The Changing Research Data Environment: a briefing paper

Introduction

Data are more easily generated than ever before and are even more fundamental to research in many fields. This has been recognised by research funders and others, generating new expectations that data will be shared and managed, will contribute to open science and will be preserved for the long term as part of a redefined scholarly record. These developments raise a number of issues and obligations for the University which are highlighted in this paper. An appendix summarises some recent local experience.

Research Data Developments

Research data have had a growing profile in recent years, not just because of major increases in volume, but also because of their greater prominence in research across many disciplines, their potential for new approaches to research and the greater interest shown in them by higher education administrators and funders. Three key developments have emerged:

- Research funders view research data as a major asset and expect it to be well managed and widely shared. Horizon 2020, for example, requires the submission of a data management plan for any funding applications relating to research which will generate data. Research data sets are also a significant asset for universities and provide opportunities for reputational gain.

- Open science (a term not limited to science) is being strongly promoted in Europe and worldwide. The key concepts associated with open science are transparency and collaboration in terms of the conduct of research, and openness with regard to the dissemination of research outcomes of all kinds, including research data.1 Greater transparency is seen as reducing both duplication and fraud.

- The scholarly record, until recently largely focused on publications such as monographs and journals, is becoming much more diverse. Interim materials generated during the research process and contributory to final publication include datasets, interactive programs and visualisations. These too are increasingly available and have the potential to enrich research if properly managed for discovery, access and use.

---

1 Open innovation, open science, open to the world: A vision for Europe (European Commission, 2016), ISBN: 978-92-79-57346-0; DOI: http://dx.doi.org/10.2777/061652.
Issues for Universities

- Research data management requires effort and resources; researchers commit themselves and their universities to this when they submit data management plans with funding proposals. Requirements include secure data storage, metadata creation and publication for open access.

- There is an increasing expectation that data will conform to the FAIR principles, making them Findable, Accessible, Interoperable and Reusable. This involves significant further processing of data, for example to achieve standard formatting for seamless use with other data sets.

- Types of data, and hence data management requirements, will vary across disciplines.

- Access to data may need to be restricted for legal and other reasons, raising potentially complex data protection issues; data may need to be managed differently from a data protection perspective in an open access environment.

- Long-term preservation of data may require intervention over a period of time, for example to migrate to newer formats.

- Awareness by researchers of evolving data management requirements and practices, including funder policies, needs to be promoted and ensured on an ongoing basis.

- Universities and their researchers need to be appropriately credited for their generation and publication of data sets.

- New sources of data are emerging via the field of ‘learning analytics’ in which student performance, activities, interactions, and other information in both individual and aggregate (eg cohorts, discipline groups, etc) forms are being compiled. These constitute large data sets which are potentially useful for educational diagnostic purposes, but also for large-scale systematic research. It is crucial, at this early stage of the development of learning analytics, that issues around data protection, ownership, data management, and research ethics are clarified.

- Data repositories act as a public good generating positive externalities that can potentially go beyond the immediate data providers to foster interdisciplinary research across the University.
Implications for NUI Galway

There are implications for the University in terms of practice, infrastructure and resources. The expectation from funders and others is that data will be open for reuse by others where at all possible and will be curated for long-term access. Funding bids will fail otherwise. Essential infrastructure will include appropriate data protection policies and storage solutions. Researchers will require advice from personnel skilled in creating data management plans, managing data throughout the research cycle, publishing for open access and preserving for future use. A programme of education for researchers at all career stages will be needed as data management becomes a core practice for researchers. Data sets should also be recognised and captured as University research outputs.

Other universities have established research data management services; examples are:

Stanford University: https://library.stanford.edu/research/data-management-services
University of Glasgow: http://www.gla.ac.uk/services/datamanagement/
University College London: http://www.ucl.ac.uk/library/research-support/research-data

In the emergent area of Learning Analytics, there are a number of national, European, and international developments in terms of standards, policy, and technologies. It is important that NUI Galway participates in such debate and is fully aware of the implications both for operational use of such systems (where appropriate) and the research potential.
NUI Galway needs to consider how to meet the growth in expectations and obligations around research data management. It is not feasible to absorb this within existing resources, given the extent, variety and specialist nature of the interventions required.

In the immediate term, a Research Data Working Group has been formed, with representation from a number of constituencies across the campus, as per the signatures below. It has prioritised the following initiatives:

- Drafting a separate research data policy (a review of existing related policies in the University has indicated a need for a separate policy for research data)
- Organising a public event, with internal and external speakers and a showcase of student and academic work to create awareness and engage communities
- Highlighting at an early stage through this document the key developments in research data management and their implications for the University

**NUI Galway Research Data Working Group:**
John Cox (University Librarian, chair), Peter Corrigan (Head of Digital Publishing and Innovation, Library), Trish Finnan (Digital Publishing and Data Management Librarian, Library), Peter Feeney (University Solicitor), Dr. Aidan Kane (Head of Economics), David Kelly (Digital Humanities Manager, Moore Institute), Dr. Iain MacLaren (Director of CELT), Niall O’Brochtaigh (Postdoctoral Researcher, Insight), Sean O’Farrell (Director of Information Solutions and Services), Aengus Parsons (Director of Research, Research Office), Dr. Srinivas Raghavendra (Lecturer, Economics).

26 April 2017
Appendix 1

Summary of a pilot project relating to research data management

In 2016 the Library initiated a pilot project to learn about research data management practice at NUIG, identify ways in which the Library could provide support and in the process develop some expertise. Two initiatives were undertaken:

- Trish Finnan (Digital Publishing and Data Management Librarian) met with applicants to the HRB Emerging Investigator Awards for Health (EIA) 2017 to highlight resources relevant to the application requirements for a data management and sharing plan.

- Interviews with eight researchers from a range of disciplines were conducted to learn about research data practices.

HRB Emerging Investigator Awards for Health (EIA) 2017

Applicants to the HRB Emerging Investigator Awards for Health (EIA) 2017 were required to develop a data management and sharing plan (DMP) in accordance with the guiding principles for findable, accessible, interoperable, and re-usable data (FAIR). The application guidelines pointed to specific resources relating to DMPs (see appendix iii at http://www.hrb.ie/uploads/tx_hrbgrants/EIA_2017_Guidance_notes_01.pdf)

The following resources were discussed at a meeting with each applicant: Digital Curation Centre’s DMP checklist, Digital Curation Centre’s DMP Online Tool, UK Concordat on Open Research Data as well as content and FAQs on the Library website (see http://library.nuigalway.ie/digitalscholarship/researchdata/researchdatamanagement/)

Conversations with applicants revealed the following:

- Complexity and variety of data generated (e.g. complex patient data, microarray data, clinical samples, numerical assay data) as well as the variety of formats (images, TIFFs, JPEGS, Excel spreadsheets, 1000 power microscopic images).

- Concerns relating to making data FAIR identified included guaranteeing anonymity, secure storage of patient data, managing copyright as well as the challenges relating to making data reusable e.g. datasets that are not self-explanatory or shareable.

- The question of whether datasets in Excel published with a journal article are considered FAIR was raised.

- The need for server space at an institutional level and a repository accommodating storage of and access to Excel datasets was identified. An index to people who have datasets on campus was suggested. Mechanisms for storage mentioned included USBs, PCs at work and shared drives, external hard drives.
• Awareness and use of disciplinary repositories was evident (GenBank, Cytometry Repository) though some applicants asked for help to identify appropriate repositories. The Data Citation Index (a Web of Science resource provided by the Library) provides a single point of access to research data from repositories across disciplines and provides links to datasets. It was used to identify a relevant data repository e.g. ENA European Nucleotide Archive. Other repositories identified for applicants were BioGPS for Skin Datasets and Melanoma Molecular Map Project (MMMP).

• Applicants agreed that research data management planning at the outset of their projects was important and that templates on the DMPonline tool would be useful for creating their plans. The Cancer Research template available in it would be useful to some applicants. Some applicants had experience of DMPs in applications to other funding programmes e.g. BBSRC (Biotechnology and Biological Sciences research Council) and would use a BBSRC template also available in DMPonline. One applicant had previously used a template for a Research Data Management Plan provided by University of Queensland.

• One applicant noted that institutional support provided for research data management would enhance a Data Management Plan for a funding application.

Trish Finnan, Digital Publishing and Data Management Librarian, 26 April 2017