Managing your Research Data
(Arts, Humanities & Social Sciences)

Dr. Amanda Doran, Data Manager, UCD Library
amanda.doran@ucd.ie
Overview

• What are research data?
• Why manage your research data?
• Data Management Plans
  – Data types
  – Data organisation, documentation & metadata
  – Data storage & security
  – Ethics & intellectual property
  – Data sharing & re-use

• Please ask questions at any time!
What are Research Data?

- Interviews, diaries, anthropological field notes, focus groups, answers to survey questions
- Transcribed test responses
- Coded numerical responses to surveys
- Digital audio or video recordings
- Digital images
- Database contents
- Digital models, algorithms or scripts
- Maps & geospatial data
- Ephemera
- Archival material
“data needed to validate the results presented in scientific publications”
What is data management?

- create
- preserve
- share
- store
- use
- document

Long-term management of data

Day-to-day management of data

Adapted from Activities Involved in RDM, (c) Stuart Macdonald/EDINA. Used with permission
UCD Research

Research IT

RDM support

Data Protection Officer

Office of Research Ethics

Intellectual Property – NOVA UCD

UCD Library
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”
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
Irish Funders

“The Lead Applicant is strongly encouraged to think carefully about the data to be generated and/or secondary data to be re-used during the project and describe the data cycle via a DMP in a manner consistent with relevant legal, ethical and regulatory frameworks or standards.”

Health Research Board Emerging Investigator Awards for Health (EIA) 2017

“Applicants are required to address the data management needs of their research project. As part of the application, applicants will furnish an outline Data Management Plan (DMP) appropriate to their project and, if successful, a detailed DMP will be submitted to Council within six months of the award commencement date.”

Irish Research Council Laureate Awards Programme, 2017
MANAGEMENT ≠ SHARING
FAIR Data Principles

- 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)
- Published in Nature Scientific Data, 2016

- [FAIRSharing.org](http://www.force11.org/group/fairgroup/fairprinciples)
  - A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.
FAIR Data

- **Findable**: It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- **Accessible**: It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren’t accessible.

‘How FAIR are your data?’ checklist, CC-BY by Sarah Jones & Marjan Grootveld.
FAIR Data

- **Interoperable**: Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- **Reusable**: Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.
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

http://libguides.ucd.ie/ld.php?content_id=9797218
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
• UK Data Archive Data management costing tool
  https://www.ukdataservice.ac.uk/manage-data/plan/costing
Fictional AHRC technical plan

The digital output of the project will be a database of historical recipes from throughout Britain from the 1600s to the first world war. Recipes will be geocoded based on project research into where the recipe originated and it will be possible to search for ingredients or styles of cooking based on region/town in combination with dates to show how the use of ingredients and methods changes over time and space. A highly innovative, vibrant and interactive public website will be created through which data will be plotted on historical map layers and on interactive visualisations, and an app will also be released. User interaction will be integral to the website and users will be able to post comments about recipes including images and video clips of dishes created by following the recipes. A content management system will also be developed for the project’s researchers to use to record recipe and other associated data over the course of the project and for user submitted content to be managed.

DATA TYPES

Provide a description of the data your project will capture, create or use.
Data Types

• How will data be created (captured)?
  – e.g. interview data, questionnaires etc.

• What data formats will be used?
  – For analysis (if different)
  – For long-term access

• File formats: Open, non-proprietary, uncompressed, ubiquitous
<table>
<thead>
<tr>
<th>Type of data</th>
<th>Recommended formats</th>
<th>Acceptable formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>• PDF/A (.pdf)</td>
<td>• PDF (.pdf)</td>
</tr>
<tr>
<td></td>
<td>• XML (.xml)</td>
<td>• MS Word (.doc/.docx)</td>
</tr>
<tr>
<td></td>
<td>• plain text (.txt)</td>
<td>• RTF (.rtf)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• ODT (.odt)</td>
</tr>
<tr>
<td>Audio</td>
<td>• Free Lossless Audio Codec (FLAC) (.flac)</td>
<td>• MPEG-1 Audio Layer 3 (MP3)</td>
</tr>
<tr>
<td></td>
<td>• WAV file (.wav)</td>
<td>• Audio Interchange File Format (.aiff)</td>
</tr>
<tr>
<td>Quantitative Datasets</td>
<td>• SPSS portable format (.por)</td>
<td>• SPSS (.sav)</td>
</tr>
<tr>
<td>(statistical file formats)</td>
<td>• Tab-delimited file (.tab) with setup file</td>
<td>• STATA (.dta)</td>
</tr>
<tr>
<td></td>
<td>(for SPSS, Stata, SAS, etc.)</td>
<td>• SAS (.7bat; .sd2; .tpt)</td>
</tr>
<tr>
<td>Images</td>
<td>• JPEG (.jpg, .jpeg)</td>
<td>• Photoshop (.psd)</td>
</tr>
<tr>
<td></td>
<td>• TIFF (.tif, .tiff)</td>
<td>• RAW (.raw, .dng)</td>
</tr>
<tr>
<td>Video</td>
<td>• MPEG-2 (.mpg, .mpeg, ...)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• MPEG-4 H264 (.mp4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Lossless AVI (.avi)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• QuickTime (.mov)</td>
<td></td>
</tr>
</tbody>
</table>
Useful links

- UCD Digital Library Preferred Formats for Data: [https://digital.ucd.ie/about/#technical_formats](https://digital.ucd.ie/about/#technical_formats)
- ISSDA File Format Policy: [https://www.ucd.ie/t4cms/ISSDA_Format_Policy_V2.pdf](https://www.ucd.ie/t4cms/ISSDA_Format_Policy_V2.pdf) [PDF]
- UK Data Service Recommended formats: [https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats](https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats)
- DRI Factsheet No 3: File formats: [https://doi.org/10.7486/DRI.rj43ck402](https://doi.org/10.7486/DRI.rj43ck402)
Fictional AHRC technical plan

The project will use a variety of open and proprietary formats that will best suit the needs of the project’s outcomes. These will be migrated to suitable open standards to facilitate preservation at the end of the project. Text will be transcribed as plain text with HTML markup and will take up roughly 500Mb of space. Images will be in the JPEG format and 5Gb of server space will be set aside for them. Video files will be MOV and 20Gb of space will be available for them. Visualisations will be SVG files. The map interface will be based around Google Maps. Web pages will follow current HTML and CSS standards.
DATA ORGANISATION, DOCUMENTATION & METADATA

Organising, documenting and describing data is important in order to assure quality control and reproducibility of data.
Good quality documentation allows others to find and understand your data:

- Interview protocol
- Questionnaires & interviewer instructions
- Codebook or data dictionary
- Information sheets, Consent forms, Ethical approval
- Database schemas
- Methodology reports
- Provenance information about sources of derived or digitised data

https://mantra.edina.ac.uk/documentation_metadata_citation/
Metadata

• Metadata tells us a story about the data
• Structured metadata relies on international standards, e.g.
  – Dublin Core
  – Data Documentation Initiative (DDI)

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

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

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

http://dublincore.org/
DATA STORAGE & SECURITY

How will the data be stored during the research and how will you manage access and security?
<table>
<thead>
<tr>
<th></th>
<th>RED DATA</th>
<th>AMBER DATA</th>
<th>GREEN DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LEGAL DATA TYPES</strong></td>
<td>Protection of data is required by law or regulatory instrument.</td>
<td>UCD has an obligation to protect the data.</td>
<td>Protection of data is at the discretion of the owner or custodian.</td>
</tr>
<tr>
<td><strong>REPUTATION OF UCD DATA</strong></td>
<td>Disclosure would cause exceptional or long term damage to the reputation of the University, or risk to those whose information is disclosed.</td>
<td>Could cause harm to the reputation of the University.</td>
<td>Low risk of embarrassment or reputational harm.</td>
</tr>
<tr>
<td><strong>COMMERCIAL DATA</strong></td>
<td>May have serious or long term negative financial impact on the University.</td>
<td>May have short term financial impact on the university.</td>
<td>No impact to the commercial operation of UCD.</td>
</tr>
<tr>
<td><strong>OTHER INSTITUTIONAL RISKS</strong></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><strong>APPROVED STORAGE PLATFORM OPTIONS</strong></td>
<td>Novell Storage. Encrypted Storage Device*</td>
<td>Google Drive Novell Storage Microsoft office 365 Encrypted Storage Device *</td>
<td>Google Drive Novell Storage Microsoft office 365 Storage Device *</td>
</tr>
</tbody>
</table>

http://www.ucd.ie/itservices/ourservices/documentsandstorage/options/
Encryption

• “Encryption is a secure method of scrambling messages in a format that is unreadable by unauthorised users”
• Encrypt your computer
• Encrypt individual files
• Please remember that if you email an encrypted file, then do not email the password. Please phone or text the password to the person.

http://www.ucd.ie/itservices/ourservices/security/encryption/
Backup Strategy

- 3 total copies of your data
- 2 different mediums (devices)
- 1 copy offsite
Useful links

• Which File Storage should I use?
  http://www.ucd.ie/itservices/ourservices/documentsandstorage/options/

• UCD Cloud Computing Evaluation Guidelines

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

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

• REC Data Storage and Retention Guidelines
Fictional AHRC technical plan

The website and content management system will be hosted on LAMP servers based at project partner the University of Edinburgh who will supply the project with two virtual servers: a ‘development’ and a ‘live’ server. The resource will be developed using the Joomla content management framework. Images will be edited with Adobe Photoshop and videos with Final Cut Pro.

Fictional AHRC technical plan

An initial project website will be set up by the developer during the first month of the project. This version of the website, along with project presences on social networking sites such as Facebook and Twitter, will be managed by the Co-I. During months 1-3 the developer, in collaboration with the rest of the project team, will create a scoping study for the content management system and the public website. S/he will work on a first version of the CMS during months 4-6, launching it during month 6 with further iterations which will introduce further functionality being made every few months over the three years of the project.

The 10 project RAs, who will have begun collating data on their laptops from fieldwork to libraries and archives around the UK in month 2, will receive training in the use of the CMS by the developer in month 6. Data uploaded to the CMS by the RAs will be analysed and processed by the developer to convert it into formats suitable for display on the project website, which the developer will be working on during years 2 and 3 of the project.

A ‘beta’ version of the online resource will be made available to selected users midway through year 2 of the project. This version will feature full access to the recipe records but limited search and browse functionality. The launch of the ‘beta’ version will coincide with the project symposium.

The developer will continue to refine the online resource, adding functionality such as the map interface and the visualisations throughout year 3 of the project. An official launch of the final resource that will be available to all users will take place in the final month of the project to coincide with the project conference, at which point users will be able to access the maps and visualisations and post comments on the recipes.

ETHICS & INTELLECTUAL PROPERTY

Do your data contain confidential or sensitive information?
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

UCD Office of Research Ethics: www.ucd.ie/researchethics/
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.

UCD Office of Research Ethics: www.ucd.ie/researchethics/
General Data Protection Regulation (GDPR)

• Came into force across Europe 25 May 2018

• GDPR applies to any research that uses personal data, including scientific research and studies in the arts and humanities

• Irreversibly and effectively anonymised data is not “personal data” and the data protection principles do not have to be complied with in respect of such data

http://www.ucd.ie/gdpr/
Anonymisation

• A person's identity can be disclosed from:
  – **Direct identifiers** such as names, postcode information or pictures
  – **Indirect identifiers** which, when linked with other available information, could identify someone, for example information on workplace, occupation, salary or age

• Balance anonymisation with access control to preserve data usability

https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation
Third Party Copyright

• Section 24(1) CRRA – “The copyright [...] shall expire 70 years after the death of the author, irrespective of the date on which the work is first lawfully made available to the public.”

• Public Domain: Public Domain means a work in copyright is put beyond copyright.

• Licences: Commonly used licences include: Irish Copyright Licensing Agency, Creative Commons or bespoke licences.

• Orphan works: A work is considered ‘orphan’ when the copyright holder can’t be traced. This is dealt with in Directive 2012/28/EU.
Licencing

- Creative Commons
  https://creativecommons.org/

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

By Shaddim; original CC license symbols by Creative Commons - https://creativecommons.org/about/downloads/https://creativecommons.org/policies/Original CC license icons licensed under CC BY 4.0, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=47247325
LONG-TERM PRESERVATION

What data will be kept or destroyed after the end of the project?
Benefits of data sharing

• Your data are valuable!
• Maximises impact to the benefit of scholarly communication & society
• Your data can be used in novel ways
• Your findings can be replicated and compared with other studies
• Your data can contribute to advancing knowledge

Dublin: Childhood Development Initiative (CDI).
“As open as possible, as closed as necessary”

Make the Data Available

1. Choose a discipline specific repository
   – Talk to colleagues about which repositories are available for your discipline
     - Registry of Research Data Repositories (re3data.org)
     - PLOS Recommended Data Repositories
     - Scientific Data Recommended Data Repositories

• Quantitative Data: Irish Social Science Data Archive (ISSDA)
• Qualitative Data: Irish Qualitative Data Archive (IQDA)
• Cultural Heritage Data: UCD Digital Library
Make the Data Available

2. If no disciplinary repository exists for your discipline consider depositing data in a multidisciplinary repository:

– [Dryad](#)
– [Figshare](#)
– [Dataverse](#)
– [Open Science Framework](#)
– [Zenodo](#)
Fictional AHRC technical plan

Upon completion of the project the digital outputs of the project will be migrated to open standards for preservation as discussed in Section 2a. Outputs will be tagged with appropriate metadata to facilitate their discoverability. Long-term preservation of digital data is a considerable challenge; however, how best to preserve digital data is not the focus of this project and other projects within our partner institutions are already making significant progress in how this issue can be addressed.

The website will continue to be hosted by the University of Edinburgh beyond the end of the project and user comments and contributions will continue to be enabled. In order to minimise the cost of sustaining the resource in the longer term the CMS will not be retained following the end of the project and the focus instead will be on the public facing website.
Data Papers

“The intent of a data paper is to offer descriptive information on the related dataset(s) focusing on data collection, distinguishing features, access and potential reuse rather than on data processing and analysis.”

Data Journals

• Research Data Journal for the Humanities and Social Sciences (RDJ, Brill, 2017);
• Journal of Open Psychology Data (JOPD, Ubiquity Press, n.d.a);
• Journal of Open Archaeology Data (JOAD, Ubiquity Press, n.d.b);
• Open Health Data (Ubiquity Press, n.d.c).

Where should I publish my research?: libguides.ucd.ie/ld.php?content_id=31318676
### Data Management Plan

#### Basic Quality Assurance
- Back-up strategy within the project
- Data collection and versioning guidelines
- Minimal documentation, e.g. sampling, variable and code labels
- Legal / ethical issues: informed consent for use of data within the project
- File formats that fulfill the needs of the primary research group

#### Reproducability
- Back-up strategy for storing data after the project (for 10 years)
- Metadata to describe the entire research process
- Legal / ethical issues: data storage or making it accessible to others
- File formats for keeping data & documentation accessible for at least 10 years

#### Reusability
- Plan submission to an archive for long-term preservation
- Standardization, e.g. by employing licensed scales
- Detailed documentation for reuse
- Legal / ethical issues: archiving and reuse (covered by informed consent?)
- File formats that facilitate data reuse in the future

---

Adapted from CESSDA Training [https://www.cessda.eu/Research-Infrastructure/Training/]
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.
THANK YOU

amanda.doran@ucd.ie

http://libguides.ucd.ie/data

PS: please fill in the feedback forms