Open Science & FAIR Data in Horizon 2020

Jenny O’Neill, Data Manager, UCD Library
jenny.oneill@ucd.ie
Overview

• Horizon 2020 requirements
• Open Science
  – Open Access
  – Open Data
• FAIR Data
• Questions
Horizon 2020 requirements

From July 2017

RESEARCH DATA ARE OPEN BY DEFAULT

with possibilities to opt out
Data Management Plan

Proposal stage
• Can include details of RDM under the ‘Impact’ criterion
• What standards will be used
• How data will be shared
• How data will be curated and preserved

During the project
• A DMP is a project deliverable
• This is not a fixed document and should evolve
• First version is due at the 6 month mark
H2020 Useful links

- European IPR Helpdesk Fact Sheet Open Access to scientific publications and research data in Horizon 2020: Frequently Asked Questions (FAQs)
Open Science

“Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools.”
Open Science

“Open Science is about extending the principles of openness to the whole research cycle, fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done.”
Open Science

“to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction”

Goals of Open Science

• Transparency in experimental methodology, observation, and collection of data
• Public availability and reusability of scientific data
• Public accessibility and transparency of scientific communication
• Using web-based tools to facilitate scientific collaboration

Dan Gezelter, What, exactly, is Open Science?
Benefits of openness

• SPEED: The research process becomes faster
• EFFICIENCY: Data collection can be funded once, and used many times for a variety of purposes
• ACCESSIBILITY: Interested third parties can (where appropriate) access and build upon publicly-funded research resources with minimal barriers to access
• IMPACT and LONGEVITY: Open publications and data receive more citations, over longer periods
• TRANSPARENCY and 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
Open Access

• Open Access (OA) means that electronic scholarly research outputs are made freely available on the Web to all, with no license restrictions.
  – Green
  – Gold

• H2020 Requirements:
  – An electronic copy of the publication must be deposited in a suitable green open access repository.
Open Data

• Open data is data that’s available to everyone to access, use and share.
• But open data should be easy to access.
• Open data isn’t the same as big data, but big data can be open data too.
• Oh, and it’s also not the same as ‘shared data’.

https://theodi.org/article/what-is-open-data-and-why-should-we-care/
Opting Out in H2020

• data may commercially exploited
• data are confidential in connection with security issues
• sharing would break data protection regulations
• participation would mean that the project's main aim might not be achieved
• the project will not generate / collect any research data
• other legitimate reasons
What are Research Data?

- Text documents, notes
- Numerical data
- Questionnaires, surveys, survey results
- Audio and video recordings, photos
- Database content (video, audio, text, images)
- Mathematical models, algorithms
- Software (scripts, input files ...)
- Results of computer simulations
- Laboratory protocols, methodological descriptions
- Sometimes also samples, artifacts, objects.
What data needs to be shared?

“data needed to validate the results presented in scientific publications”
Rule of thumb

“As open as possible, as closed as necessary”
Benefits of data sharing

- Impact & longevity: Your data may be cited by others. Open publications and data receive more citations, over longer periods.
- Compliance: Funders, publishers and institutions may require that you share your data.
- Transparency & quality: Your findings can be replicated and compared with other studies.
- Collaboration: creates opportunities for follow on research and collaboration.
- Re-use: Your data can be used in novel ways. Data sharing facilitates re-use of your data for future / follow-on research and discovery as data collection can be funded / collected once, and used many times for a variety of purposes.
- Efficiency: Data sharing is good research practice!

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.
General Data Protection Regulation (GDPR)

• Comes 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
Anonymisation Resources

• UK Anonymisation Network (UKAN) Anonymisation Decision-making Framework
  – http://ukanon.net/ukan-resources/ukan-decision-making-framework/

• UK Data Service
  – https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation

• IQDA Guide to Anonymising Qualitative Data
  – https://www.maynoothuniversity.ie/iqda/data-resources/resources-researchers
FAIR Data Principles

• Principles to enhance the value of all digital resources;
  – Data, images, software, web services, repositories..

• Developed and endorsed by researchers, publishers, funding agencies, industry partners.

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

www.force11.org/group/fairgroup/fairprinciples
FAIR Data Principles

• An increasing number of discoveries are being made using other people’s data
• However, *significant* effort was needed to *find* the right datasets, put them all together and *use* them
• If we are ever to realise the full potential of content we create then we must find a way to reduce the barrier to (automatically) *find* and *reuse* that content.
• To achieve this we must build a social & technological infrastructure for the discovery & assessment of digital resources.
• Increasing the FAIRness of digital resources will *increase* their *quality* and their potential for *reuse*.

Michel Dumontier, Distinguished Professor of Data Science, Maastricht University
https://www.slideshare.net/micheldumontier/are-we-fair-yet
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.
  – Naming conventions
  – Search keywords
  – Clear version numbers
  – Metadata standards

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

• Accessible: Provide documentation and tools that someone else will need to access the data.
  – How will the data be made accessible (e.g. by deposition in a repository)?
  – If certain datasets cannot be shared a clear reason why should be provided
  – 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, [EUDAT](http://eudat.eu)
FAIR Data

• Interoperable: Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.
  – Use standardised metadata and methodologies
  – Consider appropriate file formats for long term access
    • ISSDA File Format Policy
    • DRI File Formats

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

- 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.
  - Creative Commons CCO/CC-BY
  - DCC: How to License Research Data

'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT
Make the data available

• Open by default – choose open source software and open file formats where possible, and only restrict access to data when there is a good reason for doing so
• Plan to share - especially in multi-partner work, planning is a means of communication
• Document as you go – the best quality metadata is produced by the research team themselves, as close as possible to the moment of capture
• Reasons for not sharing should be recorded
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
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
Questions (1)

• Clear guidelines on how to approach Open Data
  – what materials should be made open (e.g. interview transcripts, audio recordings, informed consent).
  – There is concern about making sensitive interview material open for interview with young people (children).
Questions (2)

• Do researchers have to make sure they can track back individual data, in case an individual would like to get access to their data in a few years time?
Questions (3)

• Are there option to not share data in the case that studies get comprised with lack of participation.
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.
QUESTIONS?

jenny.oneill@ucd.ie

http://libguides.ucd.ie/data