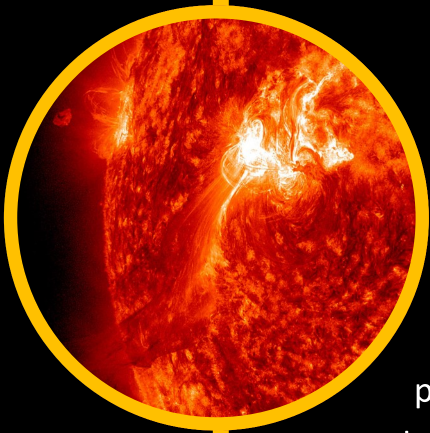
Open Science Training Initiative: Post-Pilot Report

April/May 2013



Introduction

Welcome to the post-pilot report for the **Open Science Training Initiative (OSTI)**, an educational scheme which aims to drive the production of high-utility, high-impact research in the sciences by training students in the delivery of reproducible outputs in the form of a **coherent research story**, using approaches and techniques drawn from the Open community.

This report aims to summarise the setting and outcomes of the pilot initiative, which took place at the University of Oxford in January 2013. It provides insights into how existing courses across the sciences can be modified to deliver subject-specific education whilst also training students in contemporary research methodologies and scientific working culture.

Our analysis unites diverse comments, opinions and feedback from a variety of sources and includes perspectives from the course leader, auxiliary demonstrators and student participants. We hope that these insights will prove equally useful to researchers and educators alike.

Introduction

Findings

Feedback and responses from the lead lecturer, auxiliary demonstrators and the students themselves from a highly successful OSTI pilot have provided a detailed picture of how the OSTI approach works in practice.

- Young researchers deliver work of increased utility and quality when **provision for the research user** in their output is incentivised;
- Students showed a keen **preference for rotation based learning** over traditional single-project assessment;
- A **hands-on approach** to licensing implementation and data management is instrumental in building the students' confidence in the use of Open techniques;
- The majority of research students start out with **minimal awareness of Open practice**, but quickly absorb and apply these methods once introduced to the main concepts;
- Doctoral students are keen to engage directly in debate about the evolving research landscape.

Course Outlook

The upcoming release of CC-BY licensed OSTI materials via our GitHub repository will split the resources into three main categories:

- **Core Lectures:** Lectures on reproducibility and open science; code, content and data licensing; and the changing face of publication emerged as the most useful in their current form and are recommended for use in all future OSTI courses;
- **Practical Workshops:** Version control and data management are best delivered as practical workshops delivering first-hand experience, as opposed to passive learning in formal lectures;
- **Enhancement Lectures:** Cultural awareness lectures on the work of the Open Knowledge Foundation or similar organisations can be included where possible.

Further information on the structure and aims of the Open Science Training Initiative can be found on the course website, www.opensciencetraining.com.

Motivation & Aims

Traditional approaches to graduate training frequently reinforce the role of **research producer**, while neglecting to promote the role of **research user**. A joint study by JISC and the British Library in June 2012, entitled, “Researchers of Tomorrow” [1], identified several difficulties faced by modern PhD students over the course of their doctoral research.

According to the report, Generation Y students:

- have a strong inclination towards “*face-to-face support and training*” and favour subject-specific teaching over generic content;
- hold many misconceptions about open access publishing, copyright and intellectual property rights;
- are slow to utilise the latest technology and tools in their research work, despite being proficient in IT;
- are heavily influenced by the methods, practices and views of their immediate peers and colleagues.

These findings highlight the need for **subject-specific training** across the sciences, utilising a **hands-on approach to learning** and delivering an **integrated approach** to data, licensing and open science as part of the natural research process. Such training needs to be delivered at a pre-doctoral stage wherever possible, if we are to overcome the broad variability in ethos across the research groups young scientists initially join.

Reproducible Science

Ever greater focus is being given to the problem of **reproducibility** in scientific research. As an example, Begley & Ellis [2] outline the severe difficulties encountered in preclinical cancer research, citing two separate studies by Amgen and Bayer Healthcare. Each of these studies attempted to reproduce the findings of several landmark papers in the field, with only 11% and 25% success respectively. The authors call for greater publication of negative results; greater credit for teaching and mentoring within the science disciplines; and reduced reliance on publication metrics as the sole means of measuring research contribution.

Introducing OSTI

Young scientists need to be taught not only that reproducibility is an issue, but also how they can practically achieve this in their own research outputs. Emerging techniques in Open Science have much to contribute to this area, and it is vital that our newest generation of researchers is equipped with this knowledge and feels fully engaged in current discussions surrounding changing research practices.

The **Open Science Training Initiative (OSTI)** aims to address the reproducibility issue by harnessing open science techniques to improve research clarity, delivered via a novel, dynamic style of graduate teaching.

Awareness of reproducibility issues

- Provide first-hand experience of research from the user perspective;
- Instil an awareness of the value of good research communication;
- Promote the delivery of a coherent research story involving code, data and written output, rather than the report alone;

Acquisition of technical skills

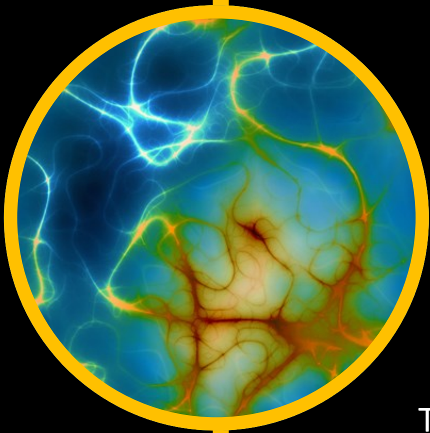
- Educate students in the technicalities of licensing, data curation, data preservation and open science;
- Provide first-hand experience of licensing and data management;

Knowledge of scientific working culture

- Equip PhD students with an awareness of modern research practices and encourage them to critique the prevailing research culture and the viable alternatives;
- Empower students to make considered choices about their own research careers, by developing an awareness of open science and its impact on their publication options.

A novel teaching structure, which we term **rotation based learning (RBL)**, is central to the OSTI approach. RBL is vital in placing equal emphasis on the research user/producer roles and in stimulating the need for reproducibility of research outputs. The following pages provide details of the basic course pattern for OSTI and its outcomes in practice.

Course Structure



General Structure

Rather than being a stand-alone course in its own right, OSTI is a rotation-based teaching template, accompanied by lectures in digital research techniques and open science, designed to slot over existing, subject-specific course provision.

The bulk of the timetable is given over to research-style work in small groups, save for one short (30 minute) lecture per day. The course leader also holds short supervisory sessions with each of the groups on a daily basis. The research component is delivered in two phases via Rotation Based Learning, which we shall now describe.

All discussion of "projects" and "research problems" in this report refers to subject-specific content selected by the course lecturer.

Rotation Based Learning

The student cohort is split into separate groups. When running OSTI in an interdisciplinary setting, efforts should be made to distribute the subject backgrounds across the groups. No communication is permitted between the groups at any stage during the exercise: this provides the stimulus for reproducible research by encouraging students to view their own work from the perspective of a potential user.

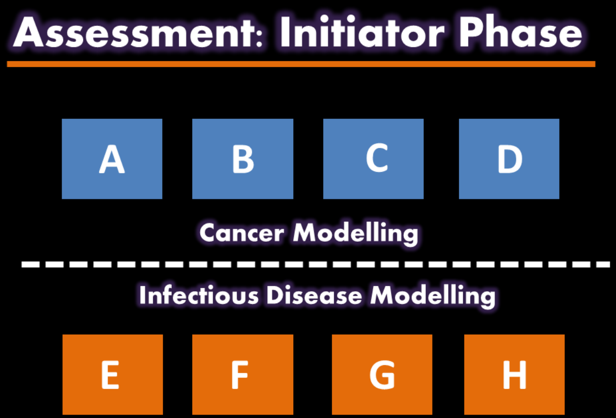


Figure 1: Example group structure for Phase 1 (Initiator Phase) showing the theme allocations.

Phase 1 is the **Initiator** phase. Groups are allocated one of two subject areas and given a list of pre-selected papers from the relevant literature (in the example of Figure 2, namely cancer modelling and infectious disease modelling). Each group must select one of the listed papers as their focus and subsequently attempt to reproduce its results. They are asked to deliver a **coherent research story**: not just a written report, but also the accompanying data, code and figures, all appropriately licensed.

Assessment: Successor Phase

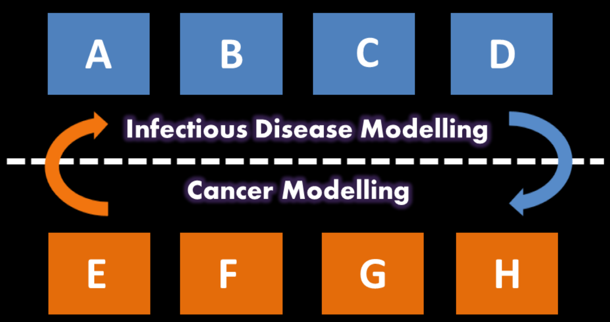


Figure 2: Example group structure for Phase 2 (Successor Phase) showing the theme rotations.

Phase 2 is the **Successor** phase in which the projects are rotated. The inherited research must first be verified by the successor group (for example, validation of code, data sets and figures) before the work can be extended, in the manner of a novel research project. The end of Phase 2 again sees a full submission of a coherent research story with suitably licensed components. Students understand from the outset that their work will be graded on the openness and reproducibility to subsequent users, as well as the quality and innovation of its research.

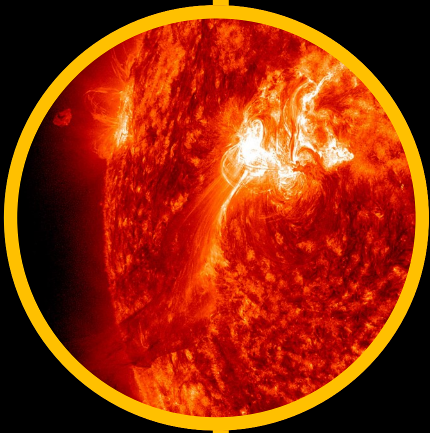
Figures 1 and 2 depict an eight-group RBL structure as used in the OSTI pilot. This structure is readily amenable to modification when working with smaller or larger cohorts; similarly the research themes should be chosen to suit the subject specificity of the underlying course.

Daily Lectures

Short, daily lectures integrate with the work timetable to provide general information about digital approaches, data management and research culture. These methods are consolidated by direct application to students' own work. Lecture content is independent of subject specificity and is designed for portability across the sciences. Lecture content is described further on page 11.

Teaching Support

The learning process is supported throughout by a team of experienced PhD students and postdocs, who remain on hand to provide technical support and advice to the students. Meanwhile, the lead lecturer/course leader acts as supervisor to the groups, meeting with each team for 20-30 minutes daily to discuss their research progress and assist with any queries about implementation of open science practices.



Pilot Scheme Intake

The OSTI pilot scheme was delivered at the **University of Oxford** in **January 2013**, within the Doctoral Training Centre (DTC) for Systems Biology, Life Sciences and the Industrial Doctorate *(www.dtc.ox.ac.uk)*. The DTC provides a pre-doctoral taught year of compulsory modules in experimental and theoretical sciences.

A cohort of forty-three students were involved in the OSTI pilot. The majority were recent graduands from four-year undergraduate masters programs, or from one-year MSc courses. A small number came to the DTC from employment in non-research sectors (e.g. finance). Academic backgrounds included all areas of the physical (maths, computer science, physics, engineering) and life sciences (biochemistry, biology, chemistry, pharmacology); two of the physicists had specific specialisation in chemical physics. The balance of subject backgrounds is shown in the bar graph of Figure 3 below.

Forty-two of the students were permanent members of the DTC, attending for the entire taught year. One additional student joined the course as a guest attendee from the Physics department; this student was a first-year PhD researcher who wanted to develop his programming skills.

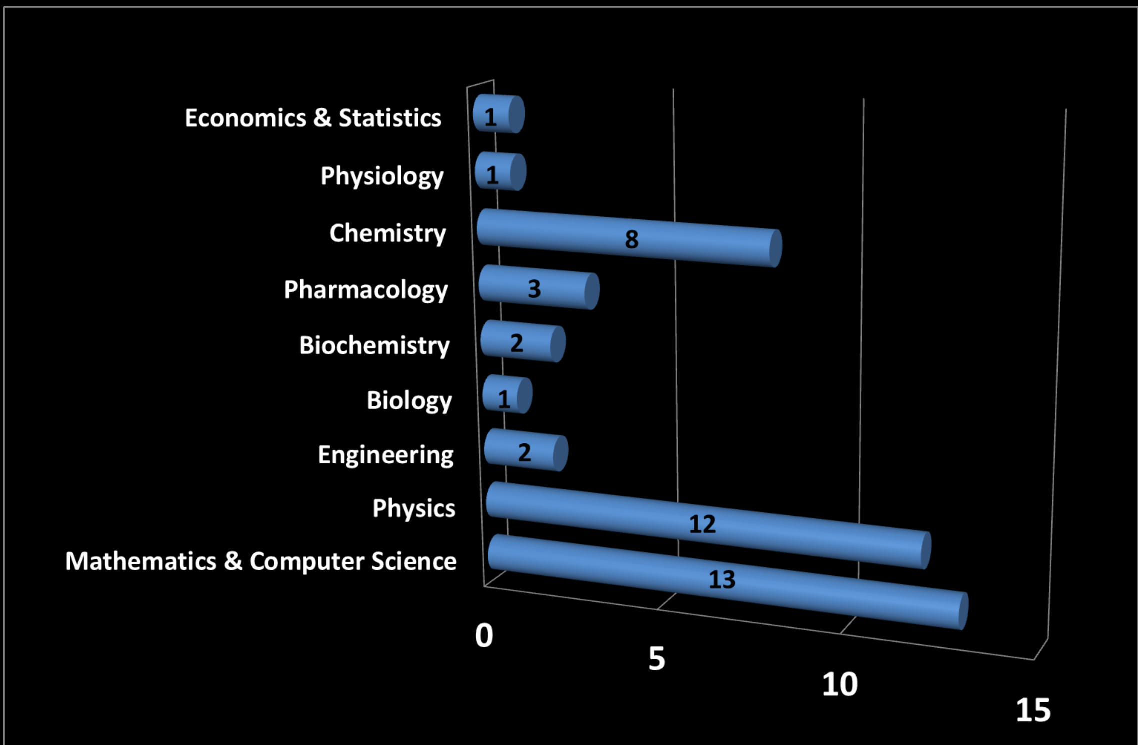


Figure 3: Subject background of the student cohort

Course Modification Using OSTI

Scheduling demands at the DTC already required delivery of a three week course in **computational and mathematical modelling**, incorporating mandatory instruction in Matlab programming.

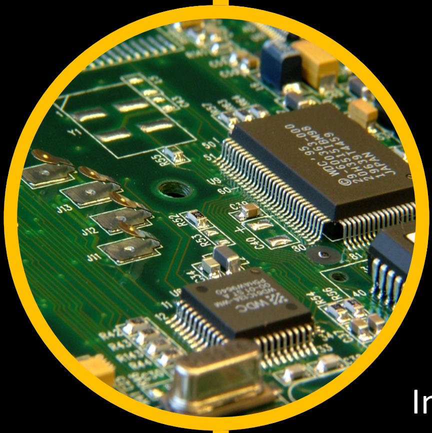
In previous years the same course had delivered two weeks of traditional teaching, followed by a one week assessment involving group work on a single project. In a departure from this model in 2013, the first half of the three-week course involved programming training in Matlab, delivered by an existing DTC director and auxiliary demonstrating team. The Open Science Training Initiative was run over the remaining time, delivering a research-style assessment in computational modelling using a combination of hands-on, rotation based learning and daily OSTI lectures.

The ideal scenario would have been to use non-proprietary software (e.g. C++, Python) to create a completely open ethos. However, the reality of teaching in all institutions dictates that training in Open may often be required to fit around whatever teaching space is available and this was the case for the pilot setting.

Consequently, this report aims to show how existing teaching programmes can be adapted to emphasise the research user role, while also introducing training in Open techniques, without compromising on subject specific provision. Furthermore, the pilot scheme showed the students that situations requiring use of commercial software need not preclude release of open reports, data etc., and that they ***can*** and indeed *should* strive to maximise the openness of their outputs wherever possible.

Research Problems

To facilitate the rotational structure of the course, research themes were restricted to two broad areas: **cell-based cancer modelling** and **infectious disease modelling**. A series of papers from the published literature in the field were pre-selected and Phase 1 required each group to select one from their designated theme as a starting point for further investigation.



Research Focus (cont.)

The two themes demanded interdisciplinary collaboration within each group, drawing on combined strengths and knowledge in mathematics, computer science and biology. Four groups were working on each theme at any one time.

In addition to delivering daily lectures, the lead lecturer acted as research supervisor to the groups throughout the course. Typically this involved seeing each team for half an hour every day to chart their research progress and answer their questions on licensing and open science (see photos below and right).

Rotations

Students were informed at the outset of the assessment period that communication between groups was prohibited throughout Phases 1 and 2.

Phase 1 (Initiator Phase, 3 working days): Students were tasked with selecting a paper from their assigned theme area and attempting to reproduce results from this existing, peer-reviewed research. From the outset, students were mandated to deliver a **coherent research story** comprising code, data and a written report, all appropriately licensed, rather than a traditional assessment based purely on a single written output.

Each group also had to form a data management plan to accompany their Phase 1 hand-in, using the “DMP20” tool which provides **Twenty Data Management Planning Questions** (available online at *http://www.miidi.org:8040/dmp/*). Students were required to version control their projects via GitHub throughout the course via the GitHub for Windows interface, to assist in collaboration and project handover.

Rotations (cont.)

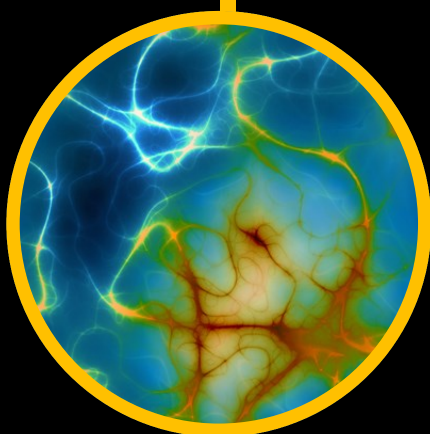
Phase 2 (Successor Phase, 3 working days): Projects were rotated such that cancer modelling groups inherited work in disease dynamics, and vice versa. Communication between the groups was still prohibited; each team had to develop their new project as a research problem, building on the work of their predecessors. Trained demonstrators (postdocs and PhD students with teaching experience) were invaluable in the smooth running of both Phases and remained on hand in the main work area to assist the students with technical queries. Phase 2 required submission of another coherent research story and culminated in **presentations from each group** on the final day.

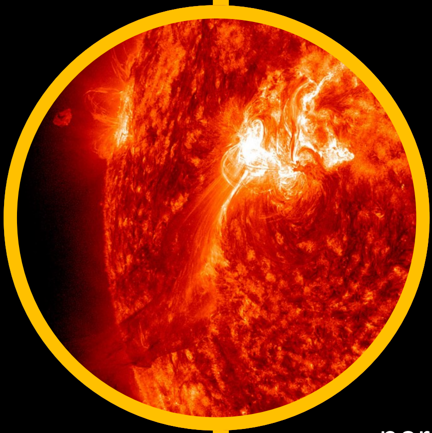
Lecture Content

The simulated research environment provided by the rotation structure and supervision sessions was supplemented by a series of daily half-hour lectures on topics in open science, on the following themes:

- **Introduction: Reproducibility and Open Science**
- **Version Control Using GitHub**
- **Data, Code and Content Licensing**
- **Data Management Plans** (followed by a guest talk from Jun Zhao on scientific workflows)
- **The Changing Face of Publication**
- **The Work of the OKFN** (session co-presented with Jenny Molloy)

At the request of the DTC, we also accommodated external (non-OSTI) lectures in **How to Write Mathematics** and **Interfacing between Matlab and Open Source Software**.





Practical Implementation

Several students voiced enthusiasm for the RBL approach from the outset and appeared to be highly energised and motivated by this new approach to learning. Nonetheless, some students were already set in the role of research producer before the OSTI course and initially found it difficult to adjust their perspective whilst also learning new techniques. The **daily supervision format** provided an excellent forum to discuss these issues and help the students to develop their outlook and critique their own methods. A four or five day rotation length may help to ease this progression in future instances, where timetabling space allows.

Version Control

It was difficult to determine an appropriate level for the Version Control (VC) lecture, given the broad range of scientific backgrounds amongst the students. Computer scientists in the cohort were already highly skilled users of version control in program development and expressed personal preferences for specific VC systems such as Subversion or Mercurial. On the other hand, many students had not even heard of version control before the lecture and were therefore new to VC using Git.

The collaborative RBL environment necessitated use of the same VC system for all participants, in a way that would be accessible to less confident programmers. Git was selected as the VC system for the course and students were expected to create a public repository on the GitHub website to manage their team's project. This enabled students to see VC as a means of streamlining project development for writing ***and*** code, as well as facilitating project swaps at the start of Phase 2.

Unfortunately, difficulties were encountered in the technical implementation of version control, owing to the use of **GitHub for Windows**. It was hoped at the outset that its user –friendly interface would assist the less confident students in implementing VC, whilst established programmers could choose to use the command line. Despite excellent IT support during the course, the GitHub for Windows program proved highly volatile on the DTC network and significantly hindered the learning process, even when used correctly, and had to be dropped midway through Phase 2.

In future instances we would **recommend direct use of command line Git** for ***all*** students, possibly providing a short practical session on VC for new programmers at the start of the assessment. The command line approach can then be used in tandem with a GitHub repository to manage project creation, development and handover seamlessly. Use of an online repository also facilitates assessment by the course leader, who can synchronise their own local copy of each group's repository with the master copy at each hand-in deadline and thus obtain an entire research story with ease.

Data Management Planning

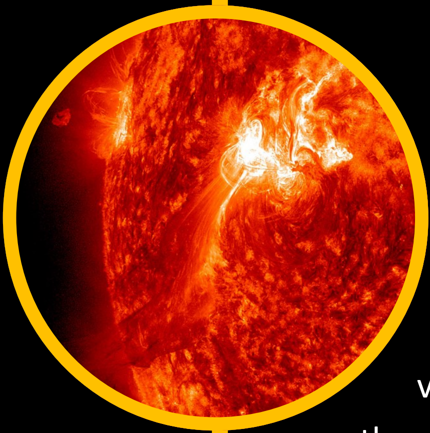
The project hand-in at the end of Phase 1 also required completion of a data management plan (DMP) by each group. The online tool, **"DMP20",** by Tanya Gray and David Shotton at the University of Oxford, was used. This provided an accessible way of introducing the students to creation of simple DMPs and was easy to integrate into the course when used in tandem with a short lecture on data management. Questions relating to long-term data management proved more challenging for the students to answer and they would have benefited from extra time in the course to allow for group discussions. The success of this part of the course could be increased still further by extending the rotation phase length by one or two days.

Licensing

Direct, practical application of licensing is one of the strengths of the OSTI approach. The initial lecture on the subject was necessary to introduce the main concepts, but it was through applying these methods to their own work that students consolidated this knowledge.

Students quickly grasped the principles of licensing and in several cases began to ask sophisticated questions about the finer points of licensing implementation when multiple forms of content were involved. Furthermore, the Licensing lecture was the most recommended session of the entire lecture series.

It is likely that this **hands-on approach to learning** was a significant factor in students' confidence with Open techniques by the end of the course (see page 18 for analysis). We would recommend this aspect be included in all future OSTI courses.



Phase 1 Findings

The objective of reproducing research from the published literature provided a strong goal to shape the students' work throughout Phase 1. One of the chief aims of this exercise was to demonstrate that although peer-reviewed research may provide interesting and potentially groundbreaking scientific claims, the utility of that research may be limited if the findings are not communicated with sufficient clarity.

Early on in Phase 1, many students reported verbally that the broad ideas of their chosen paper were easy to grasp, but that implementation of the same methods—or, in some cases, even ascertaining what methods had been used in the research—was significantly more difficult, owing to missing information. Critiquing the published literature in this way helped them significantly in shaping their own approach to delivering a coherent research story at the end of the Phase.

Still from the Reproducibility lecture at the start of the course

Phase 2 Findings

Rotation of projects at the start of the Successor phase was significantly helped by the use of online GitHub repositories but would have occurred much more smoothly were it not for ongoing problems with the GitHub for Windows interface. The majority of groups spent most of the morning session of day one of the Phase reading through their inherited projects and deciding how they were going to develop them further. Given the demands for novel research in this Phase, we would recommend allocating more time to this phase in future OSTI instances, as some of the students commented that they would have liked more time to explore their ideas.

Research Output & Innovative Use of Data

The Successor phase in particular provided groups with the freedom to explore their own research ideas. Overall standards of work were extremely high, despite students' tendencies to take a highly critical approach to their output. Some groups delivered particularly innovative work which sought out data from external sources in developing their research projects.

Some examples of student innovation and creativity are given below.

- One of the groups who inherited an influenza model for virus-host dynamics in Phase 2 chose to create a spatial embedding of the model to explore population-level epidemiology. They **sought out large data sets** detailing flight frequencies between major US cities and used these data to create a diffusion matrix for their mathematical model, allowing them to simulate the spread of influenza between cities;
- The same group also turned their numerical outputs into a **visualisation application** that created movies of the simulations for a user-defined parameter set;
- One group included a routine in their code to **automatically write appropriate licensing statements and tags** into the data output files;
- Another group took their predecessors' code which delivered a video output to screen and **augmented this with a script to efficiently store the relevant matrices** and catalogue them for post-processing and further analysis;
- Yet another group **interfaced their code with an external open-source library** to speed up large-scale implementations of their code for future users.

These applications all arose from students' own ingenuity and knowledge, inspired further by data-related discussions in lectures and daily supervisions. A combination of RBL with integrated training in data techniques holds strong potential for fostering creativity and good practice in our newest generation of researchers and information specialists.

Course Leader Perspective

Participant Feedback

Student Questionnaire

A questionnaire was communicated to each of the 43 participants on the final day of the course. 32 students completed the questionnaire anonymously and their responses to the three themes are presented below: Learning Outcomes & Impact (page 16), Rotation Format (page 19) and Lecture Content (page 22). Comments from the final two questions, which allowed a freeform response, are detailed on page 24-5.

1. Learning Outcomes & Impact

Section Aim: To determine the success of the OSTI scheme in improving awareness of scientific culture and building students' confidence in the implementation of open science methods and data/digital management approaches.

Q: How would you have rated your knowledge of reproducibility and open science BEFORE taking this course?

Students were asked to rate their awareness of open science and reproducibility issues on commencing the course, ranging from 1 (complete

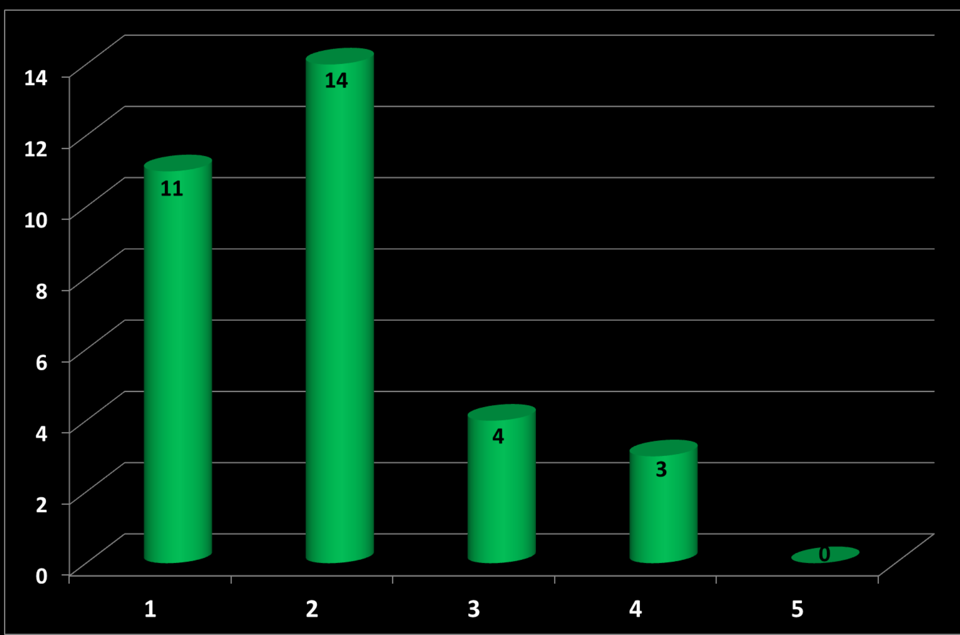


Figure 4: Students' self-rated awareness of reproducibility & open science before taking the course, rated from 1 (complete beginner) to 5 (expert)

beginner) to 5 (expert). As can be seen from Figure 4, the majority of respondents (25 of 32) rated their incoming knowledge as 1 or 2, and there were no self-rated 'experts' in the group.

Closer inspection of the responses indicates that the 3– and 4-rated students were all from physical sciences backgrounds and had significant programming experience. It is likely (though we cannot be certain of this) that these students had existing knowledge of open source code development. From the course leader's perspective, very few students appeared to be aware of content and data licensing at the very start.

Q: To what extent has this course contributed to your awareness of current practices in scientific research?

One of OSTI's key objectives is to raise awareness of research culture at a pre-doctoral level, before young researchers begin to draw on the ethos of their

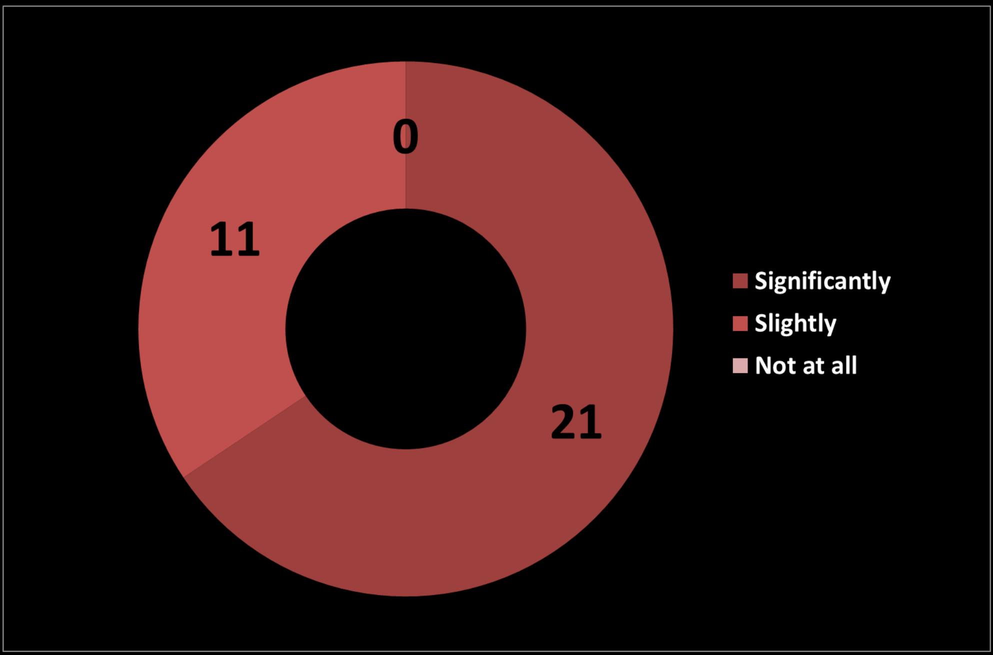
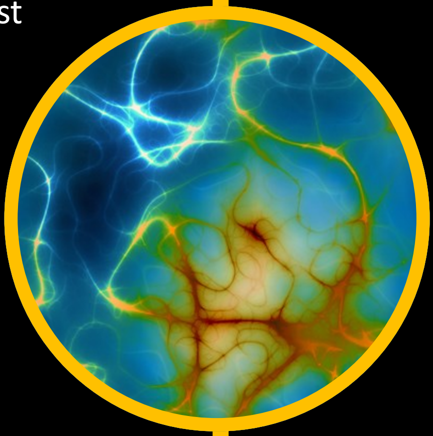


Figure 5: Students' opinions on the contribution of OSTI to their awareness of scientific culture & practices. Numbers on the plot indicate overall totals for each response.

first research group. The responses summarised in Figure 5 suggest that the OSTI scheme achieved considerable success in improving students' awareness of modern scientific practice. All respondents felt that the course had improved their knowledge of current working practices to some extent. There was no apparent correlation between a student's subject background and their response to this question.



Where appropriate, would you feel confident applying Open techniques to your work in the future, either by yourself or with help/ guidance?

A significant majority of respondents would be happy to implement open science practices in later work, as summarised in Figure 6. Of these, 13 respondents would prefer additional guidance in doing so. It is therefore vital that OSTI course lectures identify **accessible sources of support and advice on Open matters for the students**, whether this takes the form of local, in-person advisors specific to the host institution, or names and details of online information portals and advisory organisations. Ideally a handout needs to be produced as a postcard-style contact list, which is given out during the course and which the students can keep thereafter.

The three neutral responses were drawn exclusively from Physics and Chemistry backgrounds; however, given the small numbers involved, it is difficult to draw any firm inferences, particularly given that other students from the same subject backgrounds did not respond neutrally.

Drawing on Figures 4 and 6, we observe that **78% of respondents entered the OSTI course with minimal knowledge of open science,** while

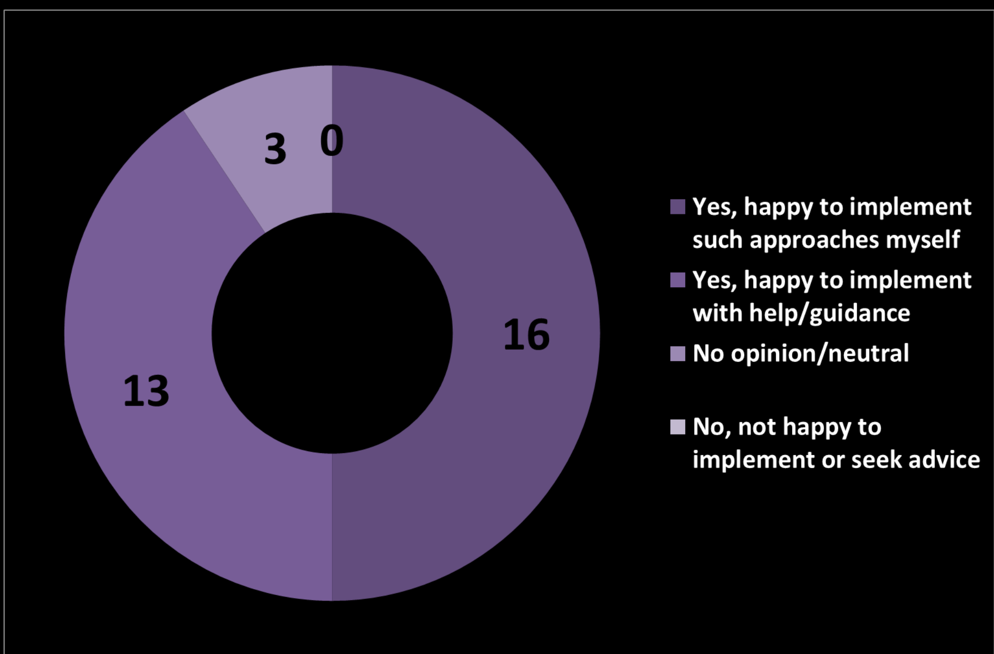


Figure 6: Confidence in applying Open techniques to future work. Numbers on the plot indicate overall totals for each category.

87.5% of respondents had acquired confidence in the use of open science approaches to research by the end of the course.

2. Rotation Format

<u>Section Aim:</u> Previously, the same DTC course had been run using a more traditional assessment method, in which students received two weeks of programming tuition before undertaking a one week assessment involving submission of a single project per group. Section 2 of the exit questionnaire addressed students' attitudes to OSTI's novel RBL structure, with regard to their preferred working pattern and the length of the rotation phases.

What style of assessment do you prefer?

Responses relating to the RBL approach overwhelmingly favour it against the traditional method of assessment (Figure 7). There was no apparent correlation between preference and subject background of the respondent.

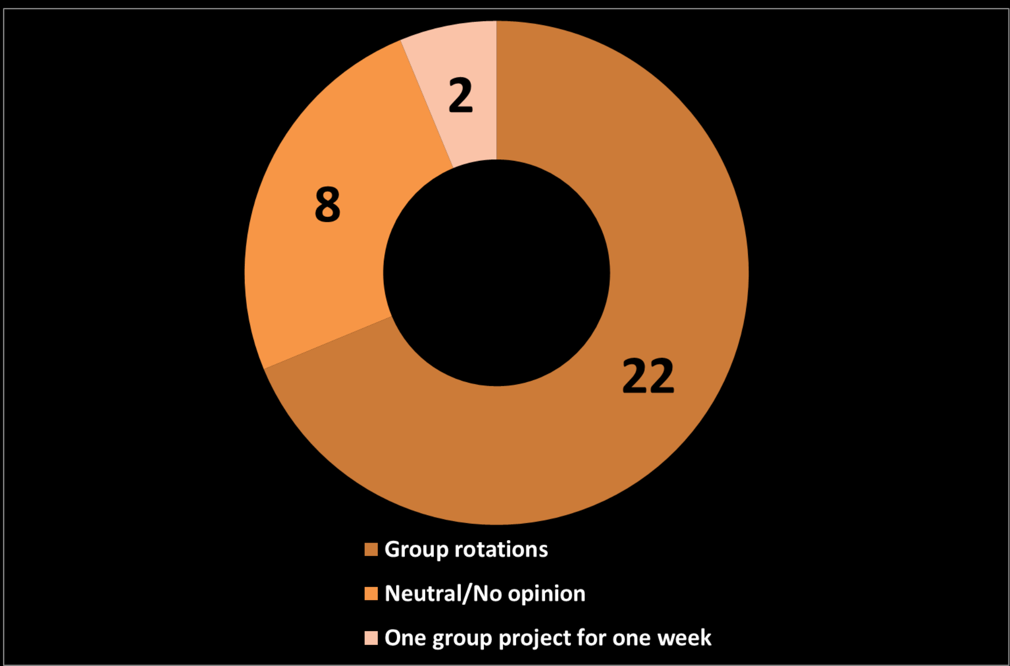


Figure 7: Preferences for assessment styles. Numbers on the plot indicate overall totals for each category.

Such results provide excellent support for the use of rotation based learning to drive student engagement with the learning process. The two respondents with a preference for a "single project" style assessment were the same two students who indicated a preference for a phase length of 6-7 days (shown in Figure 8 overleaf).

Q. Given your experiences on this course, what timeframe do you feel is appropriate for each phase in the rotation?

Addressing the diverse needs of a variety of subject backgrounds proved one of the main challenges during the OSTI pilot. Project timeframes can be quite difficult to perfect when delivering a mathematical/computational course in an interdisciplinary setting, where the more mathematical students can make rapid progress with a problem, whilst others may require more time to consolidate their

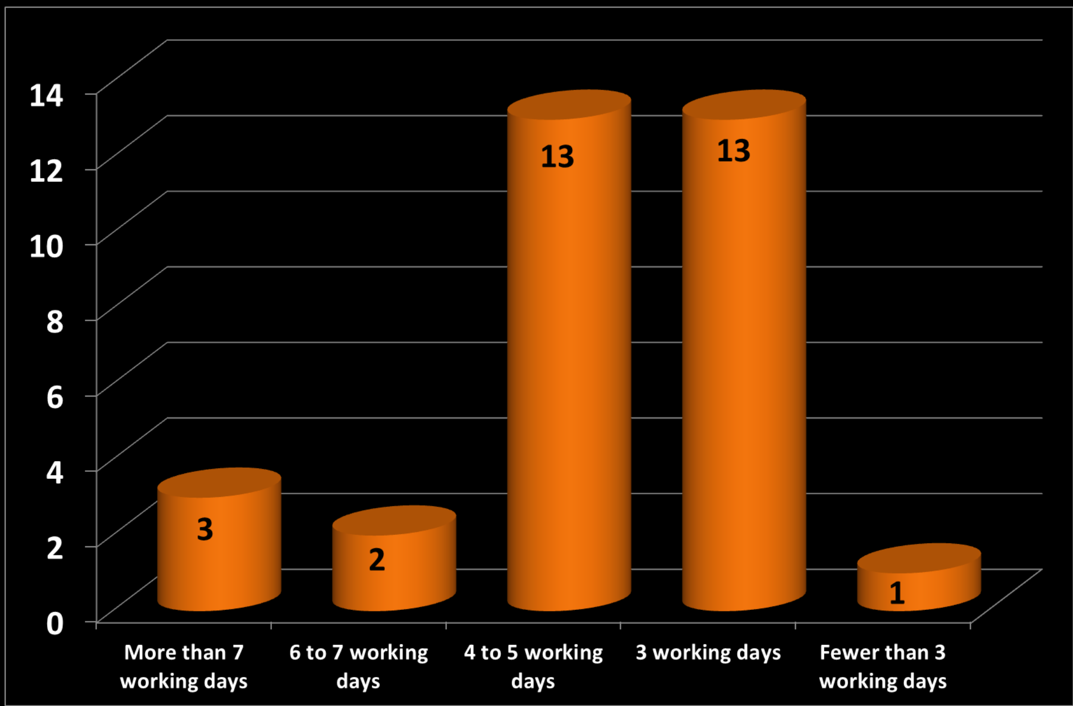


Figure 8: Students' perspectives on the length of rotation phases. Numbers on the plot bars indicate overall totals for each category.

knowledge. The responses shown in Figure 8 suggest that **a phase length of between three and five days is optimal**. The majority of requests for a longer rotation time came from physical science students, many of whom wanted more time to explore the finer detail of the mathematical models, as opposed to having any problems grasping the major concepts in the time allocated.

Prospective OSTI leaders may therefore find a 4-5 day phase length appropriate when working in an interdisciplinary setting. Alternatively the scope of the research problems could be scaled down if timetabling capacity is limited. Single-subject cohorts, on the other hand, may be capable of a three day rotation period, given judicious selection of the initial papers.

Q. Would you recommend use of the rotation-style assessment format for this course in future years?

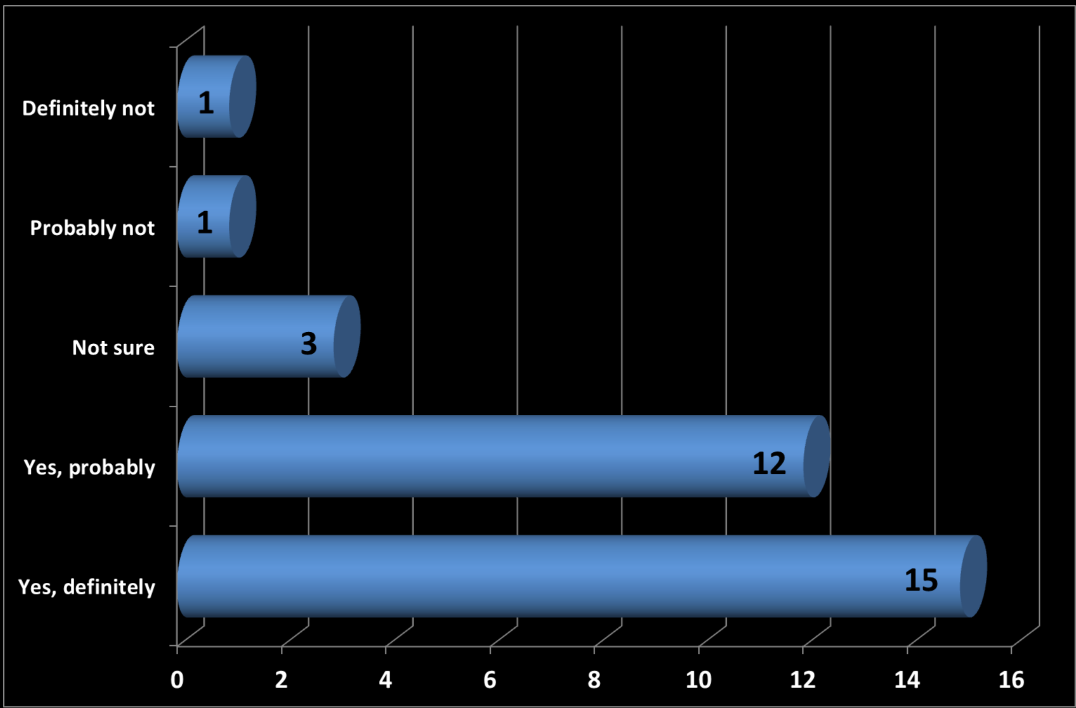
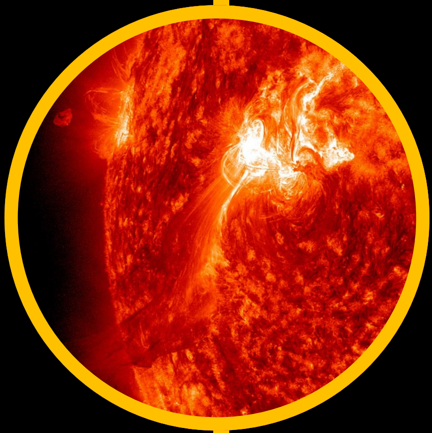


Figure 9: Student support for use of the RBL-OSTI model in future versions of the course. Numbers on the plot bars indicate overall totals for each category.

As can be seen in Figure 9, **twenty-seven respondents came out in favour of the RBL approach**, delivering either a weak or strong positive.

These data are best compared against the responses of Figure 7, in which students were asked to state their preference for either the RBL or single-project approach, or alternatively to state their neutralilty. Closer inspection of the raw responses for both questions reveals that the two students who preferred the single-project assessment model also delivered the two "probably/definitely not" replies of Figure 9. Both were from mathematics/computer science backgrounds. One of the neutral respondents in Figure 9 still indicated a preference for group rotations in the data of Figure 7. We further note that the one external guest student from Physics chose to respond with "No opinion/not sure" on the results for Figures 7 and 9; this should be borne in mind when interpreting these results.

Overall these findings present a very positive outlook on the future potential for rotation based learning approaches in either single-subject or interdisciplinary environments.



3. Lecture Content

Section Aims: Lecture slides and advice notes are expected to form a significant component of the initial OSTI materials release via the GitHub repository. Section 3 of the questionnaire addressed student attitudes to the lecture content, In the interests of refining the content for future OSTI instances.

Prospective OSTI course leaders may find these results useful if having to deliver selected lectures only, in circumstances where time restrictions prevent use of the full OSTI program.

Which of the daily lectures did you find most useful for developing your awareness of current and emerging research practices?

Lectures on licensing and publication proved most useful for "cultural awareness" education. Indeed the content of these two lectures was perhaps the easiest for the students to visualise as part of daily working practices, irrespective of an individual's subject area.

The more technological lectures covering data management plans and ver-

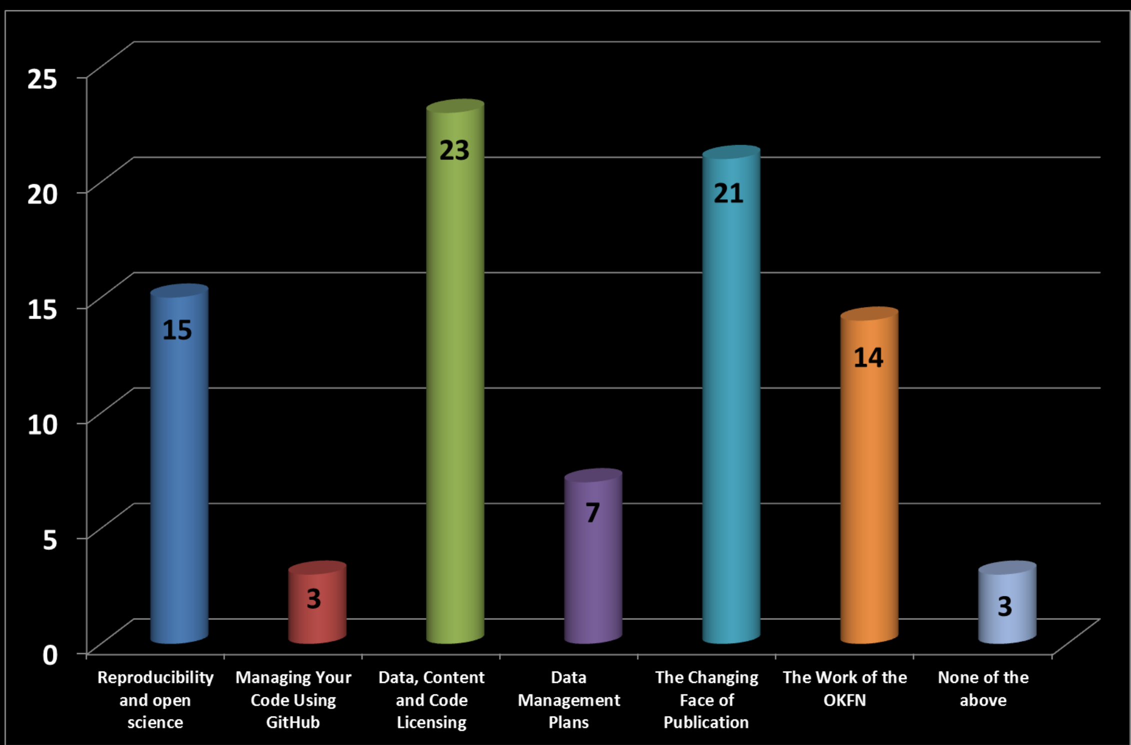


Figure 10: Numbers on the plot indicate overall totals for each category: respondents were able to select more than one lecture in their response.

sion control were not identified as contributing as much to scientific awareness; however, several students indicated that they would have preferred these topics to appear as practical learning sessions rather than lectures.

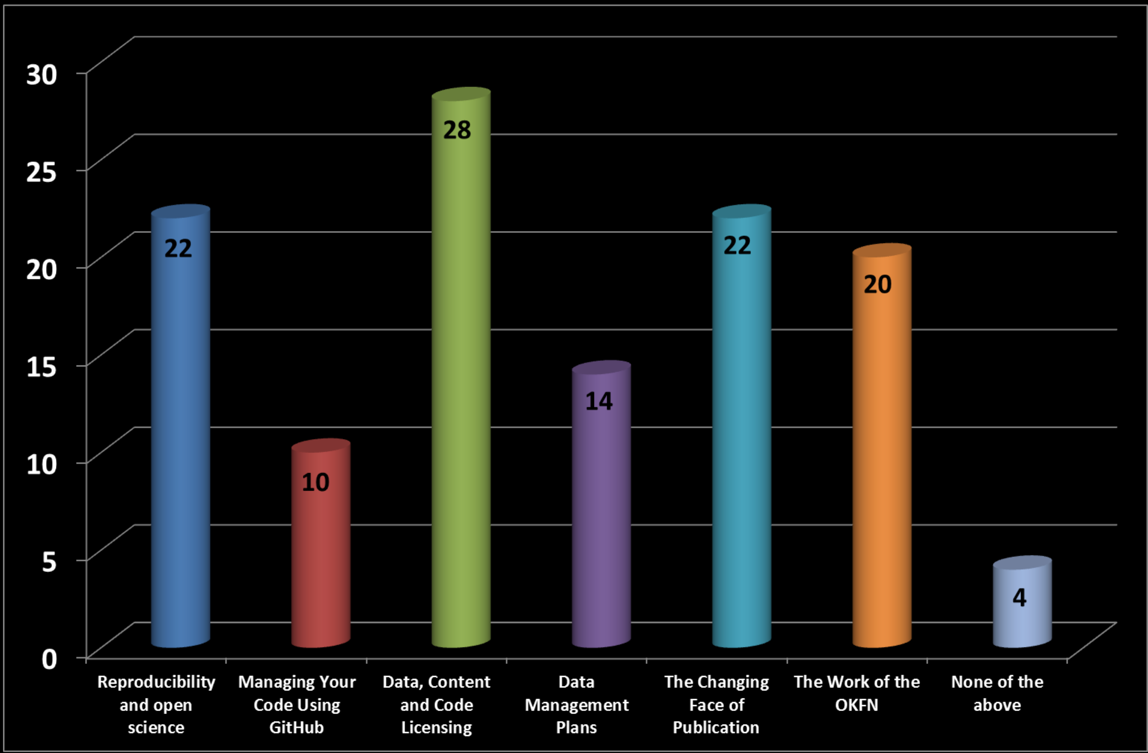


Figure 11: Numbers on the plot indicate overall totals for each category: respondents were able to select more than one lecture in their response.

Which lectures would you recommend we include in future versions of the course?

Many of the students suggested that the GitHub/version control lecture be delivered as a hands-on tutorial in the main work area in the future.

The findings depicted in Figure 11 suggest that the “big picture” themes such as licensing, publication and reproducibility have the most to contribute in their current lecture format and their delivery should be prioritised by course leaders. The enhancement lecture outlining the work of the Open Knowledge Foundation was one of the top performers and several students enjoyed the perspectives it provided.

Adaptation of the material on version control and data management for a practical learning session would perhaps be the best way to introduce these themes into an open science course. There is also a possibility that the student perception of GitHub was slightly tarnished by the technical problems with the GitHub for Windows interface during the course.

How did you find the level of information provided in the daily lectures?

All six respondents contributing to the endpoint categories were originally from either Physics or Mathematics disciplines. In general, however, the consensus was that lectures delivered an accessible level of detail without becoming overly simple.

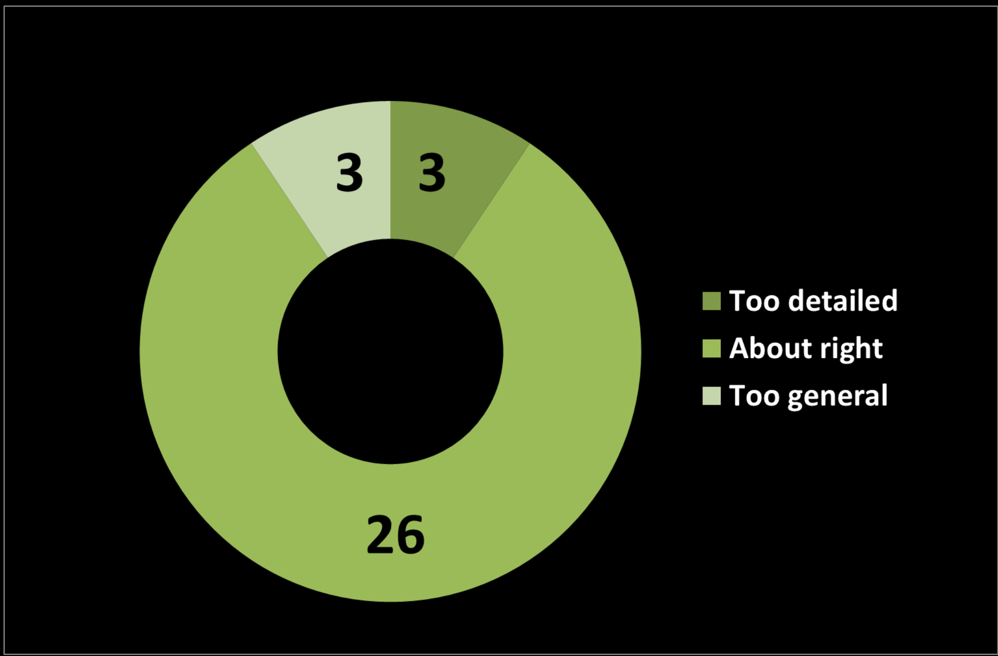


Figure 12: Students' assessment of information level in the daily lectures in Open Science. Numbers on the plot indicate overall totals for each category.

Consequently the first publically-available version of the OSTI lecture materials will maintain the existing level of depth.

4. Freeform Responses

The final part of the questionnaire provided two areas for freeform response and suggestions. These questions were not compulsory.

Are there any other subject areas in data handling, reproducibility or open science which you would liked to have seen included in the course?

A small number of students suggested additional areas they would like to see addressed in the course; one also commented that the coverage of subject areas was generally good. A mathematician on the course said that they could not suggest other areas to cover as they entered the OSTI course with minimal knowledge of the area.

Suggestions for additional topics in open data and open science included:

- Specific coverage of open source coding;
- Inclusion of a separate lecture or practical session on how to structure Matlab code;
- Practical aspects of data mining.

Do you have any other comments or suggestions for improvements and/or modifications to the course?

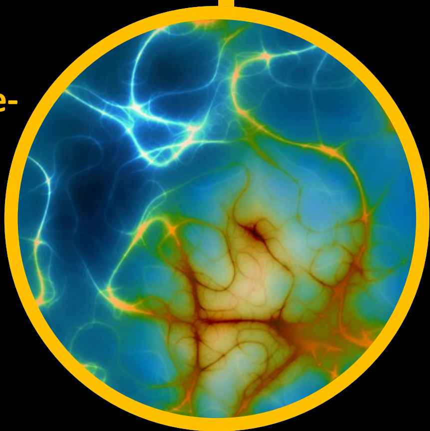
The final question drew a range of responses from the students; here we provide a summary of the main themes.

- **Funding:** A small number of respondents expressed a desire to know more about how science is funded, who makes these decisions, how different areas of research are prioritised and whether funding sources may create bias in the system.
- **Phase Lengths:** Four students indicated that they would have liked more time to work on their projects in order to deliver work of a higher standard (NB: This said, the research output during the pilot was excellent and the students appeared committed to perfecting their work). These comments support our earlier findings that a 4 to 5 day rotation phase might be beneficial, particularly for Phase 2.
- **Counter arguments:** Three students wanted greater discussion of the pros and cons of different approaches and/or the potential problems faced by Open methodologies.

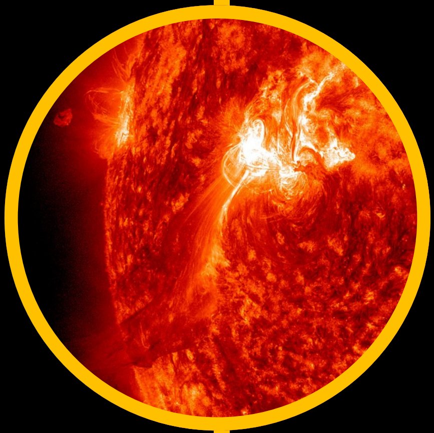
Several students used the freeform area to reiterate their enjoyment of the rotation based approach and/or lecture content.

Of particular note were the comments describing OSTI as "[amongst] **the best 10 days I've had at the DTC. Thanks to everyone who thought so hard about how to build this course."**

Another student also said that **"the projects...increased my confidence in interpreting and using scientific papers and I now feel that I am capable of doing really good research."**



Demonstrator Feedback



Tutor Questionnaire

A questionnaire was also communicated to the course demonstrators at the end of the course. Their responses are summarised below.

1. Tutors with previous experience of the DTC Matlab course

First we outline the feedback from **James Osborne, Associate Director of the Doctoral Training Centre**. Dr. Osborne has run the Matlab course at the DTC for several years previously and was present for Phases 1 and 2, in addition to the presentations on the final day of the course. He also observed many of the daily supervision sessions with the groups as the projects progressed.

He felt that ***"the overall standard of the work completed was higher and I feel the groups got further. Moreover, the students got experience working in two applications rather than one. I believe that the students also liked the change of project midway through and it was worth the extra organisation."***

Dr. Osborne also commented that ***"the open science component really added to the Matlab course"*** and that he would be happy for the same approach to be used again. His only improvement for future years would be to reappraise the version control implementation by avoiding the use of problematic software.

A second demonstrator, anonymous for the purposes of this Report, was present for the second phase of the OSTI rotations. The tutor in question is a postdoctoral researcher currently working as a Junior Research Fellow in Applied Mathematics and has several years' experience of demonstrating on the same course.

Contrasting the OSTI approach with previous years, he approved of the change to a rotation based structure and felt that it shifted the emphasis to writing more understandable code. He did however feel that this arrangement meant that students may have had to restrict their overall ambitions for a project in some cases. He also ***"thought teaching was very hands on and that the students benefited from this. The work load seemed manageable and there seemed to be enough time for students to get stuck into their projects."***

2. New tutors to DTC demonstration

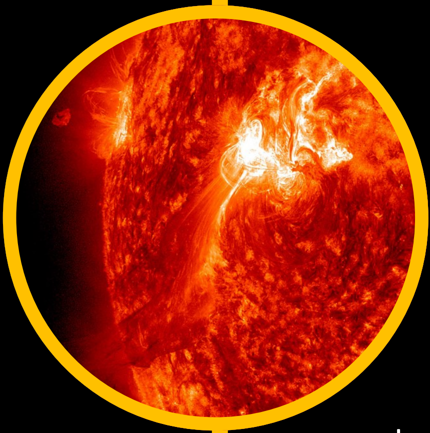
Our remaining two responses come from course demonstrators who were teaching on this Doctoral Training Centre course for the first time. Both were heavily involved in the auxiliary teaching in the main work area on a daily basis. They joined the course for both rotation phases and worked with each of the eight groups throughout the OSTI pilot.

David Robert Grimes is a postdoctoral researcher, originally trained in Physics and currently working on interdisciplinary projects involving tumour modelling. He identified the teaching methods as one of the successful aspects of the course and noted that, ***"working with students to explain the methods and mathematics behind [their] examples...went well".*** He would improve the course by providing demonstrators with earlier access to the programming content that preceded OSTI in the timetabling (NB: We note that this relates more to general infrastructure than to the OSTI-specific components of the course).

Mark Gilbert is a DPhil student in Applied Mathematics. When asked to highlight the aspects of the course that had gone well, he responded that teaching and content were useful—both to the students' understanding of Matlab and to the assessment phase of the course—and well paced.

He also felt that students seemed to find the problems very interesting and engaging. Mr. Gilbert commented that the timings of the rotation structure were well judged and that ***"the amount of time each group spent on the two projects seemed sensible; enough to be engaged with the problem, but not too much."*** He highlighted the technical problems with GitHub for Windows as an element that would need to be improved in future versions of the course.

Conclusions



Summary of Findings

The Open Science Training Initiative and its method of Rotation Based Learning provides a powerful means of educating students about reproducibility and the intrinsic value of a coherent research story. Equal emphasis upon, and incentivisation of, the producer and user roles during the research process have major roles to play in aiding the progress of open practice within academia. It falls to us as a research community to decide how best to address this issue in the coming years.

Students entering the course detailed in this study initially rated themselves as having minimal awareness of reproducibility and open science practices. Indeed, an informal show of hands in the first lecture suggested that only one student out of the cohort of forty-three already knew about open science on commencing the programme. Such sobering statistics as these indicate how far we have yet to go if unilateral openness and highly repeatable, repurposable research are to become the norm.

That 29 of 32 respondents to the exit questionnaire said they would be confident in implementing open practices in the future is a highly encouraging reflection on the OSTI approach. Such a result represents a substantial improvement on the level of student awareness at the start of the initiative. OSTI's successful use of hands-on learning for practical topics such as licensing and data management holds vast potential for driving large-scale uptake of these techniques.

Course leaders of future OSTI instances may wish to extend the rotation phase lengths, to provide more time for the students to develop their ideas and techniques. We also recommend that prospective OSTI leaders consider scheduling a debate at the end of their course, possibly in place of the group presentations. This would provide a forum for students to engage in discussion about current scientific practices and encourage them to form a balanced critique of the prevailing research culture from the very start of their scientific careers.

The OSTI pilot has showcased its strengths in modifying existing subject-specific provision in the sciences. The model is readily amenable to adaptation for use across the sciences at undergraduate, graduate and postdoctoral level and we highly recommend it for further application.

Future Directions

Following on from this post-pilot report, our GitHub repository will shortly be populated with downloadable course materials over the coming weeks. These will comprise: **three core lectures** in reproducibility, licensing and publication, to be prioritised in cases where timetabling space is limited; notes and slides for **practical workshops**; and additional **enhancement lectures** on further aspects of open culture, for use in longer courses where timetabling space is fairly unrestricted.

All materials (lecture slides, advice notes, handouts and exercises) will carry a CC-BY licence to enable modification, adaptation and reuse of the materials to fit the context, scale and scope of the host programme. Active use of these materials is encouraged; indeed our hope is that educators will “fork” the GitHub repository and grow the initiative over time by creating their own adaptations of the course. Interested educators may wish to refer to the official course website, www.opensciencetraining.com, for further details.

Acknowledgments

My warmest thanks to those individuals and organisations who have enabled and assisted with the founding of the Open Science Training Initiative.

Particular thanks are due to the **Open Knowledge Foundation**, whose funding in the form of a Panton Fellowship in 2012-13 provided the financial support to establish the OSTI scheme. A debt of gratitude also to the staff and students of the **Doctoral Training Centres at the University of Oxford**, whose participation in the study detailed herein has provided a wealth of insights and information on which we can begin to build in the future.

Sophie

Sophie Kershaw, May 2013

	Thurs 10/01	Fri 11/01	Mon 14/01	Tues 15/01	Weds 16/01	Thurs 17/01	Fri 18/01
9:30am - 10:30am	**Introductory Lecture: Reproducibility and Open Science**	Project Work	Project Work	Project Work	Project Work	Project Work	Project Work
10:30am - 12:30pm	Project Work	Project Work	Project Work	Project Work	Project Work	Project Work	Project Presentations
12:30pm - 1:00pm	**Version Control Using GitHub**	**Content, Code and Data Licensing**	12:30 **Data Management Plans** 12:45 Guest Lecture (Jun Zhao) "*Scientific Workflows*"	**The Changing Face of Publication**	Guest Lecture (Tom Dunton) "*Interfacing Matlab with Open Source Code*"	**The Work of the OKFN** (co-presented with Jenny Molloy)	Project Presentations
1:00pm - 2:00pm	*Lunch Break*	*Lunch Break*	*Lunch Break*	*Lunch Break*	*Lunch Break*	*Lunch Break*	*Lunch Break*
2:00pm - 2:30pm	Project Work	Project Work	Guest Lecture (Alex Fletcher) "*How to Write Mathematics*"	Project Work	Project Work	Project Work	Project Presentations
2:30pm - 5:30pm	Project Work	Project Work	Project Work	Project Work	Project Work	Project Work	Project Presentations

The following timetable shows the teaching schedule for the OSTI pilot initiative. The majority of the time is allocated to project work, shown in blue, during which time the students are expected to structure their own time. Lecture components are shown in pink, with official OSTI lectures listed in bold. Project presentations at the end of the course are shown in green.

Pilot Timetable

Textual References

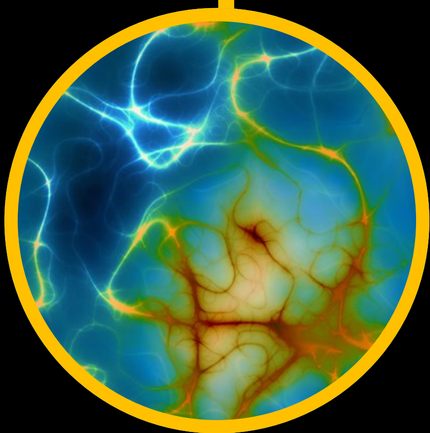
The following references are cited in the Motivation & Aims section of this Report. The first reference provides an extensive study into the research practices of UK students. The second is a commentary discussing the reproducibility issues facing science.

[1] **Researchers of Tomorrow: the research behaviour of Generation Y doctoral students.** JISC/British Library, June 2002

A PDF of the report is available from http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow.aspx

[2] **Drug development : Raise standards for preclinical cancer research**. C.G. Begley, L.M. Ellis; Nature 483, 531-33 (2012)

A PDF of the report is available via the Nature website, http://www.nature.com/nature/journal/v483/n7391/full/483531a.html



Licensing & Credits