This Finishing Touches module focuses on publishing research outcomes which include the supporting data. On completing this section you should be more aware of the considerations that should be made when choosing where to publish and what needs to be considered before publishing your research data.

These materials could be used as an introduction to scientific writing; continue by demonstrating the differences in the writing style used in journal papers and magazines. An activity could then be to write a 200 word summary of their project. Alternatively, this module can be used to highlight the benefits of putting details into the publication and making the research data that supports the research open access in an archive. You could use this as an introduction for publishing research data and ask the audience to select data to archive or demonstrate your institutional archive.

Section 1: Publications

Focus: Publishing options

Main points:
- Journal papers aren't the only option.
- More than one medium can be used.
- There are importance issues to consider; these are discussed in the next few slides.

Highlight that publications are important for your professional reputation, but they can work against you if you do not acknowledge the contributors, the data, all opinions, and describe your work fully.

Focus: Journal Impact

Main points:
- High impact journals are more esteemed, and therefore more difficult to publish in.
- Nature and Science are magazines so they will edit your manuscript and they are not openly available.
- Check your subject for suitable journals in the web of knowledge citation reports.

Highlight that subject journals are targeted by your peers and can improve your papers citation.

Focus: Expenses

Main points:
- Publishing is not free; either you pay to submit, your audience pays to read, or both.
- Charges may be per page, with additional charges for colour pages – do you need colour print or just online versions?

Highlight that you need to include publication costs within your funding bid.

Focus: Accessibility

Main points:
- How discoverable is your paper? Subscribers only? Open after a set period of time? Or completely open as you pay a premium.
- Can you submit a manuscript to an open access electronic print archive?

<u>Highlight</u> that funding bodies may require that you make your work open access so this could affect your journal options.

Focus: Conference Proceedings

Main points:
- Conferences are an excellent opportunity to promote your research.
- Posters, presentations, and proceedings all advertise your work.
- Be aware of poachers – you can decline photographs or sharing of your slides, but advertise your paper coming soon.

<u>Highlight</u> that funding for students to attend can be procured from professional societies, conference organisers, and industrial partners.

Focus: Press release

Main points:
- Newspapers and university news articles can also promote your work.
- You may be asked to interview so make sure you know your stuff.

<u>Highlight</u> that they are not researchers, use basic language and be clear – you don't want them to get the wrong end of the stick.

<div style="background-color:#8dc63f;">

Section 2: Preserving Data

</div>

This section could be used as an introduction to preserving data which could continue to how to deposit in an institutional or subject archive.

The issues of what to preserve and how to do so should be considered at the beginning of the project process during the funding application and ethics approval so that provision can be made for sharing.

<u>Focus:</u>   Benefits of sharing data

<u>Main points:</u>
- Reusing existing data cuts out data gathering time and expense.
- Sharing increases the impact of your work
- Essentially a backup of your data.

<u>Highlight</u> that most funders require that researchers preserve their data.

<u>Focus:</u>   Selection

<u>Main points:</u>
- Preserving every test, version, and stage of your analysis is **not useful**.
- Keep what can be reused easily.

<u>Highlight</u> that data that helped them, may help others; processed data? Analysis code? Electronic versions of results.

<u>Focus:</u>   Examples to preserve

<u>Main points:</u>
- Raw data is crucial if it is not stored at the facility where it was acquired.
- Processed data could be more useful, but are others likely to reprocess the raw data for consistency with new data.
- Bespoke software – maybe chargeable – and scripts that are difficult or time consuming to reproduce.

<u>Highlight</u> that metadata and documentation is required for all preserved data.

<u>Focus:</u>   Supporting data

<u>Main points:</u>
- Your work needs to be verifiable, so make the numbers behind plots available.
- Electronic versions of tables can be downloaded and used as comparisons quickly and easily – you'll be cited.
- Electronic images hold extra information in the third dimension; e.g. intensity, noise measurements, background values.

<u>Highlight</u> that the more information shared, the more you can be credited for.

Notes for the trainers

Research Data Management
Anonymisation

If you have information that is not vital to the study, then don't share it, anonymise it.

- Remove direct identifiers (e.g., personal information such as addresses)
- Aggregate or reduce the precision of variables that might be identifiable (such as postcode).
- Generalise text variables to reduce identifiability
- Restrict continuous variables to reduce outliers
- Pay particular attention to anonymising relational data - some anonymised variables may become identifiable when considered in combination.

Whenever editing is done, researchers need to be aware of the potential for distorting the data. For example, deleting all possible identifiers from text or sound recordings is a simple but blunt tool that creates data that are confidential but may be unusable.

UK Data Archive

University of Hertfordshire UH   Safeguarding data with Research Data Management

Focus: Anonymisation

Main points:
- If you can anonymise for your paper, you can anonymise your data for sharing.
- Make sure your ethic approval allows you to share your data.
- Open access does not necessarily mean anyone can download it – you can apply restrictions.

Highlight that you should not breach DPA to share your sensitive data.

Section 3: Anonymisation

Research Data Management
Sensitive Data

Sensitive Data is information covering:
- The racial or ethnic origin of the Data Subject
- Political opinions
- Religious or other beliefs of a similar nature
- Membership of trade unions
- Physical or mental health or condition
- Sexual Life
- The commission of any offence or criminal records

Sensitive data must be collected using an opt-in and should be carefully handled. Other classes of data which might be regarded as sensitive are data relating to children and financial information, and also data related to national defence.

Sensitive data also includes any information that is protected by University policy from unauthorized access. This information must be restricted to those with a legitimate business need for access.

University of Hertfordshire UH   Finishing Touches of Research Data Management

Focus: Sensitive data

Main points:
- Generally these data are personal information that can lead to discrimination.
- If your research focuses on links between these data, then identifiers should not be published alongside these data.

Highlight that these data can be collected, but care should be taken in the sharing and storage of these data. This should be considered thoroughly at the beginning of the project when ethics is applied for.

Research Data Management
Personal and Confidential Information (PCi)

Personal data
Personal data are data which relate to a living individual who can be identified from those data or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller and includes any expression of opinion about the individual and any indication of the intentions of the data controller. This includes any other person in respect of the individual (Data Protection Act 1998).

Confidential data
Confidential data are data given in confidence or data agreed to be kept confidential, i.e. secret, between two parties, that are not in the public domain such as information on business, income, health, medical details, and political opinion.

The University UPR12 sets out the acceptable practice for managing PCi.

University of Hertfordshire UH   Finishing Touches of Research Data Management

Focus: Personal and confidential data

Main points:
- Personal data relates to a living individual and may include sensitive information.
- Confidential data is only given if agreed that it will not be shared.

Highlight that personal data can be collected, but confidential data cannot be shared without permission, even among colleagues.

Research Data Management
Anonymisation

A person's identity can be disclosed from:

direct identifiers such as names, addresses, postcode information, telephone numbers or pictures

indirect identifiers which, when linked with other publicly available information sources, could identify someone, e.g. information on workplace, occupation or exceptional values of characteristics like salary or age

Direct identifiers are often collected as part of the research administration process but are usually not essential research information and can therefore easily be removed from the data.

UK Data Archive

University of Hertfordshire UH   Finishing Touches of Research Data Management

Focus: Direct Identifiers

Main points:
- Obvious information like names, addresses, phone numbers and dates of birth.
- Less obvious are occupations, places of work, and age.

Highlight that these data can be collected, but care should be taken in the sharing and storage of these data and together they can identify individuals.

Research Data Management
Anonymisation

- Assign each participant an anonymity code, which would be their identifier in a SPSS or Excel database.

- Only the PI can match an individual to their code and each participant can be given their own code in case they wish to withdraw from the study.

Most data are used en masse so individuals are difficult to identify, it is the grouped responses that researchers are interested in.

When publishing the data, information that identifies participants cannot be published such as name, address, telephone and email contact information. Apply this same anonymisation to the data sets, what doesn't need to be shared?

You can ask the participants for permission to publish certain disclosed information when you're ready to archive it if you did not do so when it was collected.

University of Hertfordshire UH   Finishing Touches of Research Data Management

Focus: Anonymisation methods

Main points:
- The simplest method is to remove direct identifiers and replace with a reference number.
- This reference can be used to identify individuals if you need to contact them.

Highlight that how you're going to anonymise data should be declared when you apply for ethics approval.

Research Data Management
Ethics approved sharing

Your sensitive data is most at risk when you share it between devices and people.

If you are working with the NHS, you will be provided with an encrypted email account that you need to use to contact them.

In your everyday environment, you can;
- encrypt your emails using **Outlook**; this also encrypts the attachments but you will need ensure the recipient can decrypt it.
- encrypt your data and files using **TrueCrypt** before attaching them to your email or using the file exchange.

University of Hertfordshire UH   Finishing Touches of Research Data Management

Focus: Ethics approved sharing

Main points:
- Data is most at risk when you're passing it between colleagues.
- Use encryption to protect your data either directly in your email client, or using software.

Highlight the tools used by your institution.

Research Data Management
Open Access with restrictions

The term "Open Data" is unsettling and can have different connotations depending on who you talk to. Generally, it means that data is accessible by everyone. With respect to repositories, it is unreasonable and unlikely that all data is available without restrictions.

- If data is discoverable, then it is open to the public and accessible.

- It may not be *downloadable* by everyone

- You can restrict access to professionals, members of a particular group i.e. NHS workers.

Archives will offer different levels of access so your data is protected against misuse.

University of Hertfordshire UH   Finishing Touches of Research Data Management

Focus: Open Access for Sensitive Data

Main points:
- If the metadata can be discovered then it's open access.
- You can restrict access to authorised institutions or professional at a university.

Highlight that different archives offer different levels of restrictions. This should be considered when choosing an archive. Funders may specify which archive has to be used.

Research Data Management
NHS Data archives

Sensitive data has to be anonymised before being preserved in an archive; however, this is not a new concept as data must be anonymised for publication.

Databases with sensitive information tend to have access control so that data is not misused. The National Institute of Health and Care Excellence (NICE), provide an evidence search which is accessed using NHS Athens.

The University of Hertfordshire is a member of Athens, so if you are supporting NHS patients, you can get a NHS Athens account for free.

NHS approved databases are reviewed by their ethics committee and meet their requirements for the preservation of research data.

University of Hertfordshire UH   Finishing Touches of Research Data Management

Focus: NHS archives

Main points:
- The NHS provides an encrypted email for use when communicating with them.
- The NHS has a list of approved archives where their security and treatment of sensitive data has been ethics approved.

Highlight that the University's institutional archive may have been approved by the NHS.

Section 4: Archiving Data

These materials could be used as part of an introduction to preserving data which could continue to how to deposit in an institutional or subject archive. This includes a summary of which data should be deposited in an institutional archive, as well as using openDOAR to search for a suitable archive.

The researcher's funding body may dictate what data is deposited when and where. This should be considered when applying for funding.

<u>Focus:</u> Definitions

<u>Main points:</u>
- There is a subtle difference between a repository and an archive.
- A catalogue does not necessary mean that the objects are stored locally.

<u>Highlight</u> that the requirement to preserve data cannot be satisfied by keeping the data on your website and listing it in a catalogue as this cannot be persistently cited.
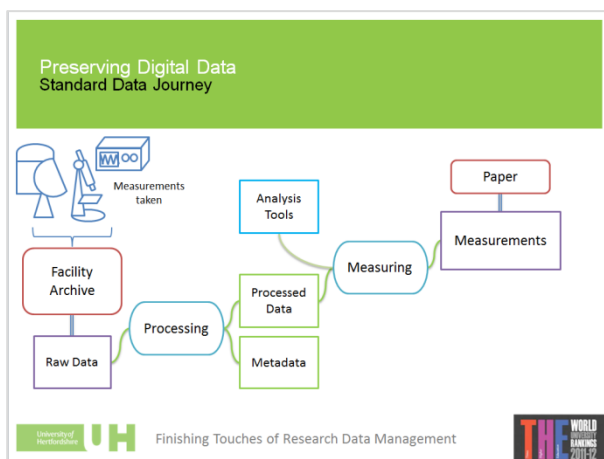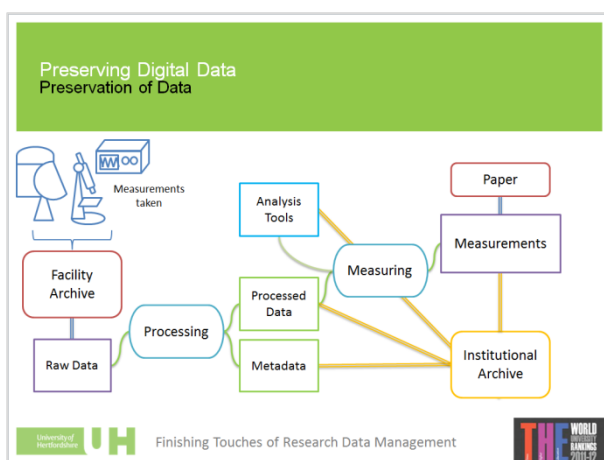

<u>Focus:</u> Standard practice

<u>Main points:</u>
- Currently, collected data is processed and analysed before the results are presented and explained in a paper.

<u>Highlight</u> that the processed data, metadata, and analysis tools are not preserved.

<u>Focus:</u> Data preservation

<u>Main points:</u>
- New emphasis on preserving data within an institutional or subject archive.
- If enough data is stored in institutional archives from a specific subject, a subject archive could be installed later due to this demand.

<u>Highlight</u> that institutional archives are there to allow you to share your data when you do not have access to a subject archive; a subject archive should be the first place people would look for data.

Focus: Archive catalogue

Main points:
- OpenDOAR is constantly growing as more archives are created.
- Refining to your language, country, or subject reduces the transformations required before you can deposit.

Highlight some archives charge for management of your data and that retention periods vary.

Focus: Institutional archives

Main points:
- You cannot redeposit somewhere else later, as you need a persistent citation; however, you can enter a record elsewhere referencing your data's location.
- Discoverability is less than in a subject repository, so ensure you declare your data's location and its unique digital identifier in your publication.

Highlight that a record of data in a subject repository should also be entered in your institutional data catalogue.