will be ensured by regular information through the executive board.

II. GOALS

With this data policy we want to initiate a research data base of all CRC-TR 211 publications to demonstrate good scientific practice and ensure long-term storage of our data. In particular, we want to provide a storage area from where data can be retrieved independently from the actual researchers that worked on the publication, as they might have left the field or the collaboration. We want to maintain reproducibility by documenting the complete workflow that lead to the published data. In addition resources are provided to share software and maintain it independently of specific members of working groups.

III. IMPLEMENTATION

Research data are distributed by publications (journals, proceedings, etc.) and in undergraduate and graduate theses. In the high-energy, nuclear, and astrophysics communities that cover all topics in our CRC the arXiv [3] is used for all publications. The platform is provided by Cornell University and supported by many German and international research institutions, universities, and societies. Therefore, this will be used as the central repository to avoid double effort. The arXiv is publicly available and also ensures that the international community gets access to the data generated in this CRC. For all additional purposes version-controlled repositories will be used. If groups do not have their own infrastructure for this purpose, git repositories are provided through Redmine [4] at Bielefeld University. Members of the CRC who acknowledge CRC funding are obliged to submit the Tex source of publications to the arXiv repository in the form described below. If the ancillary data files are too large or already uploaded to another public repository (e.g. at the corresponding

university or a research laboratory, HEPdata or similar) a link should be provided in the remarks of the arXiv submission. The common structure for the source directories is defined as follows, where some of the sub-directories are mandatory, others are optional:

- **source** (mandatory) This directory contains the original TeX source of the publication. Inside the **source** directory that will be uploaded to arXiv¹ there will be two sub-directories:
 - **figures** (mandatory) Separate pdf- or eps-files for each figure in the publication.
 - **anc** (mandatory) Separate text files with the data to reproduce the figures. The names of the .dat-files match the names of the corresponding figure files. *If data files are too large or are already publicly available the **anc** directory is not required.*
- **workflow** This directory contains information on the workflow, i.e., information on the used software (version, libraries, packages, etc.), as well as the analytical methods, possibly packages and scripts for computer algebra systems. Additional notes, i.e., handwritten calculations, as pdf, TeX documents or scans, can also be added. Information from this directory is meant to allow reproduction of the results. If wanted part of the **workflow** like useful scripts (in particular the plotting scripts to obtain the figures from the submitted data) or code parts can be added to the arXiv submission in the **anc** directory.

Responsible for the submission to arXiv are the authors of the publication. For Master/PhD theses completed by CRC members arXiv upload is not required; however in the case the author chooses not to upload their thesis to arXiv, they must upload to the CRC Theses and Dissertations Redmine. On request a sub-project containing the git repository will be created in the 'CRC Theses and Dissertations' (or 'CRC publications') area. While all members of the CRC have read access, the write access is restricted as required. The sub-project name shall contain the arXiv number and keywords that guide to the content². For theses and dissertations the directory name shall contain the thesis type and author name.

-
- [1] [DFG Richtlinien Forschungsdaten](#)
 - [2] [Data Management redmine Wiki](#)
 - [3] [arXiv repository](#)
 - [4] [Redmine of the CRC](#)
 - [5] [Ancillary file handling on arXiv](#)

¹ Authors are encouraged to mention the addition of ancillary data files in the free text field. Please note that arXiv supports ancillary files [5].

² The directory may be created even before the publication is submitted to arXiv with a working title. After the arXiv number is known it can be moved using the git mv command.