



CODATA-RDA Schools for Research Data Science Vision, Information and Progress, April 2021

Objectives and Needs

Contemporary research – particularly when addressing the most significant, transdisciplinary research challenges – cannot effectively be done without a range of skills relating to data. This includes the principles and practice of Open Science and research data management and curation, the use of a range of data platforms and infrastructures, large scale analysis, statistics, visualisation and modelling techniques, software development and data annotation. We define ‘Research Data Science’ as the ensemble of these skills.

Research Data Science skills are common to all disciplines and training in ‘Research Data Science’ needs to take this into account. For example, all disciplines need to ensure that research is reproducible and that provenance is documented reliably and this requires a transformation in practice and the promotion of the necessary culture, practice and skills.

A strategic priority shared by [CODATA](#) and the [Research Data Alliance](#) is to build capacity and to develop skills, training young researchers in the principles of Research Data Science. It is also important that Open Data and Open Science benefit research in Low and Middle Income Countries (LMICs) and do not result in even greater inequalities in research and scientific output. On the contrary, it has been argued that the ‘Data Revolution’ provides a notable opportunity for reducing research inequality in a number of respects. For this reason, particular attention is paid to the needs of young researchers in LMICs.

Mission of the CODATA-RDA Schools for Research Data Science

The CODATA-RDA Schools for Research Data Science:

- address a recognised need for Research Data Science skills across disciplines;
- use and adapt existing materials to create an accredited curriculum that is more than the sum of its parts;
- provide a pathway from a broad introductory course for all researchers;
- are reproducible: all materials will be online with Open licences;
- are scalable: emphasis will be placed on Training New Teachers (TNT) and building sustainable partnerships;
- pay particular attention to the needs of early career researchers in LMICs;
- build a network of researchers who can collaborate and continue teaching this material.

Visit our [website for more information](#).

Foundational School for Research Data Science

The foundational school provides a bedrock of introductory material, common to all research disciplines, and upon which more advanced schools can build. The schools are primarily targeted at post graduate students, but early career researchers are also admissible. The schools provide skills that are of use both to researchers and to other roles involved in the research enterprise.

The foundational school is designed to run for two weeks and covers the following topics; Programming, Information Security, Version Control, Authoring Papers, Research Data Management, Principles of Open Science, Computational Infrastructures, Analysis and Visualisation.

Physical schools

The first full introductory school ran from 1-12 August 2016 at the [Abdus Salam International Centre for Theoretical Physics \(ICTP\)](#) in Trieste, Italy. It was a huge success, delivering the school to 67 students from over 30 countries. Since then 10 additional physical schools have been run in Ethiopia, Costa Rica, Rwanda, South Africa, Brazil and Australia. Additional training events for Data Stewards, in collaboration with the EU network [FAIRsFAIR](#) have been run in Trieste, Manchester and San Jose. A selection of photos from the first school can be found [here](#). A short video about the 2018 Trieste school can be found [here](#). An interactive map of CODATA-RDA and CODATA alumni can be found [here](#).

Approximately 450 students have been taught on five continents. This corresponds to roughly \$425K worth of teaching and what is remarkable is that all this has so far been achieved using an entirely volunteer network drawn from over 20 countries.

Virtual schools

The present situation with respect to COVID-19 has not stopped the school from continuing its delivery of teaching. In August 2020 the annual Trieste school will be replaced by an online event with a focus on providing a refresher course and train-the-trainer for the Alumni of the school. **From May to July this year the [first entirely online version of the school](#) will be hosted by the University of Pretoria in South Africa. From September to November the second version will be hosted by the ICTP in Trieste.**

Our Curriculum

In December 2019 our curriculum, used to teach all our students on five continents, was endorsed by the Research Data Alliance (RDA) and officially recognised as an RDA Recommendation – *A Curriculum for Foundational Research Data Science Skills for Early Career Researchers*. The RDA is a grass-roots organisation of over 10,000 data professionals.

Our Alumni

Enabling the alumni of the schools so that they can collaborate and deliver further teaching is a key goal of the schools. Three of our co-chairs are alumni. The train-the-trainer event last year kick started a more formal recognition of the alumni and there is now a regular series of webinars being run for them.

Your support

The delivery of online schools is potentially much more cost-effective but nonetheless there are still barriers for students to attend. **For \$500 you can support the online access of 25 students.** These comparatively small sums of money can be transformative in terms of improving accessibility.

Further Information

Details about the schools can be found our [web site](#). You can contact the schools [here](#).