Managing your Research Data

Jenny O’Neill, Data Manager, UCD Library

jenny.oneill@ucd.ie
Before we Begin

Data Sharing and Management Snafu in 3 Short Acts

- NYU Health Sciences Library
- https://youtu.be/N2zK3sAtr-4
Lessons Already

- Funder requirements
- Publisher requirements
- Storage & backup
- Metadata
- File formats
- Documentation (software & codebook)
- Naming variables
Overview

• What are research data?
• Why manage your research data?
• Before your research
• During your research
• After your research
• Questions
What are Research Data?

- Documents (text, MS Word), spreadsheets
- Scanned laboratory notebooks, field notebooks, diaries
- Online questionnaires, transcripts or surveys
- Digital audio or video recordings
- Transcribed test responses
- Database contents
- Digital models, algorithms or scripts
- Contents of an application (input, output, log files for analysis software, simulation software, schemas)
- Documented methodologies and workflows
- Records of standard operating procedures and protocols
“data needed to validate the results presented in scientific publications”
What is data management?

- create
- preserve
- share
- store
- use

Long-term management of data

day to day

Adapted from Activities Involved in RDM, (c) Stuart Macdonald/EDINA. Used with permission
Why Manage your Research Data?

“Your research data is a valuable resource that will have taken a great deal of time and money to create”

https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data
Why Manage your Research Data?

- **Efficiency**: makes your own research easier
- **Safety**: protect valuable data
- **Quality**: better research data = better research
- **Reputation**: enhances research visibility
- **Compliance**: with ethical codes, data protection laws, journal requirements, funder policies
Why Manage your Research Data?

- Meeting funder requirements (e.g. HRB, IRC, H2020)
- Reduce risk of losing work
- Possibility of citations based on data
- Opportunities for follow-on research and collaboration
- Good research practice!
Data Management Plans

A data management plan (DMP) is "A formal statement describing how research data will be managed and documented throughout a research project and the terms regarding the subsequent deposit of the data with a data repository for long-term management and preservation."
DMP Templates

• Data types
• Data organisation, documentation & metadata
• Data storage & security
• Ethics & intellectual property
• Data sharing & re-use
• Long-term preservation
• Implementing your plan
MANAGEMENT ≠ SHARING
Informed Consent

• Voluntary
• Informed
• Specific
  – given for a specific purpose, including future use of data and data sharing, if applicable to the research project – researchers should anticipate how the data may be used in the future and address it in the consenting procedure
• Explicit
Useful links

• UCD Library's Data Management Checklist
  http://libguides.ucd.ie/ld.php?content_id=9797218

• DMP Online
  https://dmponline.dcc.ac.uk/

• Sample DMPs
  http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples
DURING YOUR RESEARCH
<table>
<thead>
<tr>
<th>Filename</th>
<th>Date Modified</th>
<th>Size</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>data_2010.05.28_test.dat</td>
<td>3:37 PM 5/28/2010</td>
<td>420 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.28_re-test.dat</td>
<td>4:29 PM 5/28/2010</td>
<td>421 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.28_re-re-test.dat</td>
<td>5:43 PM 5/28/2010</td>
<td>420 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.28_calibrate.dat</td>
<td>7:17 PM 5/28/2010</td>
<td>1,256 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.28_huh???.dat</td>
<td>7:20 PM 5/28/2010</td>
<td>30 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.28_WTF.dat</td>
<td>9:58 PM 5/28/2010</td>
<td>30 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.29_aaarrgh.dat</td>
<td>12:37 AM 5/29/2010</td>
<td>30 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.29_#$@*!&amp;!.dat</td>
<td>2:40 AM 5/29/2010</td>
<td>0 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.29_crap.dat</td>
<td>3:22 AM 5/29/2010</td>
<td>437 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.29_notbad.dat</td>
<td>4:16 AM 5/29/2010</td>
<td>670 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.29_woohoo!!.dat</td>
<td>4:47 AM 5/29/2010</td>
<td>1,349 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>data_2010.05.29_USETHISONE.dat</td>
<td>5:08 AM 5/29/2010</td>
<td>2,894 KB</td>
<td>DAT file</td>
</tr>
<tr>
<td>analysis_graphs.xls</td>
<td>7:13 AM 5/29/2010</td>
<td>455 KB</td>
<td>XLS file</td>
</tr>
<tr>
<td>ThesisOutline!.doc</td>
<td>7:26 AM 5/29/2010</td>
<td>38 KB</td>
<td>DOC file</td>
</tr>
<tr>
<td>Notes_Meeting_with_ProfSmith.txt</td>
<td>11:38 AM 5/29/2010</td>
<td>1,673 KB</td>
<td>TXT file</td>
</tr>
<tr>
<td>JUNK...</td>
<td>2:45 PM 5/29/2010</td>
<td></td>
<td>Folder</td>
</tr>
<tr>
<td>data_2010.05.30_startingover.dat</td>
<td>8:37 AM 5/30/2010</td>
<td>420 KB</td>
<td>DAT file</td>
</tr>
</tbody>
</table>
Naming files

- Be concise, but informative e.g. project acronym, study title, instrument, location, date, data type, version number, filetype
- Be consistent
- Use lowercase letters and avoid spaces
- Date should be YYYY-MM-DD to facilitate sorting

- e.g. 2017-09-28 Managing your Research Data (All disciplines).pdf
## Version Control

<table>
<thead>
<tr>
<th>Version</th>
<th>Responsible</th>
<th>Notes</th>
<th>Last Amended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image3_1.0</td>
<td>VK</td>
<td>Original document</td>
<td>25/07/2016</td>
</tr>
<tr>
<td>Image3_1.1</td>
<td>LD</td>
<td>Minor revisions made</td>
<td>08/08/2016</td>
</tr>
<tr>
<td>Image3_1.2</td>
<td>SM</td>
<td>Further minor revisions</td>
<td>12/08/2016</td>
</tr>
<tr>
<td>Image3_2.0</td>
<td>VK</td>
<td>Substantive changes</td>
<td>17/10/2016</td>
</tr>
</tbody>
</table>
Storage & Security

• How will the data be stored during the research?
• How will the data be shared during the project?
• Back-up: How will it be done and how often will it be done?
• How will you manage access and security?
• How will data security be guaranteed e.g. data encryption, password etc.
<table>
<thead>
<tr>
<th>LEGAL DATA TYPES</th>
<th>RED DATA</th>
<th>AMBER DATA</th>
<th>GREEN DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Protection of data is required by law or</td>
<td>UCD has an obligation to</td>
<td>Protection of data is at the discretion of the</td>
</tr>
<tr>
<td></td>
<td>regulatory instrument.</td>
<td>protect the data.</td>
<td>owner or custodian.</td>
</tr>
<tr>
<td>REPUTATION OF</td>
<td>Disclosure would cause exceptional or long</td>
<td>Could cause harm to the</td>
<td>Low risk of embarrassment or reputational harm.</td>
</tr>
<tr>
<td>UCD DATA</td>
<td>term damage to the reputation of the</td>
<td>reputation of the University.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>University, or risk to those whose</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>information is disclosed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMMERCIAL DATA</td>
<td>May have serious or long term negative</td>
<td>May have short term financial</td>
<td>No impact to the commercial operation of UCD.</td>
</tr>
<tr>
<td></td>
<td>financial impact on the University.</td>
<td>financial impact on the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>university.</td>
<td></td>
</tr>
<tr>
<td>OTHER INSTITUTIONAL RISKS</td>
<td>Information which provides access to resources, physical or virtual.</td>
<td>Smaller subsets of protected data from a school.</td>
<td>General university information.</td>
</tr>
<tr>
<td>APPROVED STORAGE PLATFORM OPTIONS</td>
<td>Novell Storage.</td>
<td>Google Drive Novell Storage Microsoft office 365</td>
<td>Google Drive Novell Storage Microsoft office 365</td>
</tr>
<tr>
<td></td>
<td>Encrypted Storage Device*</td>
<td>Encrypted Storage Device *</td>
<td>Storage Device *</td>
</tr>
</tbody>
</table>


Backup Strategy

3

total copies of your data

2
different mediums (devices)

1
copy offsite
Useful links


- Research IT http://www.ucd.ie/itservices/ourservices/researchit/

- IT Security http://www.ucd.ie/itservices/ourservices/security/

AFTER YOUR RESEARCH
Open Science

- Open Access
- Open Data
- Open Source
- Open Notebook
FAIR Data Principles

• They are not a technical specification!

• A minimal set of community-agreed guiding principles and practices to ensure that research data is:
  – Findable
  – Accessible
  – Interoperable
  – Reusable

• Initially developed by Dutch Tech Centre for the Life Sciences (2014)
• Reviewed and refined through multi-stakeholder practitioner groups, including Force11 and the Research Data Alliance
• Published in Nature Scientific Data, 2016
FAIR Data Principles

• **Findable** – Assign persistent IDs, provide rich metadata, register in a searchable resource,...

• **Accessible** – Retrievable by their ID using a standard protocol, allows for authentication/authorization, metadata remain accessible even if data aren’t...

• **Interoperable** – Use standard vocabularies, qualified references, shared and broadly applicable language for knowledge representation...

• **Reusable** – Rich, accurate metadata, clear licences, provenance, use of community standards...
Benefits of data sharing

• **Speed**: The research process becomes faster
• **Efficiency**: Data collection can be funded once, and used many times for a variety of purposes
• **Impact & longevity**: Open publications and data receive more citations, over longer periods
• **Transparency & quality**: The evidence that underpins research can be made open for anyone to scrutinise, and attempt to replicate findings. This leads to a more robust scholarly record

“As open as possible, as closed as necessary”
Plan to Share from the Start

- Negotiation on licenses and consent agreement may preclude later sharing if not careful
- Costings can’t be included retrospectively
- Useful to consider data issues at the consortium negotiation stage to make sure potential issues are identified and sorted asap
Personal Data Sharing

• Personal data cannot be shared with a third party, unless specific and explicit consent is secured.

• Even if data is de-identified/anonymised prior to sharing it with a third party, this must be covered by valid consent of the person to whom the data pertains.

• All human data sharing arrangements should be covered by a ‘Data Sharing Agreement, which should be reviewed by UCD Legal Office.
Metadata

• Metadata tells us a story about the data
  – structured: relay on (international) standards, e.g. Dublin Core, DDI etc.
  – unstructured: non-standardised documentation, e.g. field report
Dublin Core

- Title
- Creator
- Date
- Description
- Rights
- Type
- Lang
- Contributor

- Relation
- Source
- Coverage
- Subject
- Identifier
- Format
- Publisher

Dublin Core and the Digital Repository of Ireland [https://repository.dri.ie/catalog/rx91h464n](https://repository.dri.ie/catalog/rx91h464n)
Document as you go

- Good quality documentation allows others to find and understand your data
  - Codebooks
  - Data dictionary
  - Software
  - Metadata
  - Lab notebooks
Make the Data Available

• Choose a discipline specific repository
• Talk to colleagues about which repositories are available for your discipline

www.re3data.org

• Zenodo is a multidisciplinary repository

zenodo.org
<table>
<thead>
<tr>
<th>Basic Quality Assurance</th>
<th>Reproducability</th>
<th>Reusability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-up strategy within the project</td>
<td>Back-up strategy for storing data after the project (for 10 years)</td>
<td>Plan submission to an archive for long-term preservation</td>
</tr>
<tr>
<td>Data collection and versioning guidelines</td>
<td>Metadata to describe the entire research process</td>
<td>Standardization, e.g. by employing licensed scales</td>
</tr>
<tr>
<td>Minimal documentation, e.g. sampling, variable and code labels</td>
<td>Detailed documentation for reuse</td>
<td></td>
</tr>
<tr>
<td>Legal / ethical issues: informed consent for use of data within the project</td>
<td>Legal / ethical issues: data storage or making it accessible to others</td>
<td>Legal / ethical issues: archiving and reuse (covered by informed consent?)</td>
</tr>
<tr>
<td>File formats that fulfill the needs of the primary research group</td>
<td>File formats for keeping data &amp; documentation accessible for at least 10 years</td>
<td>File formats that facilitate data reuse in the future</td>
</tr>
</tbody>
</table>

Data Management Plan

Adapted from CESSDA Training [https://www.cessda.eu/Research-Infrastructure/Training/](https://www.cessda.eu/Research-Infrastructure/Training/)
Useful links

• Research Data Alliance Metadata Standards Directory
  http://rd-alliance.github.io/metadata-directory/

• FAIRSharing.org
  – A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.
  https://fairsharing.org/

• DCC: How to License Research Data
  http://www.dcc.ac.uk/resources/how-guides/license-research-data
And finally...

- Publications are not data and don’t need to be included in a DMP

- Open Access (including Green/Gold) is not an appropriate term for data, use data sharing or data publishing

- A website/blog is not suitable for long term sharing.
free clinics

RESEARCH DATA MANAGEMENT

When: 13:00 to 14:00
Location: Link 3, James Joyce Library
For dates and more information see our Library Guide:
http://libguides.ucd.ie/data

Services include help with:
- finding a suitable data repository
- ethical considerations
- publishing your data
- data organisation
- data storage
- metadata & documentation
- IT security
- informed consent
- file formats
THANK YOU

jenny.oneill@ucd.ie

http://libguides.ucd.ie/data