Embedding Open Science practice into Data Science Training.

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 annotation.

A Task Group has been set up to address the critical need for skills around data science, research data management, and open science practice by pulling together existing and emerging initiatives and resources into a sustainably run, globally applicable foundational programme in Research Data Science, particularly for Low and Middle Income Countries (LMICs).

The approach for this is to run a series of Schools which typically run for two weeks that provide a foundation in the wide range of topics described above to Early Career Researchers (ECR's), with a focus on supporting those from LMIC's.
Since August 2016 these schools have run in a variety of locations, namely Trieste, Italy; Sao Paulo, Brazil; Brisbane, Australia and Kigali, Rwanda. We have delivered the curriculum to approximately 250 students from over 30 countries in a wide variety of disciplines (e.g. Bioinformatics, Earth, Atmospheric and Climate Sciences, Economics and Physics).

The school covers a wide variety of technical topics such as Programming, Machine Learning and Cloud Computing but also spends a significant amount of time on more social areas such as Research Data Management and Authorship in the 21st century.

Open Science in particular is delivered in a way that is both implicit and explicit. Implicit in terms of the teaching of topics that enable Open Science, such as sharing tools such as Git, the use of Persistent Identifiers and the management of data in general, and explicit in terms of running specific module on Open Science. This module, however, does not take a didactic "you must do Open Science" approach as many ECR's are often isolated and without support of such approaches. The module is based on asking the students to reflect on aspects of Open Science, particularly from an ethical standpoint and hence to enable these students to approach Open Science in a more robust fashion, determining what practices they can adopt and the benefits they will gain. This ethical approach is amplified by the use of exercises throughout the school which relate this ethical aspect of Open Science with the specific module being taught. The impact of this approach has been exemplified by the comments on how useful the Open Science aspects have been to them and have spreading the practices in their home institutions.