



Discipline-Specific Data Champions

The University of Cambridge

The University of Cambridge supports a volunteer led Data Champion Programme across its colleges and institutions, to assist researchers in navigating the often convoluted world of data management.

“The scale of data production is huge and continuing to grow rapidly. With Open Access becoming the norm, we can now access all different types of data from multiple sources across the globe. The level of data being produced is fast-paced, continually evolving and difficult to grapple with.”
- Clair, Data Champion

Research data and records can include anything from surveys to sound recordings, digital images to musical scores - so it's vital that researchers understand that this data needs to be cared for properly, throughout the research process, and how they can do that. It all starts with creating or collecting the data, and from there it needs to be organised, accessed, looked after and shared.

“The more open, explicit and carefully data is managed/packaged, the more the wider research community will benefit, through reusability, replicability and understanding of data.” - Frazer,
Data Champion

Those at the heart of research data management stress that this is not a ‘tick-box exercise’ but a vital part of the research itself. This is why the University of Cambridge went further than just providing tools and established their Data Champion Programme, which embeds open communication, care and respect within their support processes.

The programme recognises the need for a combination of tools and people to properly assist researchers from all areas when it comes to research data management.

- **By embedding Data Champions within departments, the peer-to-peer support guidance, and knowledge-sharing for discipline-specific data management embeds good practice within local research communities.**
- **By promoting good data management, the programme ultimately helps research advance faster while safe-guarding trust in the process.**
- **The flexible nature of the role allows Champions to contribute as much time as their schedules permit while developing their own knowledge, benefiting their personal and professional growth.**

The Data Champion Programme serves as a valuable model for other institutions interested in promoting good data management practices and research integrity. By facilitating communication and collaboration among researchers, it creates a supportive environment that encourages learning from each other, asking questions, seeking feedback, and upholding ethical standards collectively. Equipping researchers with essential skills and resources, the programme plays a pivotal role in upholding research integrity principles and good research practice.



What is the Data Champion Programme?

The University of Cambridge Data Champion Programme is managed by the Research Data team in the Office of Scholarly Communication at Cambridge University Libraries. It is in response to the ever-increasing need to improve knowledge and confidence with research data management processes for the benefit of the researcher, the subject and the institution.

The programme seeks to educate and support researchers across all disciplines with understanding and implementing effective data management within their work, given its increasing importance, the principles of Findable, Accessible, Interoperable, and Reusable (FAIR) research. The approach sees both the development of appropriate data management resources coupled with the opportunity for peer-to-peer support.

“There are two dimensions to supporting researchers through this process: first is making sure researchers are aware of the tools available to them and second is having a person there, embedded within the environment to help guide them. Peer support is vital to creating a shift in mentality and culture of good research data management.” - Alex, Data Champion

The programme is delivered by a network of volunteers, serving as “Data Champions”, from different schools and departments across the University of Cambridge as well as affiliated institutes. There is also the option for collaborations outside the University.

“We really impart expertise and advocate for best practices. We’re there to offer on the ground support and guidance” - Clair, Data Champion

The volunteers serve as “Data Champions”, who support initiatives and offer ground support and guidance to their local research community on how to handle research data according to discipline-specific best practice. The role allows Champions to be as active within the programme as their schedules allow, making it more accessible, building a greater and increasingly diverse team.

As part of the programme, Data Champions:

Deliver workshops and training at their departments/institutions and the public

Offer one-to-one mentoring

Promote/participate in development of data management tools

Participate in open research studies

Carry out activities to support good data management initiatives

The University already has a wealth of policies, frameworks and guidelines available on their Data Management website, however as a researcher trying to navigate the various options and processes on your own, it is reportedly overwhelming - particularly when you’re working in a specific research area and need discipline-specific guidance. The presence of a Data Champion within departments has been seen as a life-saver for many researchers.



“A lot of researchers, but also professional services staff, feel that research data management is scary,” says Alex, a Data Champion at the University of Cambridge. “We act as the glue...we take information about best practices and share them with colleagues who find research data management intimidating.”

When starting new projects, it is common for researchers to underestimate the effort and tools needed to consistently manage their data properly, in order to maintain the projects' integrity. Providing the opportunity to speak with a local Data Champion and ask questions helps researchers find the management processes that best fit their research needs, and saves them a lot of time and confusion in the process. Individuals are able to feel much more confident about data management and their trusting in their research.

This approach to supporting data management stands in contrast to conventional methods employed by most institutions, where a central office handles data management education but lacks the capacity or resources to address discipline-specific areas or daily research challenges.

The overall mission of the Champions is to equip researchers with the tools to properly care for and disseminate their data in a way that upholds research integrity principles and good research data management practices, maintaining the integrity of the research, the data and the institution, for societal benefit. It brings people together to find practical solutions to challenges and share information and knowledge.

The programme aims to promote research Integrity within data management from a grass-roots level – sharing knowledge from the inside out, fostering a culture of effective communication and allowing individuals to develop solutions collaboratively, across career stages and subject boundaries.



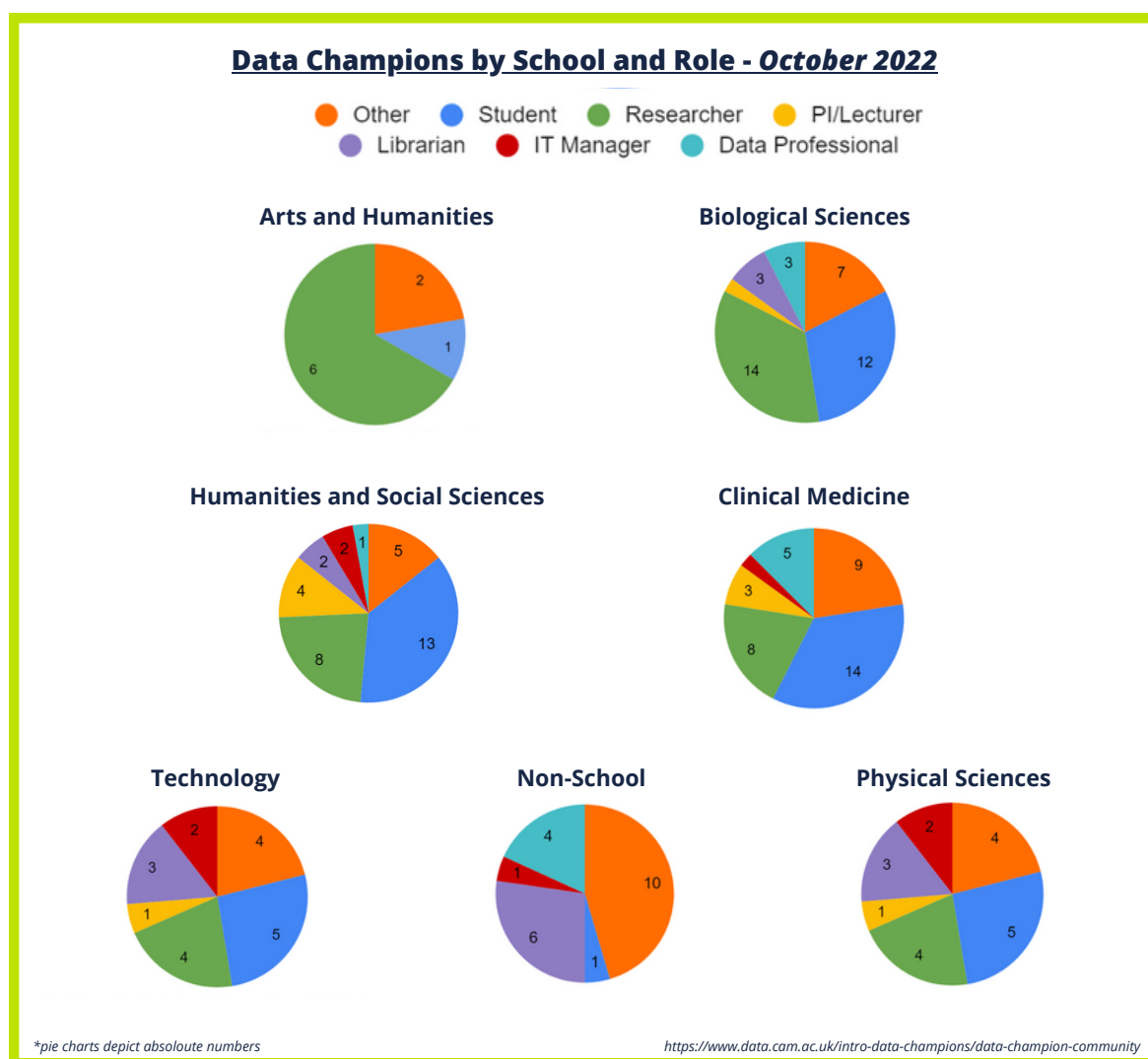
Who can become a Data Champion?

Any member of the University of Cambridge, who has an interest in best practices for managing and processing research data, can be a Data Champion. There are currently 180 Data Champions at the University, covering practically every academic and academic-related department or organisation.

Data Champions are not necessarily experts in data management. They are individuals who recognise the importance of data management in their field and want to learn more about how to improve their practices to then pass on this knowledge to their department and peers.

The Champions communicate via bimonthly Data Champion Forums, where they exchange experiences and expertise, share accomplishments, and provide further advice and support to one another.

The diverse uptake of volunteers as part of the Data Champion programme indicates a growing concern and curiosity surrounding data management across disciplines. The charts below display the range of roles of the Data Champions across various departments/schools within the University of Cambridge (the numbers represent the number of Champions of each role within each School).



It is important for the University to make sure the Data Champions cover a diverse range of backgrounds, to make sure all voices are heard, and all needs are supported under the programme.



What is the benefit of having a Data Champion in your department / school?

Clair, a Data Champion, emphasises the human value Data Champions add to their individual departments, *“when you have somebody who is a colleague or peer, researchers always take more notice of that person than just another member of staff or policy.”* The presence of Champions within various departments and colleges also allows a dynamic, engaged and informed community to blossom. By having relatable Champions embedded directly into the local research environment, volunteers encourage grassroots-level change through direct engagement and discipline-relevant examples.

Clair points out that running activities connected to research data management *“improves awareness of data management throughout the university and institutions”*.

“The embedded nature of the Data Champions is important because it allows for people to be able to relate to the general principles of research integrity and data management. Researchers can come to us with their specific questions and have conversations with Champions that have a much more detailed understanding of their needs and can explain processes in detail.” - Lucy, Data Champion

What is the benefit to the Champion, for volunteering in the programme?

Being part of the Data Champion community allows for collaboration and information sharing, maintaining a culture of integrity related to data management which ultimately benefits themselves as well as those they are supporting. It sets the stage for creating positive change based around behavioural changes, as opposed to simple redirection towards guides and policies.

The Data Champions have recognised the follows benefits following from volunteering within the programme:

- Free training to learn how to better manage your own data and improve your skills
- Grow your networks within and beyond the University of Cambridge
- Collaborate - Work with other Data Champions and the University RDM team
- Increase your impact by making a broader contribution to research at the University
- Increase your visibility and the visibility of your work within the University
- Enhance your CV

Frazer, a Data Champion, commented that *“participation in the programme offers a behind-the-scenes perspective on data archiving, copyright concerns, and protocols”*.

Furthermore, advocating for sound data management is just one aspect of becoming a Data Champion. In the words of one anonymous Champion, *“I like to share my expertise but also learn new things about it...it's also collaborative.”*

A range of Data Champions from different fields and departments within the University of Cambridge have been interviewed to get a better understanding of how and why they became a Data Champion, and the benefits and impacts they have recognised as result. There has been a particular focus on data management for the arts and humanities, as this is a commonly overlooked, yet prevalent, topic for research integrity.



Stories from Data Champions

Alex - Fitzwilliam Museum

Alex finds tremendous fulfilment in her position as a Data Champion, offering guidance on open research and open data regulations and procedures. She directs her students to resources and services available throughout Cambridge. Alex works with other Champions to present talks, demos, and workshops on tools and programmes for managing research data, increasing awareness of the value of data management throughout the university and other institutions.

"After I started the Data Champions program, I started a weekly Team Research Bulletin, that would include data and information that had been shared with me by other data Champions.

The weekly Research Team Bulletin was an email that was circulated every Thursday morning to subscribers of an internal research list. Members of this list include Fitzwilliam Museum staff on research contracts, our collections management team, and other key research collaborators (namely the Faculty Manager and Chief Secretary of the Art History and Architecture department).

It included various snippets of information, including: open funding calls, research events, research culture and community info, and data management information and guidance. One bulletin every month was devoted to highlighting open access repository requirements, to ensure the maximum number of research outputs are REF eligible and that UKRI-funded projects are meeting their requirements.

Below is an example of one of my open access focused bulletins. Drawing attention to open access reminders and updates is particularly important for guiding researchers through the data management process. Updated processes and regulations can be easily missed. Developing a succinct communications channel helps everyone keep up to date whilst promoting the implementation of processes that ultimately encourage increased transparency.

Hello,

Mornings are getting frosty and soon we will all be in hibernation for the Christmas period! I'd love to tell you that this week's Research Bulletin will be a light one, but alas... the good bits are at the bottom! (The important bits are at the top).

This bulletin can be read on the [intranet](#).

Open Access Reminder

- The time has come for my monthly reminder to you to ensure that you are putting your publications into Symplectic Elements and supplying an Open Access copy where needed! If you are unsure about what Open Access is, how to use Symplectic Elements, or what the current rules are, [you can consult our intranet page](#). You can also reach out to me or go straight to the central OA team: info@openaccess.cam.ac.uk if you have specific queries.
 - Happily, [UKRI have published a succinct round-up of what the current rules and requirements surrounding OA are](#) (specific to projects funded by UKRI). This gives a good overview of what you should be doing with regard to OA.
 - It is v. important that you deposit an Open Access copy of your manuscript into Symplectic Elements **within 3 months of acceptance**. If you do not do this, your paper will not be eligible for the next REF (you may also be in breach of funder requirements, though this is rarer).
 - Over the past month, I've been tidying up the records already in Apollo (Symplectic Elements public-facing site). Many, many thanks to everyone that has been helping me to ensure these records are all accurate.
 - Over the next month, I will be cross analysing the publications that have been reported to Syndicate with what is in Symplectic Elements. In theory, the records should be identical. In practice, they aren't. If I discover anything awry, I'll drop you an email and help you get it fixed.
-



There were a few reasons as to why I decided to create this bulletin. I often found that emailing staff about a number of research-related issues was not easily digestible, so condensing this knowledge into a single weekly digest made the information more manageable to access.

Secondly, there are increasingly strict rules and regulations around research data management, and I wanted to ensure that researchers had the tools to put their best foot forward – by giving them a monthly reminder to check their publications in our institutional repository (Apollo), I felt that I could improve a bit of the burden on them to remember that administrative responsibility. Overall, this actually adds a level of accountability for the researchers – they have access to updated information and it's now up to them to bring that on board and take responsibility for maintaining good research data management practices.

Lastly, the bulletin was a way of bringing people together: everyone in our division received this bulletin and, in an organisation of roughly 250 people it can be difficult to convey departmental identity within a large institutional structure. This bulletin served as one way to facilitate communication and share information about what the Data Champions were doing, as well as our Research Workshops and website and intranet pages! By adding a personal touch to the start of the bulletins this also helped to bring some warmth to the lists of resources attached, again just adding that human touch and being that friendly person on the other side of the screen who is there to answer any questions.

I really enjoyed creating these bulletins – it was helpful for me, as the Research and Impact Coordinator, to think big-picture about our research and its relationship to the institution and wider sector. I had many colleagues report that they found the bulletin really helpful:

“Can I just say your bulletins are amazing! Thank you.”

“Thanks for you very encouraging and supportive email- I love open access too!!”

“I love the opening of this email... really encouraging to read! (And obviously great content too!!)”

Colleagues would often reply to the bulletin and ask me questions or highlight various topics that I didn't know about. In that regard it was a great opportunity for knowledge collection as well as dissemination. Somebody has now been tasked with getting all that important information out to those that need it most!”



Stories from Data Champions

Lucy - Department of Genetics and Cambridge University Libraries

Given that the role of a Champion is voluntary, it may appear difficult to carry out so many activities due to time constraints, but Champions like Lucy find it to be immensely fulfilling. Every term, she conducts a class on research data management and offers one-on-one assistance to individuals who require it. She also aids researchers and students. She has been known as the go-to expert in her department thanks to her efforts, and researchers have even asked her to review a data management plan.

Lucy worked on a collaborative project for the University of Cambridge Press to produce a learning resource about Metadata and Metadata Standards. This project exemplifies how activities carried out by Champions integrate and support research integrity principles at various institutional levels.

"In the summer of 2022, as part of a joint project between Cambridge University Press & Assessment (CUP&A) and the Data Champions Programme, I participated in a workshop which aimed to produce a learning resource about Metadata¹ and Metadata Standards. This followed on from a discussion at a previous Data Champions Forum on the subject, investigating what kinds of information we thought were important when it comes to metadata, why metadata standards might be useful and how we might go about creating standards that align with best practices and principles.

After this discussion, I volunteered to be involved in a collaborative workshop along with two other Data Champions, the Research Data Coordinator and three members of CUP&A staff. I thought it was important to be part of this project as datasets are virtually useless without accompanying metadata, even if you share them openly. However, it's really hard to know exactly what information to include when processing or depositing your data to ensure they're reusable and reproducible. So, coming up with a standard set of guidelines for the metadata that should be produced to accompany datasets seemed like a really worthwhile endeavour.

The workshop itself was highly collaborative – each Data Champion worked with a member of the CUP&A team to discuss how we use data and what kind of metadata we thought was needed to make datasets truly reusable. As a librarian, I'm aware of the importance of good metadata for discovery purposes, and so I provided that perspective. Taking part in such a collaborative workshop that encouraged open and honest communication with various stakeholders demonstrated a dedication to fostering ethical research practices, ensuring the usability and reproducibility of datasets becomes engrained within their processes.

We discussed the need for case studies to give examples of good metadata, and what level of detail we felt was needed to achieve a good standard. We also wondered if this information should just be included in a normal Data Management Plan, which would increase transparency and responsibility for this process. We also flagged the difficulty of creating standards when data differs vastly by subject discipline, highlighting the importance of collaboration, communication and accountability between colleges and departments.

However, by the end we'd agreed that some further guidance for researchers would be helpful, so CUP&A took away our discussion points with the aim of creating a document to move forward. This document would acknowledge the contributions of the Data Champions and form part of CUP&A's guidance for authors. Whilst the document is yet to be produced, the commitment to transparency, responsibility, and accountability in providing researchers with necessary guidance exemplifies the research integrity principles upheld by both CUP&A and the Data Champions Programme."

¹ - Metadata: information that describes other information in order to help you understand or use it (<https://www.oxfordlearnersdictionaries.com/definition/english/metadata?q=metadata>)



Stories from Data Champions

Frazer – Scott Polar Research Institute

Frazer finds tremendous fulfilment in his position as a Data Champion, offering guidance on open research and open data regulations and procedures. Frazer directs his students colleagues to resources and services available throughout Cambridge and spends much of his time volunteering on additional data management projects that are important for the progression of research within his discipline:

“One of the projects in progress that I’m currently working on is the digital preservation of Scott Polar Research Institute’s the Arctic airborne radio-echo sounding datasets. This is a data legacy project where we are volunteering our time to archive and store historical data collected from over various regions of the high Arctic over the last 30 years.

In many cases the data collected represent the first and only measurements of precisely how was the first to measure and help us understand how thick the glaciers and ice caps were are in those variousthese highly remote locations regions. A place like tThe high Arctic is an extremely hostile environment so gathering that data was a great challenge for the scientists involved. Several high airborne flights traversed over the ice and imaged measured the thickness using radar imaging techniquesradio detection and radio echo sounding. This created a huge amount of data critical that was crucial for understanding how climate change is impacting the planet and in particular the polar regions.

Much Since collection, much of this historical data is has been stored on floppy disks, CDs and hard drives. , and a lot of these are just sitting underneath people’s desks or stashed away in boxes gathering dust. This data is incredibly valuable because in polar research, historical data can be used as a comparison to against current satellite and other geophysical observationsreadings. By comparing with this datathis historical data with modern observations, we can measure just how much the glaciers have melted and how the landscapes have changed. The issue of course is that back when this data was collected there was no such thing as a Dropbox, Google Drive or a Data Champion! Data stored on such old technology is liable to degradation (and potentially permanent loss!), so preserving it digitally is an important undertaking huge challenge.

As a Data Champion I had have had a lot of discussions about what we shouldto archive, what could be in a made freely or open-accesssly accessible and what couldn’t be made accessiblemanner. As a Data Champion I advise on best practices relating to cleaning the data, presenting the data and even how metadata associated with the datasets themselves are stored.

There’s a lot of back and forth between myself, the principal investigator and the project researchers who were employed back then to collect and process the data where we discuss these sorts of challenges and make decisions.

All of my work on this project is voluntary, I set aside time to address some of these data challenges but also work on the physical archiving itself. What you have to imagine for the physical archival side of the project is quite literally a process of sitting with a stack of CD’s and a CD player and combing through lots of raw data and associated files. We have to take that raw data and convert it into a format that is more compatible with today’s standards. It’s a time-consuming challenge, of course and it does take time, but it is an incredibly valuable and important one nonetheless - not just for long term storge, but for the progression of open climate science.



It really goes back to research integrity principles. *By making this data open access it promotes transparency and fosters communication and honesty between researchers. Sharing data becomes a self-sustaining cycle of reusability and recyclability. It does away with this old mentality of “these are my data, and they can collect dust under my desk until I retire” - science doesn’t advance effectively that way, people need to be using available data as much as possible. The fact that data these records are open for anyone to use encourages people to interact with the data, and even though it’s historical data, you can find important new insights.*

Many scientists and other interested users We’ve had lots of people who have contacted us about using who have used these historical data in their own research. That research ultimately feeds its way converts itself into new scientific publications and research proposals, it and in turn breeds good science precisely because we made the data findable, accessible, interpretable and reusable (FAIR). Making the information accessible, keeping it open and, ensuring that we continue to follow best practices allows these data to live on and be useful to the wider scientific community. There’s trust built into that process. It becomes a sustained way to facilitate communication,; it’s done with a degree of care, respect and rigour that, at the end of the day, helps science progress.”

We’ve had lots of people who have contacted us who have used these historical data in their own research. That research feeds its way into new publications and proposals, it breeds good science precisely because we made the data findable, accessible, interpretable and reusable (FAIR). Making the information accessible, keeping it open, ensuring that we continue to follow best practices allows these data to live on and be useful to the wider scientific community. There’s trust built into that process. It becomes a sustained way to facilitate communication, it’s done with a degree of care, respect and rigour that, at the end of the day, helps science progress.”

You can find more information about the project and the papers produced containing the Arctic RES survey data here: <https://www.repository.cam.ac.uk/collections/e8a2940e-9c65-4995-a448-ce79340981de>



Stories from Data Champions

Clair – Chemistry Department

Clair, a Data Champion and Librarian, was driven to become a Champion because she wanted to make sure she could help people complete their work. Working to guide people towards relevant resources, help review data management plans and taking anyone who needs assistance through the intricacies of file storage and organisation, Clair is on hand to support. With a wealth of experience having also worked for the Office of Scholarly Communication, Clair uses her networks, creativity and passion for data management to drive projects with university wide impact.

"The Research Data Team has established a convenient 'advocacy' resource in a shared Google Drive folder for Data Champions. This resource features captivating cartoon designs, crafted by a talented volunteer who is also an artist. These designs can be used on postcards, posters and during training sessions, adding vibrancy and impact to various communication materials.

One of the main ideas behind the cartoons is that the Research Data Team wanted to create an 'advocacy' resource in the Data Champions' Google Drive. Data Champions can then use them in posters, training sessions etc. that they would design themselves. The first use for the cartoons was on postcards to promote the Data Champions Programme and the RDM services that the Research Data team offer.

We tried very hard to be as diverse as possible, in order to represent the Data Champions inclusively. Our artist's inspiration for the characters has tended to come from real-life people she has encountered.



I devised a final set of scenarios to be illustrated, including just three boxes in a strip for ease, keeping the scenarios concise and imagining ones that would be immediately understood.



Scenario 6			
Stakeholder	Objection	Objection	Response
Research Support Staff	"How much does data storage cost?" "How do I back up my data?" [emails/phone calls from researchers in speech bubbles]	The researchers I support need answers but it will take me too long to learn research data management principles. [stakeholder looking despondent]	Have a look at the FAQs on the Research Data Management Website for quick answers. [DC, screen showing name of website]
Scenario 7			
Stakeholder	Objection	Response	Response
VC	Managing research data costs too much money and involves too much red tape. [VC in ceremonial dress in a function room, with food/drink in hand]	Good research data management means ensuring that data is handled ethically and stored securely. [DC with food/drink in hand]	The university's reputation could be damaged by a significant loss or breach of sensitive data. [DC imagining bad newspaper headlines]
Scenario 8			
Stakeholder	Objection	Response	Response
Postgraduate student - HASS	Check out this course on research data management! [DC shows colleague the email/poster/flyer advert]	I don't have data! [stakeholder shrugging]	You probably do - data can mean anything you produce. Even if you have no data you still need to manage your files [DC]
Become a Data Champion!			
	Data Champion	Stakeholders	Data Champion
	Are you interested in looking after and sharing data properly?	[Two stakeholders talking to each other] Do you want to be the local expert in research data management in the department?	Become a Data Champion!

I used the characters and scenarios to update the Research Data postcard that the team had been giving out at training sessions and other events. This provides links to sources of help and guidance on sharing research data, and to the Research Data Management website and social media accounts. It would now also include a link to the Data Champions programme.

Our cartoons had a nice 'hand-made' feel and our prototype postcards received positive feedback.

The final designs were ready for the beginning of the new academic year when we knew Data Champions would be inducting new students and staff and undertaking RDM training in their institutions. Data Champions were able to request printed copies of each of the designs.

The initial print run was 100 of each design and we needed to make a further print run of 50 each for "Check out this course on research data management" and the "Data Champion Wanted" designs, as they proved to be particularly popular for use at induction and training events.

The Research Data team now distributes the postcards at all RDM training sessions and have proved more popular than other versions without the added vibrant cartoons."



Stories from Data Champions

Lucy - Department of Genetics and Cambridge University Libraries

"I ran a session on open data and research data management aimed at graduate students and early career researchers. About 40 out of 100 members of the department attended the event, with some watching live and others in the lecture hall.

We really focused on the basics of research data management, covering areas like backing up data, file organisation, and available storage options. We also looked at how do you actually share your data and what do you need in order to do that? This is where we found a definite knowledge gap in terms of the specific requirements people had related to their unique research. Researchers are fully aware of how complex this process is, it's so dependent on funding, how data has been stored in the past, whether repositories are safe to use and are accredited – and it's changing all the time. I completely understand how researchers find it so difficult to choose between all the management options.

After the event, we got some really positive feedback. I made sure to share the slides with those who attended and the wider department as well. I had quite a few emails saying they found the slides and session useful and even had people come up to me a few days after the event with specific questions or asking for a review of a data management plan."

When it comes to the intricacies of different types of data, Lucy also highlighted how showing and not telling can help influence and guide researchers:

"Everyone has unique pieces of research, and everyone's data is different. I did a talk about the general principles and basics of data management before bringing in a volunteer who presented a specific case study on the way he had approached the research data management process. He was working with around two terabytes of movie data he had made from these tiny microscopy images. He ended up compressing some of the data to make it easier to store. Some data went into tape storage at the University of Dundee, and it was shared through a bio image archive. Afterwards, people made comments about how they really appreciated seeing imaging data being used as a case study because it's hardly ever used in a general sense and a lot found it useful, even if they weren't working with imaging data specifically."

Champions also offer one-on-one support to help researchers, taking a hands-on approach which is crucial for shaping the local research ecosystem. Clair talks about the importance of this type of support, *"it's all about saving people time and making sure they're not having to reinvent the wheel when going through the research data management process"*. Lucy also notes the level of trust that comes with this guidance - *"peer-to-peer learning is often so much stronger a teaching tool than just someone telling you a piece of information. Having that one-to-one option in place is a really strong learning tool when approaching this information"*.

Often it is by working together that Champions drive change and spark a culture shift. As one champion from the Arts puts it: *"It's a community where people can share information and offer solutions."* As another champion notes, *"informal, friendly, open conversations about practices is generally a positive."* By engaging directly with colleagues and researchers, Champions demonstrate the importance of responsible data management and show just how much impact it can have on research outcomes.

It is ultimately undertaking the data management process itself that will define the success of the Data Champions guidance. To achieve this goal, the Champions provide clear directions and relevant resources to university members navigating the five stages of data management: creation, organisation, access, upkeep and sharing. Each stage is grounded in research integrity principles that the Champions consistently emphasise and elaborate upon.



How does Data Management act as an important tool in supporting and improving research integrity?

Creating & Collecting Data

The process of creating and collecting data is intricate and varied, and it can take many different forms, each with its own special opportunities and challenges. Data are generated in a myriad of ways across many different sectors and disciplines, from the meticulous observation of natural events to the production of elaborate computer models. In the context of arts and humanities, one may also often use the term ‘records’ rather than ‘data’.

In the sciences, collecting data frequently entails careful testing and observation. Experiments are conducted by scientists to test theories, gather information on the physical and chemical characteristics of materials, or monitor an organism's activity in a controlled environment. Such information gathering may involve the use of advanced tools like telescopes, microscopes, and particle accelerators.

In the arts and humanities, creating data frequently entails deciphering and analysing intricate cultural artefacts. To find patterns and themes that illuminate human history and culture, researchers examine historical writings, literature, art, and other forms of cultural expression. To gather information on social interactions and human behaviour, they might also employ surveys, interviews, and other social science methodologies.

Data production adopts a more innovative and experimental style in the arts. In addition to using more conventional methods like painting and sculpture, artists can also design immersive, interactive installations that engage the senses. Intricate digital worlds and experiences that push the boundaries of what is possible may also be created through digital tools like 3D modelling software or virtual reality platforms.

Data creation is a crucial step in the research process regardless of the subject matter or discipline. It calls for originality, knowledge, and a profound comprehension of the complicated systems and processes being researched. Researchers and practitioners from a variety of sectors might get novel insights and push the bounds of what is possible by creating data in novel and perceptive ways.

Organising Data





By keeping data and records organised, researchers – irrespective of discipline – can more easily access and analyse findings, saving time and increasing productivity. *“There’s no point reinventing the wheel,”* notes Clair, *“It’s not just about complying with standards, it’s about caring for and organising your data in a way that will help you and others further down the line”.*

This also ensures that researchers can conduct their work ethically, legally and professionally, whilst supporting a research environment underpinned by a culture of integrity based on good governance and best practice. Examples of good data and records management practices include:

Consistent and informative file naming

Organised folder structures

Detailed documentation and metadata

Appropriate reference management, clear and consistent file naming and organisation are crucial in all disciplines. This implies that all files are stored in a logical and hierarchical directory structure and given descriptive and meaningful names that represent their content. In doing so, researchers may easily locate the files they require, ensuring that their judgements are based on accurate information. When sharing files with others, this type of arrangement is useful because it makes it simpler to monitor and handle any modifications.

The methodologies, processes, and source materials used by researchers should be fully documented. Depending on the discipline, such documentation could be captured in a Project Wiki or Database that acts as a central repository. These should also include relevant metadata which provide details about specific files, and must be stored consistently. Such information increases interpretability and makes it simpler for researchers to comprehend and analyse the data.

By putting the above data management principles into practice, the data are not only accurate and reliable but also open to replication and can be built upon by other researchers, fostering a culture of openness and transparency within the research community.

“In the Genetics Department, a lot of researchers work with fruit flies. Data is created and organised from a diverse set of sources. This might include something like genetic or phenotypic data which can be organised in special databases and specialised software before being stored and analysed. Once that data are organised and labelled correctly, researchers can then start to work on their analysis.

Just making sure that datasets are organised, metadata are included, files are named properly, all helps to keep the findings accurate and reliable. If the data are organised well, others can understand the processes the researcher took. When they eventually get to a point of uploading the data to repositories and sharing it with others, comprehensive data organisation improves data exchange and facilitates collaboration.” - Lucy, Data Champion

By putting the above data management principles into practice, the data are not only accurate and reliable but also open to replication and can be built upon by other researchers, fostering a culture of openness and transparency within the research community.



Organising data and records in the humanities can come in a variety of forms including cataloguing, archiving, reference management and survey response filing. Implementing good organisational practices allows researchers to keep a precise record of their sources, which helps to make their findings and thought process transparent to users of their research.

Cataloguing large databases can often be challenging, particularly when dealing with multiple data streams that span everything from statistics to images and audio recordings. Alex from the Fitzwilliam Museum details how data organisation plays a role in her area of work:

“Data organisation at the Fitzwilliam Museum involves things like collecting and cataloguing information and objects. We have a huge collection of objects, art and other artefacts that we need to systematically catalogue. For each item we have, records need to be kept containing relevant information, dates, notes about the artist, materials and more. By organising all of this data, staff can keep information on objects up to date and access information when needed.

Because we’re public facing, we also make sure to organise any information we have on exhibitions, educational content and research projects. It’s really important to ensure all of the different types of data we work with are properly organised following good research management practices such as file organisation and chronological ordering. A lot of what we do relies on the fact that we know what objects and information we have so that we can maintain the integrity of the objects and curate meaningful exhibitions.” - Alex, Data Champion

Therefore, data and records organisation is crucial for departments and researchers in the humanities as well as the sciences. By naming and organising files, creating extensive documentation and metadata, and categorising, researchers can more easily access and understand their findings, leading to better decisions and outcomes.

Accessing Data

Each researcher will inevitably find themselves working in various locations, collaborating with colleagues from all over the world, and accessing existing sources of data, whether they are a student researching the mysteries of the universe, a researcher probing the complexity of the human mind, or a librarian cataloguing ancient texts.

Many researchers need remote access to their data and records to do this, enabling them to work whenever and wherever they choose. This is especially crucial in the modern world, where hybrid working has become the norm. Therefore, it is crucial to have dependable and secure access to data as more researchers choose to work remotely. In the humanities, this could involve access to digital archives including historical documents or literature, whereas in the sciences this may include access to databases or programmes storing enormous amounts of raw data.

Paul Griffiths, a Data Champion and post-doc within the Department of Chemistry and National Centre for Atmospheric Science, connects the significance of accessible data back to why he originally applied to become a Data Champion. *“I help people work more reproducible and more openly. Incorporating ideas of reproducibility and openness into my own work has really helped.”* Making sure data is accessible is important because *“it saves time, lets me track my work more easily, and I can more easily build on work I did a while ago.”*



Data access and sharing go hand in hand. In a world that is reliant on increasingly collaborative research, ensuring that data are available and accessible requires the sharing of files and datasets with colleagues and potential collaborators.

“Of course, it’s about making data accessible,” notes Clair “but it is also about taking those correct data management steps to make the research reproducible. This means other people can use the findings with confidence and know that it’s of good quality.”

Ensuring that data can be accessed through collaborative centres and management systems such as data repositories encourages transparency and openness and helps to preserve the data for usage by future generations. Other researchers can also reproduce and build upon shared information, further propelling a management process that is built on honesty, trust and open communication.

Looking After Your Data

The right management tool for research data is only one aspect of this complex puzzle. Any data that have been created for an individual's purposes is also valuable to the wider academic community. Therefore, building in a plan to store, backup and preserve data is an important part of data management.

Selecting a storage solution suited for the type and quantity of data during the collection period is essential for the project’s long-term viability. Data can be kept on a personal computer or external hard drive for minor projects, but for bigger projects, cloud-based servers or repositories like Dataverse or Figshare are a more reliable and safe solution.

Alex expands on what data storage looks like at the Fitzwilliam Museum:

“Data storage is a big part of what we do at the museum. A lot of data is stored digitally, but in a museum setting data storage can encompass a range of non-digital data like paintings which haven’t been digitised. Our collections database is growing but there are a lot of different types of data that have to be accounted for within that digitisation process.

Because storage was such a large part of what we do, a sustainable service storage space was built into any research projects and proposals. Now, any research proposals we deal with are a part of the submission process - everyone submitting a proposal, submits a data plan.” - Alex, Data Champion

Sharing Data



"Data Champions Cartoons" by Clare Trowell is licenced under CC-BY-NC-ND

Sharing research data is essential to the research process and goes beyond simply checking a box at the end of a project. Once data have been created, organised and analysed, the data can be shared with a wider audience.

Different disciplines have different data storage needs and data sharing restrictions, which is why having Data Champions who are well-informed about their own disciplines is essential. One of the best ways to guarantee data are shared in the most secure and accessible way is to upload it to accredited repositories. This also prevents pointless data duplication, helping to raise the intrinsic value of the data at hand.

By making data available for secondary use, other researchers may gain access to tools they might not otherwise have had access to, thus helping to advance development and innovation throughout the academic environment. Any data made accessible to the public can have greater impact outside the academic community by influencing legislation, education, the media, and creative uses.

Alex describes the importance of sharing data and information, again within a museum context and how this is a way of sharing information:

"One way we share data is through object labels and within the gallery space. This is also quite an important way of sharing and disseminating data or information about the given artefact." - Alex, Data Champion (Fitzwilliam Museum)

Lucy describes the importance of sharing data and information within a Genetics and Biological Sciences context, noting how this practice can help those working with specific requirements or restrictions to increase the visibility and transparency of research:



“My area of expertise is genetics and related Biological Sciences. The researchers I engage with are highly specialised. It is precisely because their work is so interdependent that it’s important for them to share their data and results with others.

There is a general understanding that sharing your data and other research is important for the field as a whole, and within the University, there has been a concerted move towards researchers wanting to share data as part of their career progression.

Take the labeling of files as an example, if you haven’t done this correctly no one else will be able to find what you’ve worked on. If you haven’t shared the data within those files, then it becomes incredibly difficult for others to go through that reproducibility process which is an important part of understanding the results you have.” – Lucy, Data Champion

Sharing data involves a sense of responsibility and commitment to research integrity. Sharing data appropriately can have far-reaching benefits, enabling the research community to build on work and expanding the impact and reach of your research. *“If people can see that you’re sharing and you’ve got this particular expertise”,* says Clair *“then they might want to work with you...depending on the type of data you have and any restrictions, it could be used outside academia for use in policy”.* *“It links back to those principles of openness and transparency,”* she adds, *“Of course, sharing data is a way of protecting yourself, but it builds up a moral side of the work, The more open you are about sharing your data and the processes you’ve been through, the more honest and complete the research can be.”*



Closing Statement

The University of Cambridge Data Champion Programme successfully incorporates research integrity principles within its design, implementation and outputs.

Imparting advice to other individuals or institutions looking to learn from the Data Champion Programme, Frazer emphasises:

“A programme like the Data Champions is crucial for encouraging good data management practices and research integrity because communication is key.

What’s integral to the programme is that it brings together like-minded people from very different disciplines, developing a healthy, natural and self-sustaining process of data management and research integrity.

Making space for people who are interested and knowledgeable about the challenge at hand is a way other institutions can have an almost immediate impact. It’s a no-brainer to me, this programme pays for itself.” - Frazer, Data Champion

We would like to thank the Data Champions that contributed their personal experiences and stories to this article¹.

To find out more, go to Cambridge University’s Research Data Management website, where you can access information on the programme as well as an impressive array of tools, resources and guides to help anyone doing research, no matter the subject.

<https://www.data.cam.ac.uk/intro-data-champions>

¹ All kindly gave permission to be named in this document.