CODATA-RDA Research Data Science Summer Schools

Vision, Information and Progress, June 2019

Objectives and Need

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 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 Research Data Science Summer Schools

The CODATA-RDA Research Data Science Summer Schools:

- 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;
- be reproducible: all materials will be online with Open licences;
- be 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.

Foundational School in 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 to researchers and to other role involved in the research enterprise.

The introductory 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.

Progress and Resourcing for the first 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. The second school, also based at the ICTP in Trieste, ran in July 2017 was a similar success, running an additional advanced school in the week following it.

A selection of photos from the first school can be found at **http://tinyurl.com/yddlbm7o .** A short video about the 2018 Trieste school can be found at **https://www.youtube.com/watch?v=AK8Y_UigQcc&feature=youtu.be .**

In December 2017, the first regionally run example of school was held in São Paulo, Brazil. Delivered to roughly 30 students from South and Central America, this represented an important step as roughly half of the material was delivered by instructors from the region, demonstrating a mechanism for how the schools will grow in the future. Schools ran in 2018 in Trieste, São Paulo and Kigali, Rwanda - the first African school.

Scaling up

In 2019 we will run three schools (Addis Ababa in Ethiopia, Trieste and San José in Costa Rica). At the end of this year **approximately 400 students will be taught on five continents.** This corresponds to roughly $600K worth of teaching. **What is remarkable is that all this so far has been achieved using an entirely volunteer network.**

Reproducing the School: Fellows as helpers and training new tutors

We are putting in place mechanisms to ensure that the schools are as pedagogically effective as possible and that they can be easily reproduced elsewhere. To achieve this, we invite students who attended previous schools to take part as helpers in later schools. Through this approach, such students can become more involved and ultimately become involved in running instances of schools in their region.

Sustainable Coordination

To efficiently run continual schools of Research Data Science some level of coordination will need to be maintained. This includes minimal staff tasked with securing funding, organizing curriculum and materials, online content and learning technologies, administrative operations and overall coordination. This dedicated staff coordination will be key to sustaining and increasing the impact and capacity for additional schools outside of the foundational ICTP Schools.

A Task Group

Given the success of the schools CODATA has approved support for a new Task Group to guide the development of the schools. Its members are taken from a wide variety of institutions and countries. The Task Group leads on developing applications for funding, determine the ideal sustainability model for the schools, oversee the development and maintenance of the curriculum.

Further Information

Details about the schools can be found at http://indico.ictp.it/event/7658/, http://www.ictp-saifr.org/?page_id=15270 and http://www.codata.org/working-groups/research-data-science-summer-schools .