Sharing research data: a funder perspective

2017 Council of Science Editors Annual Meeting
23 May 2017

David Carr, Wellcome Trust
Programme Manager – Open Research
(d.carr@wellcome.ac.uk)
ORCID: 0000-0003-1435-307X

Slides made available under http://creativecommons.org/licenses/by/4.0/
Sharing research outputs: our position as a research funder

• the Wellcome Trust is a global research foundation dedicated to improving health for everyone by helping great ideas to thrive

• committed to ensuring research outputs (inc. papers, data, code and materials) can be accessed and used in ways that will maximise health & societal benefit

• passionate long-term advocate of open access & data sharing
Sharing research outputs holds the potential to...

- enable outputs to be accessed, combined and re-used in ways that accelerate discovery and its application
- help ensure research findings can be validated and reproduced
- increase efficiency in research - reduce duplication and waste
- make research outputs more readily accessible to a range of users - including policymakers, healthcare professionals and the wider public
A gathering policy momentum...

• High level political drive:
  • G8 Science Ministers’ statement (2013)
  • US OSTP memorandum on expanding public access to results of federally-funded research (2013)
  • EU focus on open science

• National and international principles – for example:
  • Joint funders’ statement on sharing research data to improve public health (2011)
  • UK Concordat on Open Research Data (2016)

• Spotlight on research reproducibility and waste
A growing funder consensus...

- major challenges associated with increasingly vast & complex datasets, but also tremendous opportunities
- strong policy convergence between research funders:
  - data should be preserved and shared in a way that maximizes value
  - requirement for data management plans in grant applications
  - data underlying published findings must be accessible (and open wherever feasible)
And emerging expectations for specific data types

• Clinical trial data
  • **WHO and funder consensus** for clinical trial registration and disclosure of results
  • **ICMJE requirements for trial registration** and proposals on release of individual-level data under discussion

• Public health emergencies
  • **Funder and journal statement** (Feb 2016): Quality assured research data relating to public health emergencies should be shared as rapidly and widely as possible
  • **Cross-funder principles** developed by GLOPID-R group (2017)
  • Strong examples of funder-journal policy alignment and coordination
Wellcome’s data management and sharing policy

- Policy published 2007 (updated in 2010):
  - expects all researchers we fund to maximise access to research data with as few restrictions as possible
  - requires data management & sharing plan, where the research is likely to generate data of value as a resource for the community
  - commits to review and support costs of plans as integral part of grant
- Associated guidance sets out elements for inclusion in data management and sharing plans
- Challenges remain around clarifying expectations, allowable costs and post-award tracking

https://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing
Key challenges to overcome

• There are significant barriers to data sharing:
  • changing reward and assessment structures in research
  • developing the resources, tools and skills required for preserving, sharing and re-using data
  • addressing ethical and governance issues
• Different disciplines at very different places in developing resources and practices to support data sharing
• Funders need to work with journals, researchers and institutions to address these challenges
Concerns persist around fear of misuse of data, loss of publication opportunities and resource & time required to share data…

Van den Eynden, Veerle; Knight, Gareth; Vlad, Anca; Radler, Barry; Tenopir, Carol; Leon, David; Manista, Frank; Whitworth, Jimmy; Corti, Louise (2016): Survey of Wellcome researchers and their attitudes to open research. figshare. (https://dx.doi.org/10.6084/m9.figshare.4055448.v1)
Open Research - exploring new opportunities for Wellcome

• following a development phase project, Open Research team established at Wellcome in early 2017

• take forward funder-led and community-led activities to pilot and evaluate innovative open research approaches, and develop policies and practices as a research funder

• will actively forge partnerships with researchers, funders, journals and other key communities
Developing our policies and practices

• **Policy development** - expanding our data policy to software and materials; strengthening monitoring and enforcement

• **Machine-actionable output management plans** – piloting an approach to enhance our ability to track, update and link plans to outputs

• **Incentives** – exploring mechanisms to increase recognition and reward for open practices

• **Resource sustainability** – developing our approach for long-term funding of key infrastructures

• **Evidence base** – developing approaches & metrics to assess added value of open practices
Supporting innovative pilots

Wellcome Open Research

The Open Science Prize

Pre-prints

ASAPbio  bioRxiv  beta
Wellcome Open Research

A new publishing platform where Wellcome-funded researchers can publish any results they think are worth sharing

http://wellcomeopenresearch.org
What problem are we seeking to solve?

• Make the sharing of research outputs faster

• …More transparent
  
  All reviews are signed and public; everything (one passed initial screening) is published

• …Easier for researchers to provide information that supports reproducibility
  
  All supporting data must be shared (or explanation provided how to access it)

• …Encourage the sharing of all research outputs
  
  Address “file drawer” problem - by publishing negative and non-confirmatory results, as well as protocols, data notes, software notes, case studies etc & reduce publication bias

• …And keep costs affordable
  
  Average APC at WOR is £780; typical average charged to Wellcome is £2044
Post publication peer review model

Median time from submission to publication: 19 days
Median time from publication to passing peer review: 31 days
Data and software availability

“All articles should include the submission of the data underlying the results, together with details of any software used to process results. It is essential that others can see the raw data to be able to replicate your study and analysis of the data, as well as in some circumstances, reuse it”

https://wellcomeopenresearch.org/for-authors/data-guidelines

---

Data availability

Sequence data used for analysis in this study is publicly archived at the European Nucleotide Archive (ENA) accession code ERR1898537. Files contain high quality sequence data, as well as associated alignment data.

---

Data availability

ALSPAC data used for this submission will be made available on request to the ALSPAC Executive via this website and also provides full details and distributions of the ALSPAC study variables: http://www.bristol.ac.uk/alspac/researchers/access/. The ALSPAC data management plan (available here: http://www.bristol.ac.uk/alspac/researchers/data-access/documents/alspac-data-management-plan.pdf) describes the policy regarding data sharing. A sampler set of similar data containing relevant ALSPAC variables is available from the European Genome-phenome Archive (accession number: EGAS000001000090).

https://www.ebi.ac.uk/ega/studies/EGAS000001000090.

---

Software availability

SeqPlots is distributed as user-friendly stand-alone applications for Mac and Windows or Linux, and is available as an R programming language package from the Bioconductor repository. SeqPlots can also be deployed as a server application, which is useful for data sharing within laboratories, collaborative usage and remote work. SeqPlots is an open source and open development project: source code wiki, bug tracker and pull requests are available via GitHub.

Software is available from:

- http://przemol.github.io/seqplots (Mac, Windows, Linux, full documentation)
- http://bioconductor.org/packages/seqplots (R/Bioconductor)
- http://przemol.github.io/seqplots/#installation--server-deployment (server deployment)
- https://github.com/Przemol/seqplots (latest source code, open development tools, including wiki, bug tracker, and pull requests)

Archived source code as at the time of publication:


Reproducible code

- F1000Research added Code Ocean widgets to articles
- Code Ocean is a cloud-based platform that makes the computational code used in research both accessible and usable. Researchers and software engineers can share and run code with a single click
- Look to extend functionality to Wellcome Open Research
Closing thoughts

• Funders are committed to maximising the value of research outputs – including data, software and code

• Value of open practices increasingly recognised but many challenges remain - funders and journals have key roles in addressing these

• Welcome the leadership and innovation demonstrated by many journals and keen to further develop funder-journal collaboration

• Would call on journals to:
  – establish policies on data and code sharing
  – require data availability statements
  – develop consistent standards and approaches for data citation
Any Questions?