

Research computing and data science training – towards a College-wide offering

Katerina Michalickova

Senior Teaching Fellow, Research Computing and Data Science Programme Leader,
Graduate School

December 1, 2021

Action requested

In consultation with the RCS governance group and the founding members of the training activity, we believe the College would benefit from:

- Expanding the training activity to meet the current demand and making it available to students, researchers, and staff at all levels.
- Explicitly connecting the training activity and other research computing activities, such as the Research Computing Service and the Research Software Engineering community.

We seek suitable road map on how to best push forward this ambition.

Executive Summary

This document concerns the Research Computing and Data Science Programme in the Graduate School. It summarises training achievements so far and outlines the priorities for future development that aim to maximise our contribution to the College's strategic mission.

The training provision is currently part of the Graduate School where it primarily supports the doctoral students. Before moving there, the activity was attached to the Research Computing Service and the Computational Methods Hub where it was about 50% smaller and open to everyone at the College. The Graduate School provided a productive environment for the training team and enabled the staff to develop a high-quality suite of workshops that cover essential topics in research computing and data science. Our research computing doctoral training programme is ahead of any other UK institution and this advantage is likely to generate returns in the scientific output of the College. We need to maximise this opportunity by enabling the rest of the College students, researchers and staff to gain access to the same high-quality training. It is also essential that the training activity is presented as an integral part of the overall research computing strategy for the College. The paper presents a possible model for such expansion and gives examples of collaboration between the Graduate School and other parts of the College; collaborations that could serve as a blueprint for a College-wide training activity.

Background and Current Status

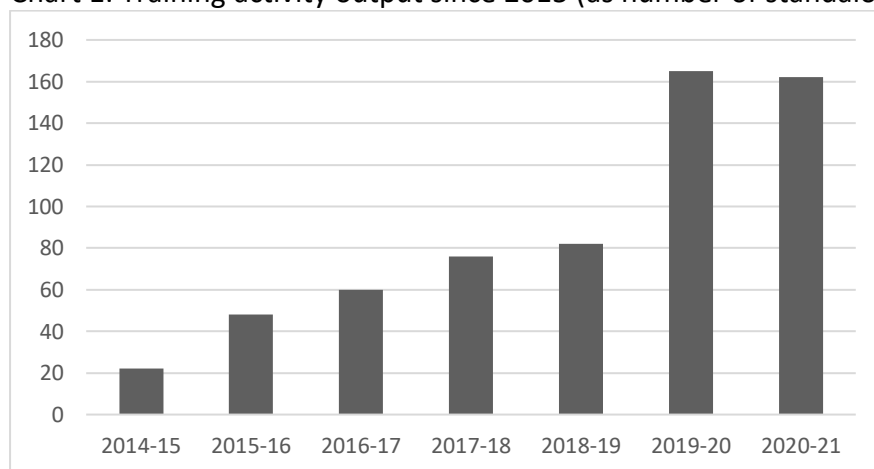
The need for a comprehensive research computing and data science training programme at a leading STEMM university is indisputable – and underpins our [strategic ambitions](#) for transdisciplinary, impactful, technology-enhanced research.

The central training provision was developed with guidance from the [Centre for Computational Methods in Science and Engineering](#) and the [Research Computing Service](#) (RCS). Between 2014 and 2019, the training activity was supported by 1.5 FTE (K Michalickova and two successive 50% positions shared with Mathematics), and attached to the RCS and the [Computational Methods Hub](#). This activity laid foundations for a training programme that exists today.

Since 2019, the [Research Computing and Data Science Programme](#) (RCDS) has been part of the Graduate School (GS) and consists of three GS Teaching Fellows (TFs) and one TF funded by the Business School, who divides their time between the GS and the Business School. The team members have a strong background in science, research computing and data science. The new home brought much-needed stability and career progression for the training staff and enabled us to develop a high-quality provision. As members of the learning and teaching family, we undertake part-time teaching training (PGCert, PGDip and MEd) or coaching accreditation. As part of the GS, we have also been exposed to various aspects of PhD student pastoral care that informed our Programme design decisions.

Since 2015, the training output has been steadily growing and, currently, the Programme delivers over 160 stand-alone courses per year (Chart 1). We cover almost 30 [topics](#) in statistics, data analysis, machine learning, programming, software engineering and large-scale computing. Our activity accounts for about 40% attendance of the Graduate School's Professional Development Programme for doctoral students; a significant proportion that was achieved in just two sessions and with courses that have multiple parts (as opposed to the rest of the GS courses that are scheduled as a single session).

Chart 1: Training activity output since 2015 (as number of standalone workshops)



The GS Programme has been shaped by multiple inputs: the team’s expertise and experience in consultation with the RCDS steering group, student partners, the RCS, the academic community, and student and employer surveys. Our teaching activity is primarily directed towards PhD students; every academic session, about 2,000 students gain a GS credit through the RCDS Programme, and the feedback has been overwhelmingly positive.

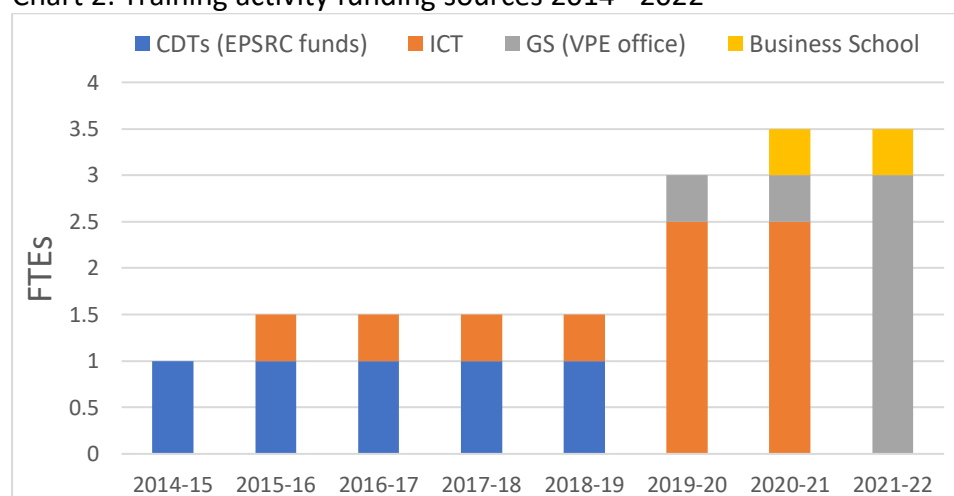
We are just about meeting the demand for the current workshops, however, with a portfolio of almost 30 workshops, our capacity to add new topics is minimal. To a smaller extent, we also support MSc students (via cohort delivery) and 2nd-year undergraduates (via an I-Explore module in Interdisciplinary Research Computing delivered in collaboration with Jeremy Cohen from DoC). We have also coordinated a local [Software Carpentry community](#) of volunteer instructors.

Challenges/Opportunities

Our next challenge is to find a way to expand the training activity beyond the Graduate School to support students and researchers of all levels – the same way that the Research Computing Service supports the whole institution. We have made significant progress, but the existing programme is having difficulties to reach an audience other than doctoral students.

This is an organisational challenge that will require coordination from representatives of Faculties, departments, and Administration and support services. Since 2015, the training activity has had been associated with and funded by various parts of the College (Chart 2). The funding sources, to a large degree, determined our primary audience; that is more true now than before. The founding members of the activity always strongly believed that comprehensive research computing training should be open to all students and researchers, and we would argue that the time is right to start discussing a true College-wide provision.

Chart 2: Training activity funding sources 2014 - 2022



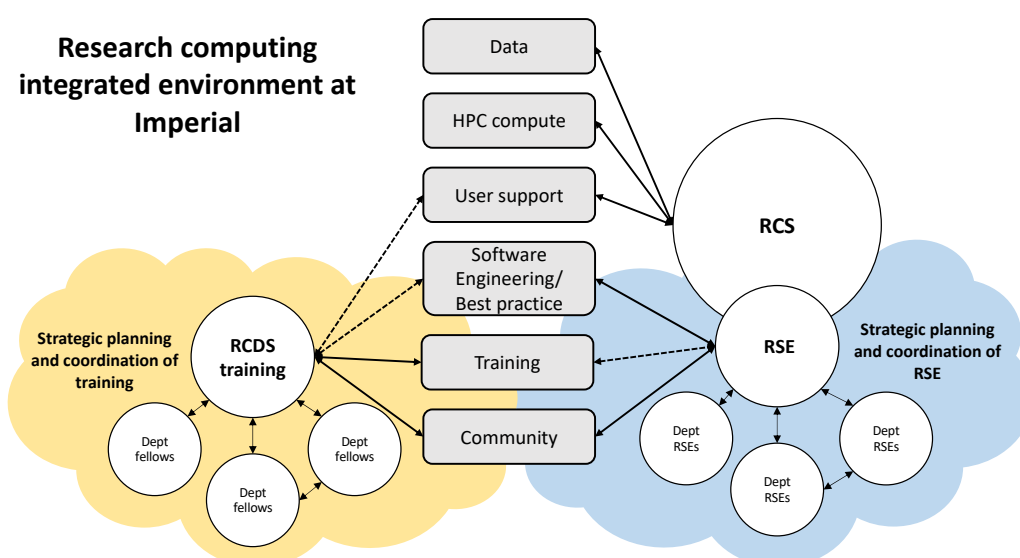
The Graduate School has been supportive and open to the idea of training non-PhDs. However, with finite resources, it became challenging to position the Programme as a flexible College-wide provision and, at the same time, satisfy Graduate School's rigorous quality standards and process. The course format became specialised with a strong focus on active learning; a format that requires extra time and may not suit everyone. Overall, we have been partially successful in keeping the Programme open, currently, approximately 15% of course occupancy is offered to non-PhDs and the GS administrative team manages the process. Since the team is at maximum capacity, we don't actively advertise or map the interest. Under this arrangement, we fill about 60% of non-PhD spaces.

Next steps

We believe – and invite you to support – our mission is to offer accessible, appropriate research computing and data science training to all students, researchers, and staff at the College. The provision needs to be tightly connected with RSE activities and RCS support – to provide a one-stop, seamless pathway and to maintain technical excellence and expertise for the staff involved.

Figure 1 presents the above ideas as a schema of an integrated College-wide research computing environment that includes the existing entities: RCS, RSE and RCDS training; the latter two enriched by departmental participation. The schema shows the contribution of each entity to six fundamental functions that underpin a successful research computing provision. While some functions are covered by a service (data, compute, user support), some go beyond – training and good software engineering practice require active participation from stakeholders and their community.

Figure 1: Research computing integrated environment at Imperial (grey - the essential functions, white - existing entities, blue and yellow – function spans a central group and departments).



Closer proximity and regular communication between all research computing staff (RCS, RSEs and RCDS) will benefit all and will enable us to develop our expertise and tackle tasks that require combined input. For example, the user support and training staff are essential to developing effective user support resources. Similarly, training can be delivered in collaboration with RSEs.

A successful research computing and data science training provision is likely to span the Central Faculty as well as the departments. The central team is well placed to deliver fundamental and cross-cutting concepts. More specialised topics are best learned via realistic disciplinary exemplars coupled with effective explanations – an activity that is best housed in departments. To make this an effective mechanism, strategic planning and strong communication must take place between the central and local staff. Perhaps, combined appointments and shared space may offer a solution.

Funding models for expanded provision

To inform future discussions about College-wide training, we would like to describe an agreement that GS reached with the Business School who were interested in hiring a dedicated research computing TF. To ensure that the Teaching Fellow benefitted from the established learning and teaching fellow pathways in the GS as well as its extensive networks, we have agreed that the TF will be employed through GS and funded by the Business School. This fellow divides their time between module teaching and delivering GS courses. This arrangement reflects the GS' contribution to line managing and professionally supporting and developing the TF. The arrangement has been working well and we are now looking to create one more equivalent position.

Other models are possible. The GS can host a TF dedicated to a Faculty, or a departmental fellow can work part-time with a central team. In the past, we have successfully used the latter in the Computational Methods Hub that shared a fellow with Mathematics.

Objectives for next 12 months

The key parts of the research computing ecosystem at Imperial are distributed across various entities at the College. The path towards one-stop seamless provision and support will require College mechanisms that may not yet exist. On the training side, we will actively pursue transforming our activity into a College-wide comprehensive offering. We sincerely hope that this document will open a conversation between key academic and administrative representatives about ownership and funding of research computing and data science training, and excellence in research computing support at the College.