Introduction to Research Data Management

Hello, I'm Holly Ranger, and I am the Research Data Management Officer in the Research & Knowledge Exchange Office at the University of Westminster. In this short talk, I'll introduce you to the basic concepts of research data management, I'll talk you through some of the ways in which research data management is relevant and beneficial to your doctoral research, and I'll show you where you can find further help and guidance for managing your research data both within and beyond the university.

So what exactly is research data, and what does 'managing' research data mean? The UK research councils define research data as "*the evidence that underpins the answer to the research question, and which can be used to validate findings regardless of its form (e.g. print, digital, or physical)*" (UKRI Concordat on Open Research Data, 2016, available at: https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/].

Research data are therefore all the physical or digital materials that underpin the findings or argument of your final doctoral thesis, and this will vary from discipline to discipline:

- if you are a humanities researcher, for example, your research data might be photographs of archival documents, reading notes and bibliographies;
- if you are a practice-based researcher or a visual artist, your evidence might be the documentation of your creative practice in a portfolio or sketchbook, it might be a ceramic glaze recipe, or a video of a performance rehearsal;
- if you are a social scientist, your data might be questionnaire results, and transcripts or recordings of interviews with study participants;
- and if you are a STEM researcher, your data might be what we typically think of as data: spreadsheets, code, instrumental readings, electronic lab notebooks, or 3D models.

Research Data Management describes the processes by which you will manage or 'curate' all of this evidence that underpins your final thesis both during and after the lifetime of your doctoral programme: from gathering your evidence and storing it securely, to preserving it for the long term – perhaps to create a personal archive or to share your datasets for other researchers to validate and reuse. This is referred to as the research data lifecycle (Jisc, 2020, available at: https://rdmtoolkit.jisc.ac.uk/research-data-lifecycle).

Research data management is a central component of good research governance and organisation. If you have to submit an ethics application for your doctoral research project, for example, you'll be asked to describe how you plan to manage your research project data through the lifecycle; and if you decide to pursue a career in academia, you'll find that describing your data management or curation plan is a standard component of funding grant applications – across all disciplines. But it is also good practice to think about how you will gather, organise, store, and curate your research before you begin any new research project. Looking after your research evidence effectively will:

- improve the integrity of your research;
- help you keep your research organised;
- and mitigate against the risk of accidental loss of your research or the inappropriate release of sensitive information.

So research data management encompasses areas such as information security and research integrity, but it is also about helping you to make sure that your research is efficient and effective. Thinking about research data management early can save you time and effort later, as it will help

you make the right decisions from the outset about creating, storing, and sharing your research data. Let's take a look at some of the things you should think about when you start a new piece of research.

Before beginning your research, you should think about what it is that you will be creating and collecting in the course of your doctoral research.

- What type of material will you be working with?
    o If you will be working mostly with physical materials or artefacts, how will you ensure that your research is not accidentally damaged or lost? Think about making a plan to regularly digitise your research or sketchbooks, taking photographs, or typing up notes. As well as taking photographs, anything stored on paper can be scanned and sent to your University e-mail account.
    o If you are collecting audio, you should make a plan for transcribing speech, where appropriate, or transferring or automating copies of recordings made on a device to your cloud storage.
- If you are working with born-digital material, what kinds of data will you create? And how much data will you create?
    o If you will be generating large quantities of computational data, will you need to apply to your department for additional storage space? Is there any specialist software you will need to download?
- If you are collecting instrumental or experimental data, what methods will you use to capture your data and how will these ensure that your data are high quality? Will you use standard protocols, replicate your research, or automate data capture?
- Will you be working with human or animal participants, or gathering personal data or sensitive data from research participants?
    o You will need to think about ethical issues such as gaining consent, and your obligations under the UK Data Protection Act to keep this information safe: this will affect how you collect your data, and where and how you store this data. Your supervisor will help you with this.
- Will you be working in the field?
    o You should make a plan for how you will back up and transfer the data you collect on a laptop, Dictaphone, or a stills or film camera to your cloud storage as soon as is practicable.
- Will you be re-using material or datasets that already exist?
    o You should be mindful of copyright and any licensing restrictions: and you can find detailed guidance on the Library webpages and libguides about finding, reusing, and citing existing material: https://libguides.westminster.ac.uk

You should also think about how you will organise all of this material so that it is easily identifiable and easy to find again in the course of your research:
https://www.westminster.ac.uk/research/researcher-support/research-data/working-with-data/organising-your-data-files

    o If you'll be working in archives or libraries, think about how you could best organise your reading notes, and what file naming conventions you will use: you can name and organise your files using the archive accession numbers or the convention of author surname+publication date.
    o You can add data tags and keywords to Word documents and OneNote notes to help you identify themes and retrieve relevant notes later with ease.

- o And use reference management software such as Mendeley or Zotero to keep track of everything you're reading – you don't want to have to write out a 3,000 item bibliography by hand when you're due to submit your thesis! You can find detailed guidance on the Library webpages https://libguides.westminster.ac.uk/referencing/zotero
- o If you're capturing a creative process, you could make sure that your digital copies are stored by date.
- o If you are using particular software, or if you are running specific protocols, what contextual information do you need to store alongside that data for you to understand it in the future? Do you need to record methodologies, equipment settings or abbreviations used?
- o In many STEM subjects, there are also community standards for naming, describing, and structuring your data files: you can find more detailed guidance on this on the website of the Digital Curation Centre: https://www.dcc.ac.uk/guidance/standards/metadata

You should also think about where you will store all your physical and material research data. The university provides you with access to Microsoft OneDrive and Google Drive: using these cloud storage spaces ensures that your raw data or digitised material is always automatically backed-up and accessible from any device at any time, protecting your research in case your laptop is lost or stolen. If your research contains personal or sensitive information, however – this could be health care data, or the names of participants – you must store this data in OneDrive. Google Drive does meet the security standard required by UK law to protect personal and sensitive information. If you use an operating system such as Linux, talk to your supervisor about the best place to store and back up your files and code.

As you reach the end of your doctoral research, you should think about what material or research evidence you will retain and preserve after your project ends. If you have been funded by a research council, they may even expect or mandate you to do this. The easiest way to preserve your data long-term is to transfer it to a dedicated data repository. A data repository is essentially a data warehouse that specialises in storing and preserving the quality of data long-term. There are generic data repositories that anyone can use to preserve their data, such as Figshare (https://figshare.com/) and Zenodo (https://zenodo.org/). Some research funders, such as the ESRC, have a dedicated repository: the UK Data Archive (https://www.data-archive.ac.uk/); some disciplines and publishers do, too: the scientific Nature journals recommend using the Dryad repository (https://datadryad.org); and there are also dedicated repositories for genomics data or crystallography: so you should find out if there are any standard repositories for your research community (use the Registry of Research Data Repositories, https://www.re3data.org/). You might have shared some work in progress over the course of your research via platforms such as Wordpress, vimeo or YouTube: although these are massive platforms, their data can still become corrupted over time: depositing copies of these videos and websites in a data repository will ensure that these videos remain in a secure and stable location long-term. Preserving your research evidence ensures that you can access all your data after you leave the University or if you transfer institutions. You won't need to or want to preserve everything, but you can find more guidance about how to choose what you keep on our webpages:

https://www.westminster.ac.uk/research/researcher-support/research-data/archiving-data/data-repositories-and-archives

https://www.westminster.ac.uk/research/researcher-support/research-data/archiving-data/appraising-and-selecting-data

Finally, you should also consider whether to publish and share your research materials and datasets. The University of Westminster is committed to open-access as a good research principle, and your accepted thesis will be made available for download free-of-charge from WestminsterResearch (https://westminsterresearch.westminster.ac.uk/), the University's online repository of publications from Westminster's academic community. Increasingly, it is standard practice for researchers to publish their underlying datasets alongside the journal article that discusses their results – this ensures that your results can be validated by others. So you might want to think about doing this for your thesis: if you have deposited your research data in a repository, you can choose how open or closed your data is, and you can license your data for reuse by others. Adding a license allows others to clearly see what they're allowed to do with your data and what they're not allowed to do. You can find out more about Creative Commons Licences and copyright on the library webpages: https://libguides.westminster.ac.uk/copyrightresearchers

But there are also good reasons for not sharing your research evidence or data: you might need to protect your intellectual property; you might have been working with commercially-sensitive data from an industrial placement; or you might have personal and sensitive data. But you can still preserve these data in repositories and restrict access to them.

In sum, research data management encompasses all the things you should think about as you complete a research project: from planning and designing your project; through creating or collecting your research and storing it securely, to preserving and sharing your research.

Your research supervisor can help you with many of these considerations. You can also find detailed guidance on the Library websites and dedicated libguides (https://libguides.westminster.ac.uk), as well as the Research Data webpages on the main University website: https://www.westminster.ac.uk/research/researcher-support/research-data

There are also dedicated websites and services that can help you, such as the UK Data Service (https://ukdataservice.ac.uk), and the Digital Curation Centre (https://www.dcc.ac.uk).

In March, I'll also be running a workshop to supplement this short talk. This workshop will discuss the basic principles of research data management across the research data lifecycle in a little more detail, from creation to publication. It will also provide you with an opportunity to write your own data management plan for your doctoral research project, and I'll be on hand with subject specialists to answer any questions you have about managing your research data. Sign up in the VRE here: https://research.westminster.ac.uk/v090x/write-your-own-data-management-p