Data publication – a funder perspective

Conference on open access scholarly publishing
20 September 2012

David Carr
Wellcome Trust
d.carr@wellcome.ac.uk
Data sharing – Wellcome Trust view

• we are a global charitable foundation dedicated to achieving extraordinary improvements in human & animal health
• committed to ensuring research outputs (inc papers & data) are made available in a way that maximises the benefit
• leading advocate of open access & data sharing:
  – long-standing policies on open access publishing & data management and sharing
  – funding to implement policies and build key resources (e.g. Europe PMC)
  – strong supporter of reforms to UK copyright legislation to enable text and data mining
Data sharing – a growing consensus

• major challenges associated with increasingly vast & complex datasets, but also tremendous opportunities

• increasing policy convergence between major funders in promoting sharing of research data
  – expectation that data outputs be preserved and shared in a way that maximises value
  – requirement for data management plans as integral part of the application process

• but implementation and enforcement of these policies is rather variable
Major challenges…

- there are significant barriers & constraints to overcome:
  - Infrastructural issues
  - Cultural issues
  - Technical issues
  - Professional issues
  - Ethical issues

- different disciplines at very different stages; different types of data raise distinct issues

- challenges will require funders to work in partnership, with each other & other key communities
Increasing access to public health and epidemiology research data

• current situation is often that:
  – lack of access to data limits comparative analysis or even checking for accuracy
  – only immediate research interests are addressed; broader questions are neglected
  – lots of duplication in data collection
  – data not archived effectively for long-term

• recognise there are legitimate concerns, and that limits and controls are essential…

• … but growing recognition among funders that we need to find ways of widening access to this data
Sharing research data to improve public health

• Joint statement of purpose was published in January 2011, with 17 initial funder signatories

• the statement set out:

  – **one vision** – to increase availability of research data generated by our funding to accelerate improvements in public health

  – **three principles** – share data in ways that are **equitable, ethical and efficient**

  – **seven goals** – immediate goals and longer term aspirations
Three Principles

**Equitable:**
Any approach to the sharing of data should recognize and balance the needs of researchers who generate and use data, other analysts who may want to reuse those data, and communities and funders who expect health benefits to arise from research.

**Ethical:**
All data sharing should protect the privacy of individuals and the dignity of communities, while simultaneously respecting the imperative to improve public health through the most productive use of data.

**Efficient:**
Any approach to data sharing should improve the quality and value of research and increase its contribution to improving public health. Approaches should be proportionate and build on existing practice and reduce unnecessary duplication and competition.
Seven goals

**Immediate goals**
1. Data management standards support data sharing
2. Data sharing is recognized as a professional achievement
3. Secondary data users respect the rights of producers and add value to the data they use

**Longer-term goals**
4. Well documented data sets are available for secondary analysis
5. Capacity to manage and analyse data is strengthened
6. Published work and data are linked and archived
7. Data sharing is sustainably resourced for the long term
A broad range of funder partners

Signatories…

Supporting organisations…
Implementing the joint statement

• partner funders have formed a **Public Health Research Data Forum**, with a joint workplan
• activities in three main areas:
  – capacity and skills
  – culture and incentives
  – infrastructure and tools
• challenges will require key inputs from researchers, institutions, funders, repositories and publishers
• committed to engaging publishers and developing opportunities to work together
Data Citation in public health research

- Forum identified the potential of data citation to track use of datasets and to help build incentives
- we have commissioned the Digital Curation Centre in the UK to:
  - review the current status of initiatives to enable the citation of datasets generated through research
  - identify opportunities and challenges for applying data citation initiatives in the public health and epidemiology
- outcomes will inform policy discussions on whether funders could usefully do more to foster uptake of these tools
Building on emerging good practice

Journal policies that clarify expectations of authors regarding data access (e.g. PLoS One, BioMed Central)

Support for repositories to enable access to, and preservation of, data underpinning published research (e.g. Dryad)

Emerging efforts to enable data citation through use of DOIs (e.g. DataCite)
Great opportunities

• clear statement in papers on where & how underlying data can be accessed, with links where feasible
• promulgate good practice in data deposition & use of recognised repositories (where available)
• use data citations in a consistent manner that enables downstream use of datasets to be tracked
• enhance ability to access and visualise data underlying published figures
• build on gathering momentum toward open access
How to move forward - a few questions…

- what do journals see as the potential and the limits of their role as regards data?
- how can we foster consistent practice across journals on these issues? which groups should we engage?
- which elements need to be discipline-specific, and which are cross-cutting?
- what are the relative roles of funders and publishers (and institutions) in encouraging/enforcing data sharing?
Further information

http://www.wellcome.ac.uk/publichealthdata