Research data in Ireland: from policy requirements to researcher incentives

Rebecca Grant

twitter.com/beck_grant

#meettheexpert
Research data in Ireland: from policy requirements to researcher incentives

- Irish context (2014-2020)
- Mapping drivers, policy change, institutional practice

- Global researchers’ perspective
- How do researchers react to policy change

- Case study at Springer Nature
- Moving from policy mandates to incentivisation
What drives change in research data management in Ireland?

- Global context for research data management and sharing – e.g. Open Science, Open Access
- Legislative and policy change, e.g. EU Directives
- National legislation and initiatives, funder/publisher policy change
- Institutional policy change
- Changes in researchers’ own practice
- Development of reward mechanisms/credit?

* Open Government data, public service data sharing
Factors for change: Horizon 2020

Beneficiaries “must aim to” deposit the research data needed to validate the results presented in the deposited scientific publications, ideally via a data repository.

Participating projects required to develop a Data Management Plan (DMP).

Research teams must develop DMPs and plan for data deposit for the first time in many cases.

Institutions support compliance through guidance, policy, repository infrastructure.
Factors for change: General Data Protection Regulation/GDPR

Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

Enacted into Irish law as part of the *Data Protection Act 2018* May 2018

Describes individuals’ rights over their data, and the basis by which data can be collected by researchers

Institutions support compliance for researchers and their research
Factors for change: General Data Protection Regulation/GDPR

GDPR was the biggest change to happen in data privacy in 20 years

3. Minimisation of processing

4. Data accuracy / quality

7 Data Management Principles of GDPR

GDPR Terminology

Personal Data

Personal Data & Research

https://www.ucd.ie/gdpr/
Factors for change: the National Principles for Open Access Policy Statement, 2012

“Research data should be deposited whenever this is feasible, and linked to associated publications where this is appropriate.”

Endorsed by Science Foundation Ireland, the Irish Research Council and the Health Research Board; institutions including the Royal College of Surgeons Ireland (RCSI) and Dublin Institute of Technology; and government agencies including the Environmental Protection Agency and the Higher Education Authority.
Factors for change: National policy statement on Ensuring Research Integrity in Ireland, 2014 and 2019

Irish Universities Association

Research data storage and good practice in data retention

“Research data is a valuable resource that should be organised, curated and appropriately stored.”
## Mapping national policy change

### Table 1. UK Institutional RDM policies in date order

<table>
<thead>
<tr>
<th>Institution</th>
<th>Policy name</th>
<th>Date released / revised</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Birmingham</td>
<td>Code of Practice for Research</td>
<td>October 2009</td>
</tr>
<tr>
<td></td>
<td>(see Research Data section)</td>
<td></td>
</tr>
<tr>
<td>University of Edinburgh</td>
<td>Research Data Management Policy</td>
<td>May 2011</td>
</tr>
<tr>
<td>University of Northampton</td>
<td>Research Data Policy</td>
<td>June 2011</td>
</tr>
<tr>
<td>University of Hertfordshire</td>
<td>Data Management Policy</td>
<td>September 2011</td>
</tr>
<tr>
<td></td>
<td>(see s.7 on research data and the appendix  'Guide to RDM')</td>
<td></td>
</tr>
<tr>
<td>University of Warwick</td>
<td>Research Data Management Policy</td>
<td>November 2011</td>
</tr>
</tbody>
</table>
Ireland’s university policies

2014
2/8 policies
1/8 guidance

2020
5/8 policies
8/8 guidance
Providing guidance

Research Data Management: Introduction

Bringing together University resources and services to facilitate researchers in the production of high quality data.
Ireland’s funder policies

2014
2/6 recommend data sharing

2020
5/6 require data sharing

*Supporters of the National Principles for Open Access, 2012.

CC-BY-NC
Ireland’s funder policies: Health Research Board

The HRB is partnering with the GoFAIR International team to build awareness amongst the research community of the FAIR Data principles (Findable, Accessible, Interoperable, Reusable) and to support effective data stewardship.

https://www.hrb.ie/funding/policies-and-principles/open-research/
Standards for sharing: the FAIR Data Principles

**F** Findable
by leveraging metadata and persistent identifiers

**A** Accessible
through free and open communications protocols

**I** Interoperable
by using controlled vocabularies, implementing machine-readability and including references where appropriate

**R** Reusable
by highlighting clear licence statements that enable the greatest possible reusability

Standards for sharing: the FAIR Data Principles

https://www.force11.org/group/fairgroup/fairprinciples

CC-BY-NC
119 organisations endorse the FAIR Principles, including...
Online survey aiming to gather data on organisational data management practice; data-related policies; motivations; and whether a recordkeeping professional is involved.

Sent to 28 Irish organisations in October 2017 (purposive non-probability sampling)

11 responses received - relatively high response rate but small sample size.

https://doi.org/10.6084/m9.figshare.11365598.v1
Who responded

- Third-level institution: 4
- Non-commercial state sponsored public body*: 4
- National Cultural Institution: 1
- Local authority: 1
- Government department: 1

*n=11

*e.g. the Broadcasting Authority of Ireland, Waterways Ireland

CC-BY-NC
What are Irish organisations doing with research data?

- Create or receive data: 11
- Allow access: 11
- Store data: 10
- Appraise data: 7
- Create metadata: 7

n=11
Why do Irish organisations support research data management?

- Obliged to due to:
  - Freedom of Information (FOI) requirements
  - PSI Directive (Public Sector Information regulations)
  - Funder requirements (2)
  - GDPR was not mentioned (but did come up later, after May 2018)
What policies do Irish organisations have for their data?

- Access policy: 7
- Records management policy: 5
- Collection policy: 5
- Retention/disposal policy: 5
- Preservation policy: 4
- None: 3
- Appraisal policy: 2

n=11
What policies do Irish organisations have for their data?

- Access policy: 7
- Records management policy: 5
- Collection policy: 5
- Retention/disposal policy: 5
- Preservation policy: 3
- Appraisal policy: 2

"We will develop policies in the future"
EU policy and legislation

National strategy

Researcher facing policies

Organisational policies

Researcher practice?
Researchers are increasingly sharing their data

The State of Open Data report shows steady growth in the number of researchers sharing their data, up consistently year on year to 64% in 2018.
How researchers commonly share their data

62% have shared data both privately and publicly

36% have only shared data privately

2% of respondents have only shared their data publicly

The three most common methods of private sharing were:
- email (65%)
- USB or flash drives (41%)
- file sharing services (39%)

The three most common ways of public sharing were: (n=569)
- supplementary information to journal articles (51%)
- lab or personal website (27%)
- subject specific repository or data archive (25%)

Source: Springer Nature researcher survey “Practical Challenges for Researchers in Data Sharing” 2018. Total respondents: 7719
How familiar are you with the FAIR Principles?

<table>
<thead>
<tr>
<th>Year</th>
<th>Familiar</th>
<th>Heard of</th>
<th>Never heard of</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>15%</td>
<td>25%</td>
<td>60%</td>
</tr>
<tr>
<td>2019</td>
<td>18%</td>
<td>27%</td>
<td>54%</td>
</tr>
</tbody>
</table>

2019 n = 8423
The Five Essential Factors

To accelerate data sharing, we propose five essential factors:

1. Clear policy
2. Better credit
3. Explicit funding
4. Practical help
5. Training & education

https://doi.org/10.6084/m9.figshare.7807949.v2
What motivates researchers to share their research data?

- Journal policy: 44%
- Institutional policy: 38%
- Funder policy: 33%

(RDA Meet the Expert series 1 May 2020)


CC-BY-NC
Credit as an incentive

- 9% Think researchers get sufficient credit for sharing data
- 46% Would be motivated to share by “getting proper credit”
- 55% Would value a data citation as much as an article citation
- 60% Note that re-use that resulted in credit as a co-author would motivate them “quite a lot” or “a lot”

[https://figshare.com/articles/Five_Essential_Factors_for_Data_Sharing/7807949](https://figshare.com/articles/Five_Essential_Factors_for_Data_Sharing/7807949)
Standard research data policies at Springer Nature journals

• Rolling out across our journals since 2016.

• More than 1,600 (~65%) Springer Nature journals have one of these policies.

• Approach is practical and pragmatic, enabling all journals to adopt a policy even if they are new to data sharing.

Policy Types

- **Type 1**
  Data sharing and data citation is encouraged but not required

- **Type 2**
  Data sharing and evidence of data sharing encouraged

- **Type 3**
  Data sharing encouraged and statements of data availability required

- **Type 4**
  Data sharing, evidence of data sharing and peer review of data required
Standard research data policies at Springer Nature

**Policy Types**

- **Type 1**: Data sharing and data citation is encouraged but not required
- **Type 2**: Data sharing and evidence of data sharing encouraged
- **Type 3**: Data sharing encouraged and statements of data availability required
- **Type 4**: Data sharing, evidence of data sharing and peer review of data required

**All policy types:**

- Recommend **sharing of data via repositories**
- **Allow citation of public datasets** in reference lists/bibliographies
- Encourage use of publisher helpdesk
What our journal policies require

29% require a data availability statement

- Type 1: 627
- Type 2: 552
- Type 3: 489
- Type 4: 8

n = 1676
What is a data availability statement (DAS)?

“The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].”

“The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.”

“All data generated or analysed during this study are included in this published article (and its supplementary information files).”
Policy impact: prevalence of DAS at BMC journals


CC-BY-NC
Open data badges at Springer Nature

Three aims:

1. To assess whether Open data badges motivate authors to share the data underpinning their published papers in repositories.

2. To assess whether reader engagement increases for papers which are awarded a badge.

3. To establish the resources required to undertake badging at scale.

The pilot ran from September 2018 to January 2020
Badging for journals: the Center for Open Science Badges to Acknowledge Open Practices

Authors complete two disclosure items for each Open Data badge application:

- Provide the URL, doi, or other permanent path for accessing the data in a public, open access repository
- Is there sufficient information for an independent researcher to reproduce the reported results? If no, explain.

Peer review (certifying organisation conducts a formal review of the disclosure) can also be undertaken

RDA Meet the Expert series  1 May 2020
Center for Open Science Badges

Launched in 2014 with the journal Psychological Science

- Increase in data sharing to 23% from 3% prior to offering badges in the first year. This increased to 39% of papers receiving badges in the second year.
- Increase in data sharing in PS over other comparable journals
- Badges were mostly awarded on an honor system (unless independently verified by editorial staff)

As reported in Kidwell (2016), the number of articles actually providing available, correct, and complete was fewer than those who reported they’d met the criteria for an Open Data Badge.
Piloting the Open data badge

Aims and scope

*BMC Microbiology* is an open access, peer-reviewed journal that considers articles on analytical and functional studies of prokaryotic and eukaryotic microorganisms, viruses and small parasites, as well as host and therapeutic responses to them and their interaction with the environment.

Collection: Veterinary antimicrobial resistance and antimicrobial use
Piloting the Open data badge

• Author is alerted to the pilot on the journal website and manuscript submission system

• The paper, including a DAS, is published by *BMC Microbiology*

• Research Data team checks the DAS for the published paper and approves the paper for badging

• Research Data team adds the paper’s details to badgr.io; the journal Production team add the badge to the paper
Awarding the badges

BMC Microbiology

Putrescine biosynthesis and export genes are essential for normal growth of avian pathogenic *Escherichia coli*

Priscila R. Guerra, Ana Herrero-Fresno, Victor Ladero, Begoña Redruello, Teresa Pires dos Santos, Malene R. Spiegelhauer, Lotte Jelsbak & John Elmerdahl Olsen

*BMC Microbiology* 18, Article number: 226 (2018) | Cite this article

1644 Accesses | 1 Citations | 1 Altmetric | Metrics

Abstract

Background

Avian pathogenic *Escherichia coli* (APEC) is the infectious agent of a
Initial pilot results (September 2018 – May 2019)

Papers receiving badges

- Pre-pilot papers: 20% (n = 40)
- Pilot papers: 31% (n = 210)

CC-BY-NC
Testing our assumptions

Were authors aware of the pilot?
Working group and Editor for journal had concerns about whether authors were aware of the badges.

Were badges *actually* influencing author behavior?
Concerns were raised that journal published a lot of genetics content (a community with mandates about data sharing), so data sharing may be due to this rather than the badges.

Did authors know if their paper received a badge?
Authors were not being notified of badge status – were they aware of their badge?

Surveyed authors at the journal & explored other metrics related to the benefits of badging
In 2019 we sent a survey to every author whose paper had been assessed for an Open data badge, with a 14% response rate.

**Survey of BMC Microbiology authors**

Only 1 author knew that their paper had received a badge.

**How did you learn about Badge pilot?**

- Learned about pilot in letters from Editorial Manager: 8
- Website/Submission guidelines: 6
- Unaware of pilot: 1

*CC-BY-NC*
Other metrics: reader engagement with badged articles

**Average clicks to download (published papers)**

- Papers with badges: 53
- Journal average: 12

**Average unique page views**

- Papers with badges: 695
- Journal average: 402

RDA Meet the Expert series 1 May 2020
Data badging as an incentive for sharing

- We saw an increase in data sharing, but it was not clear that this related to badging. Other factors cannot be ruled out.
- There appeared to be a positive impact for papers which received an Open data badge, in relation to access and download.
- Authors generally felt positive about badging and journals which offer badges.
- Incentives for data sharing are still not clear.
- Mandated actions (e.g. strong policy plus checking) increases compliance more than incentives.
Other initiatives to promote credit and recognition for data sharing

Data citation

Journal policies promote citation of datasets in reference lists and publishers are collaborating to implement better ways to measure and track reuse of data through citations

Data publishing options in peer-reviewed journals

Data journals, such as *Scientific Data*, provide a means to gain publication credit specifically for publishing datasets
Data management policies have developed significantly in Ireland since 2014.

Researchers do not necessarily feel motivated by stakeholder policies for data sharing.

The concept of credit is more motivating than policy compliance.

What form should credit take?
The story behind the image

Alan Turing (1867–1934)

The scope of the achievements of Alan Turing, computer pioneer, wartime code-breaker and polymath, cannot be overstated. Renowned as the man who broke the Enigma code, Turing is also considered the father of computer science and artificial intelligence. His legacy is represented here with a visualisation of a “Turing Machine”, a hypothetical device he devised to represent the logic of a computer. The binary code depicted translates to one of Turing’s memorable quotes: Science is a differential equation. Religion is a boundary condition.

Thank you

Rebecca Grant
Research Data Manager, Springer Nature

Rebecca.grant@springernature.com

For more information on data-related activities at Springer Nature:

Email: researchdata@springernature.com
Website:
http://go.nature.com/ResearchDataServices
Rebecca Grant
Doctoral student

University College Dublin,
School of History,
James Joyce Library Building
Belfield
Dublin 4

Supervisor: Dr. Elizabeth Mullins

Contact: ann.grant@ucdconnect.ie/beck.grant@gmail.com
ORCID: 0000-0002-7614-0806
Image credits:

Saturday Sun, Unsplash, CC0
Robbie de Vries, Unsplash, CC0
Ru Recovery Ministeries, Unsplash, CC0
Kelly Sikkema, Unsplash, CC0
Beatriz Perez Moya, Unsplash, CC0
National Cancer Institute, Unsplash, CC0
Kimberly Farmer, Unsplash, CC0
Plush Design Studio, Unsplash, CC0