

Cambridge Data Champions: Demystifying RDM through Collaboration

Dr Sacha Jones

Research Data Manager, Office of Scholarly Communication, Cambridge University Libraries

Dr Kiera McNeice

Research Data Manager, Cambridge University Press & Assessment

INTRODUCTION	1
ABOUT THE CAMBRIDGE DATA CHAMPIONS.....	1
COLLABORATION WITH CAMBRIDGE UNIVERSITY PRESS & ASSESSMENT.....	2
USE OF FUNDS	3
<i>Forums</i>	3
<i>Workshops</i>	5
ACHIEVEMENTS.....	6
CHALLENGES AND ADVICE.....	6

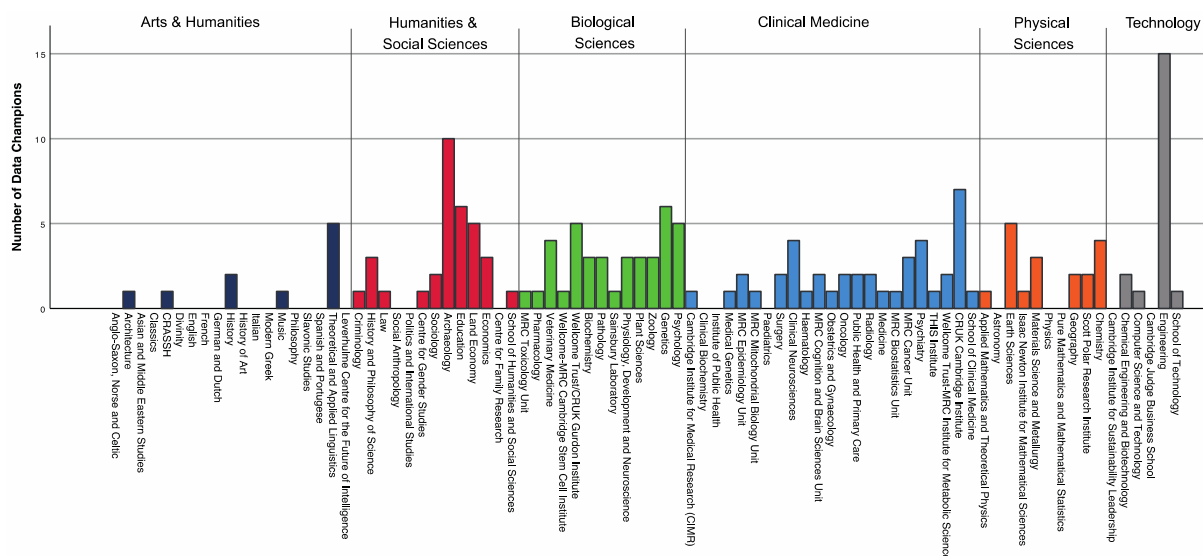
Introduction

Demystifying Research Data Management (RDM) has been a collaborative initiative between the University of Cambridge [Data Champions](#), Cambridge University Libraries' [RDM team](#) and Cambridge University Press & Assessment. The funds awarded to the RDM team were used to support forums and workshops, all aimed at exploring key topics relating to the management and sharing of research data. Part of the initiative has been to craft user-friendly RDM resources (or guides) for public dissemination and reuse by others. Several intersecting themes encapsulate the purpose of this joint endeavour. These are to contribute to improving support for researchers in matters relating to: effective research data management, open research, research reproducibility and transparency, research excellence and integrity, and publication of scholarly outputs.

About the Cambridge Data Champions

The Data Champions are advocates of good research data management practices. They have joined the Data Champion Programme, which is run by the RDM team at Cambridge University Libraries (CUL), to act as a local source of support and advice for research staff and students in their departments or institutes. The [Data Champion network](#) is diverse in terms of disciplinary coverage and data-related specialisms – this diversity and collective

expertise is one of the community’s greatest strengths. The network is comprised of PhD students, post-doctoral researchers, established academics, data and IT professionals and librarians, as well as a number in a variety of other roles. In 2022, we welcomed 70 new Data Champions to join the programme’s existing community, and by the end of the year the total size of the network comprised 179 individuals distributed across all six Schools as well as some non-School institutions (Cambridge University Libraries, Cambridge University Press & Assessment, Fitzwilliam Museum, UIS). The figure below illustrates the distribution of Data Champions across the University’s six Schools.



Number and distribution of Data Champions across the six Schools. An additional 21 Data Champions are located within various Non-School Institutions, such as Cambridge University Libraries, Cambridge University Press & Assessment, Fitzwilliam Museum, and UIS. (Data as of 21/09/2022.)

Collaboration with Cambridge University Press & Assessment

Funding awarded in this second round of the University Collaboration Budget (UCB) allowed the continuation of collaborative activities with Cambridge University Press & Assessment (CUP&A) that began in 2019 as part of support provided through UCB Round 1 for [Jointly Championing Good Research Data Management](#). The current initiative – [Cambridge Data Champions: Demystifying Research Data Management](#) – was led by Dr Curtis Sharma (formerly Research Data Coordinator) and Dr Sacha Jones (Research Data Manager) in the Office of Scholarly Communication at Cambridge University Libraries, in collaboration with Dr Kiera McNeice (Research Data Manager) at Cambridge University Press & Assessment (Scholarly Communications R&D). Together, we worked to generate forum and workshop content and devise group activities, also providing spaces for effective collaboration. Central to the success of this initiative has been partnering with the Data Champions who have participated in the forums and workshops, and contributed to the content and format of RDM resources. Cambridge University Press & Assessment have taken this content to design RDM guidance materials that are user-friendly, accessible, and open to use and reuse by all.

Use of funds

The RDM team at Cambridge University Libraries was awarded £2736 to fund this initiative. Specifically, these funds were used to cover the cost of five bimonthly Data Champion forums and two small group workshops. The following table provides a breakdown of costs and number of attendees at each event. These were held at various University sites in Cambridge. Both workshops were held in person and the forums were hybrid events except for the March 2023 forum, which was attended solely in person. The scheduled November 2022 forum was cancelled due to UCU strikes.

Type of event	Date	Location	In-person attendees	Online attendees	Total number of attendees	Cost (£)
Forum	May 2022	University Library	23	24	47	356.64
	July 2022	West Hub	15	28	43	234.40
	September 2022	Faculty of Law	17	20	37	322.06
	November 2022	<i>Forum cancelled due to UCU strikes</i>				
	January 2023	CUP&A	21	15	36	176.00
	March 2023	Engineering	35	– *	35	485.29
Workshop	July 2022	University Library	10	– *	10	150.36
	December 2022	CUP&A	9	– *	9	33.25

*not applicable, in-person attendance only.

Forums

The table below summarises the content of the five forums funded through UCB Round 2. These covered a variety of topics, and involved talks from Data Champions and invited speakers. Group brainstorming activities, general discussion sessions, and training and information sessions were also held. The latter aimed to provide Data Champions with knowledge and resources to assist them in their role as Data Champions for their departments or institutes.

May 2022 (hybrid)	<p>Licences in open research</p> <p>Talks:</p> <ul style="list-style-type: none"> • ‘Academic outputs and industry’, Dr Lyuba Bozhilova, Researcher, Department of Clinical Neurosciences and MRC Mitochondrial Biology Unit (Data Champion) • ‘Introduction to Creative Commons licences’, Claire Sewell, Research Support Librarian, School of Physical Sciences, University of Cambridge (Data Champion) • ‘Cambridge University Rights Retention Pilot’, Professor Stephen Eglen, Department of Applied Mathematics and Theoretical Physics (Data Champion alumnus) <p>Group activity: Metadata and metadata standards</p>
-----------------------------	--

<p>July 2022 (hybrid)</p>	<p>Sustainability in open research/open data</p> <p>Talks:</p> <ul style="list-style-type: none"> • Peter Sutton-Long, Repository Manager, University of Cambridge • <i>'Digital sustainability'</i>, Dr Megan Bushnell, Digital Humanities Content Producer, Fitzwilliam Museum (Data Champion) • Alastair Downie, Head of IT and Digital Transformation, School of Biological Sciences (Data Champion alumnus) • <i>"'Gone but not forgotten", tackling data longevity'</i>, Dr Amy Foreman, Research Project Manager/Information Scientist, The Gurdon Institute (Data Champion)
<p>September 2022 (hybrid)</p>	<p>Big data in research</p> <p>Talks:</p> <ul style="list-style-type: none"> • <i>'Mapping archaeological heritage: from paper to big data'</i>, Dr Rebecca Roberts, Research Associate, McDonald Institute for Archaeological Research and Fitzwilliam Museum (Data Champion) • <i>'What I wish I'd known: Some learnings from 3 years of mistakes with big data'</i>, Dr Julian Gilbey, Department of Applied Mathematics and Theoretical Physics (Data Champion) <p>Training session and questions:</p> <ul style="list-style-type: none"> • <i>'How to manage your research data, Part 1'</i>, Dr Bea Gini, Training Coordinator, Office of Scholarly Communication
<p>January 2023 (hybrid)</p>	<p>Challenges of working with qualitative data</p> <p>Talks:</p> <ul style="list-style-type: none"> • <i>'Summary of the Report of the Open Qualitative Research Working Group'</i>, Dr Meg Westbury, Academic Services Librarian, Human and Social Sciences (Data Champion) • <i>'My intimate interlocutors: more than just data'</i>, Lutfi Bin Othman, PhD student, Faculty of Music (Data Champion) • <i>'Rethinking interviewing as a migration research method'</i>, Chen Qu, Researcher, Department of Geography (Data Champion) <p>Group activity: Qualitative data – challenges, support, resources and training</p>
<p>March 2023</p>	<p>Data Archiving</p> <p>Talks:</p> <ul style="list-style-type: none"> • <i>'Choosing a data repository and using Apollo'</i>, Dr Sacha Jones, Research Data Manager, Cambridge University Libraries • <i>'Better late than never, but early is better: digital preservation to prevent data loss'</i>, Caylin Smith, Head of Digital Preservation, Cambridge University Libraries • <i>'The carbon footprint of computational science (and the data it relies on): how bad is it and what can we do about it?'</i>, Dr Loic Lanelongue, Cardiovascular Epidemiology Unit, University of Cambridge <p>Group discussion: Data archiving</p>

Workshops

The funds were also used to support two workshops, held in July and December 2022. The purpose of both workshops was to bring together two small groups of Data Champions to work with the RDM team and Kiera and colleagues at CUP&A to focus in depth on two distinct topics. The aim was to combine knowledge and ideas to produce content that would be compiled by collaborators at CUP&A to produce user-friendly guidance materials, or creative works, for broad dissemination and reuse. In doing so, duly acknowledging and recognising the Cambridge Data Champions for their contributions.

July 2022 workshop

Metadata and metadata standards

This workshop built on the results of a brainstorming session that was held in the May 2022 forum, where the Data Champions worked in groups to address these questions on metadata and metadata standards:

- What would you include in a definition of “metadata”?
- What do you think is the purpose of metadata?
- What do you think someone outside your research group, or even research area, might need to help them understand what your data are about and what your data folders and files contain?
- Why should we use metadata standards?
- Do you think a Metadata Management Plan would be useful? In what way? What might this look like?
- Creating good metadata can take time – what strategies or tools could we use to help?
- What questions would you like to be answered in a resource titled “Metadata and metadata standards”?

This was followed up by a small group workshop held in July, attended by four Data Champions (Simon Carrignon, Irene Fabry-Tehranchi, Kevin Symonds and Lucy Woolhouse) and representatives from CUL and CUP&A. The workshop focused on the format of a question- and answer-based resource on *Metadata and Metadata Standards* as well as potential questions to be addressed in the resource. The content was subsequently finalised and passed on to CUP&A for formatting and production.

December 2022 workshop

Selecting repositories

The workshop focused on how to find a suitable repository for sharing data or code, and what factors need to be considered when choosing. The workshop was attended by four Data Champions (Lutfi Bin Othman, Alexandra Gushurst-Moore, Tania Maxwell, Simon Stoddart) as well as CUL and CUP&A staff; however, it was preceded by a survey that was sent to all Data Champions with the aim to gain a broad impression of practices, pitfalls, decisions made and guidance sought when it comes to selecting a suitable data repository.

The workshop examined this topic in greater depth by asking the Data Champions present to discuss their thoughts on the following:

- Have you shared data in a repository? If ‘yes’, how/why did you find/choose that repository? If ‘no’, what's put you off?
- Have you ever struggled to find an appropriate place to share data? If ‘yes’, where did you look for advice/information?
- Can you summarise what are the main criteria that are important to you when choosing a repository?

To begin outlining a new resource on *Selecting Repositories* we first reviewed current information on this topic on the [RDM website](#) (CUL) and [author guidance](#) (CUP&A) to ask what is good about the support provided and what is missing. Lastly, we focused on the look and content of a new and user-friendly output.

Achievements

At the start of this initiative we set out to improve RDM practices, resources and tools at the University of Cambridge and beyond. Central to achieving this has been collaboration between Data Champions, the Research Data team at CUL and Cambridge University Press & Assessment. Together, we identified key areas of greatest need and started the production of concise, accessible and user-friendly resources to assist researchers and students to embed good RDM and data sharing in their research practice. While the guidance materials on *Metadata and Metadata Standards* and *Selecting Repositories* are still in production we expect these to be released under a Creative Commons (CC BY) licence when finalised later this year. One of the less tangible, but no less important, outcomes of this initiative has been expansion and consolidation of a relationship between the Research Data team at CUL and at CUP&A (Academic). In this we share a common purpose, which is to embed good RDM and data sharing in research practice, throughout the life cycle of the research project (not only at the point of publication) and across all disciplines (arts, humanities, social sciences, sciences) to produce outputs of the highest research integrity. This is where collaboration and conversation with the Data Champions, via events such as forums and workshops, is critically important; their 'on the ground' local knowledge, experience and networks within the University's departments and institutes is invaluable.

Challenges and advice

We started this initiative with an ambitious plan, perhaps overly ambitious, and achieving all that we set out to achieve with the time and staffing resources at hand has been the main challenge. More important is to celebrate what has been achieved already and a major benefit of this initiative, as well as the preceding Round 1 UCB Data Champions initiative, has been a sustained and ongoing collaboration between CUP&A, CUL and the Data Champions. Our advice to future applications is to nurture and grow such collaborative relationships while keeping common purposes and priorities in mind. For us, collectively, these have been centred broadly on excellent and open research practices and high quality research dissemination.