CRC-TR 211 Data Policy

(Dated: July 05, 2023)

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

1 Introduction

There are many benefits to managing and sharing your research data. Sustainable research data need to fulfil the FAIR principles, i.e., the data need to be **F**indable, **A**ccessible, **I**nteroperable, and **R**e-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 project leaders 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-TR 211 and will be ensured by regular information through the executive board.

2 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.

3 Implementation

Research data are distributed by publications (journals, proceedings, etc.), in undergraduate and in 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 free distribution service 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.

3.1 Enforced common structure for upload to the arXiv

Members of the CRC-TR 211 who acknowledge CRC funding are obliged to submit the LATEX source of publications to the arXiv platform in the form described below. You are encouraged to use a public service (e.g. at the corresponding university [6, 7, 8] or a research laboratory, HEPData or similar) to publish your data. In such case, a link to the data publication must be provided in the remarks of the arXiv submission. If this is not possible, as last resort, you should use the anc illary data folder on arXiv to publish your data. 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 LATEX source of the publication. Inside the source directory that will be uploaded to arXiv[1] there will be two sub-directories:

📂 figures ← MANDATORY

Separate PDF or EPS files for each figure in the publication.

📂 anc ← MANDATORY – *if data are not published using a different service*

Separate text files with the data to reproduce the figures. The names of files must match the names of the corresponding figure files.

- It is possible to include data points belonging to several figures in the same file, but it should be unambiguously clear which data points belong to which figure and where data points for each figure are and file names should still refer to figure files. For example, if data of `Figure-1.pdf` and of `Figure-2.pdf` are in the same file, this should be e.g. called `Figures-1-2.dat`.

- If desired, the text file with data points can be replaced by a script to recreate the figure file. The script name must match the name of the figure(s) it produces and authors must make sure to provide information about how the script should be run (cf. the `workflow` folder).

- *If data files are too large or are already publicly available the **anc** directory is not required.*

📂 workflow ← OPTIONAL

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, LATEX 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 in the above described common format are the authors of the publication.

---

[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].

3.2 Types of publications

All CRC publications must have the format already described. Depending on the type of publication, minor tweaks in behaviour are accepted and these are described in the following table.

Type of publication	Remarks
Regular article	If authors needs a digital space to collaborate, the Redmine shall be used. On request a sub-project containing a `Git` repository will be created in the "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 content. In the preparation phase (i.e. before the publication is submitted to arXiv) a working title for the sub-project may be used and, after the arXiv number is known, it can be renamed by the authors. *If the article does not contain any figure, the `figures` and `anc` sub-folders can be omitted.*
Review articles	Previously published figures that are used in review articles do not need to have any ancillary text file with data point associated. Therefore, if no original figure of the authors is included, the `anc` sub-folder can be omitted.
Conference proceeding	As conference proceedings often contain plots from preliminary data, a real data publication e.g. at the university library is discouraged and the arXiv ancillary file system should be preferred. However, for this type of publication a data upload to arXiv is not required. Still, if the authors decide not to upload the data points corresponding to the figures to arXiv in the `anc` sub-folder, the publication should be uploaded to the Redmine platform in a dedicated sub-project, as possible for regular articles and described above.
Master/PhD thesis	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 Redmine platform. On request a sub-project containing the `Git` repository will be created in the "CRC Theses and Dissertations" area. While all members of the CRC have read access, the write access is restricted as required. The sub-project name shall contain the thesis type, year and author name.

References

[1] DFG Richtlinien Forschungsdaten
[2] Data Management redmine Wiki
[3] arXiv repository
[4] Redmine of the CRC
[5] Ancillary file handling on arXiv
[6] Data publication at Bielefeld library
[7] Data publication at Darmstadt library
[8] Data publication at Frankfurt library