Briefing Paper

The value of open data to research institutions, April 2017

This guidance note presents the benefits of open data practice to institutions supporting a research programme and is aimed at research managers with responsibility for steering institutional policy, strategy and infrastructure development. This includes, but is not limited to Vice Chancellors, Deans, Heads of department and members of academic steering committees.

1.   Introduction

The surge in affordable digital technologies has led to an explosion in digital materials. Coupled with the increasing sophistication of networking technology, novel opportunities for collaboration and engagement are arising which carry huge potential for governments, citizens and business.

Open data generates 0.5% more GDP than closed data, according to research performed by Lateral Economics for the Open Data Institute[i].

Institutions and their researchers are well-placed to contribute to and exploit this new data economy by producing high-value data products and by using their unique skills to realise the possibilities of others' data. The European Union identifies four primary areas in which increased availability of data and scientific publications contributes to a stronger economy[ii]:

- Improved quality of research (by building on previous results and avoiding duplication of effort)
- Encouraging collaboration (particularly through cross-disciplinary visibility)
- Increased speed of innovation (faster progress to market of ideas)
- Improved involvement of citizens and society with research (greater transparency and familiarity increases society's support of research funding)

The ability to realise the value of the data assets produced by research rests on their 'openness', the notion that data should be findable, accessible, interoperable and reusable.[iii] Fortunately, this is not a new notion in research which has long held the expectation that the results of research should be visible, cogently argued and reproducible.

"Much of the remarkable growth of scientific understanding in recent centuries is due to open practices; open communication and deliberation sit at the heart of scientific practice."

Royal Society, "Science as an Open Enterprise[iv]

2.   The policy landscape

Increased recognition from governments and research funders of the value of open data has led to a huge increase in policies across Europe aiming to safeguard and amplify that value.[v]
Central to this philosophy for research funders in Europe is the notion that the outputs of publicly funded research are a public good and should be made as available as possible. The European Commission addresses this in its introduction to Horizon 2020:

'The Commission considers that there should be no need to pay for information funded from the public purse each time it is accessed or used. Moreover, it should benefit European businesses and the public to the full.'[vi]

Primarily, national-level policies have tended to emphasise the responsibilities of the individual researcher and their disciplines over those of the researchers host institution, as in the DFG's (Deutsche Forschungsgemeinschaft) Principles and Guidelines for the Handling of Research Data:

'The handling of research data is largely shaped by the conventions of the different scientific disciplines. The DFG therefore calls on the various sections of the scientific community to reconsider their handling of research data and develop appropriate guidelines for the discipline-specific use of such data and, if appropriate, open access to it.'[vii]

One differing approach is that adopted by the UK's Engineering and Physical Research Council (EPSRC)[viii] . The EPSRC is one of the UK's seven major research funders, all of whom have both signed up to an overarching set of RDM principles[ix] and produced data management policies tailored to suit their disciplines. The EPSRC has used its own policy to explicitly place the responsibility for enabling open data practice on the universities receiving its funding. This shift in focus has provoked a powerful response at the institutional level, much greater than that achieved by the other research council policies. As a direct result of this policy, there has been a major increase in local RDM support infrastructure developed by universities, primarily RDM policy, data publication and discovery platforms.

While the European funder landscape is less developed than in the UK, lessons will be learned from the impact of the EPSRC approach and may see an increasing emphasis on organisations to support FAIR data.

Regardless of any given policy's emphasis, researchers typically lack the capacity and capability to fully meet their open data responsibilities. This is where institutions can support research and enable open research practices by implementing local open data policy.

3.   The benefits to the institution of supporting open data

Enabling open data practices and good research data management (RDM) brings direct benefits to all participants, including the university hosting research, and is not limited to a share of societal improvements. An important factor to consider is what future benefits do the institution hope to enjoy and what actions must be taken now to ensure they can be effectively exploited in due course? Some of the key returns that institutions may expect from investing in open data support include improved ability to:

- **Attract research collaborators** – both academic and commercial
- **Maintain a competitive advantage over peers**
  The quality of institutional support for data sharing and RDM has a direct impact on the visibility of its academics' research and, therefore, on an institution's ability to attract high quality researchers.
- **Showcase high-quality research to a global audience**
- **Maintain quality of grant bids**
  With open data a key focus of many major research funders improving researchers' knowledge and engagement here will strengthen their ability to plan credible open research. In turn, this will enable them to bid more successfully for research funding.
- **Amplify the impact of research**
  By extending the lifetime and reach of research data assets, their ability to influence society increases. Evidence also shows that journal articles associated with open data enjoy higher levels of citation[xxi][xii].
- **Improve research integrity** – transparency enables validation of research results
- **Reduce the risk of inappropriate data release**
  By adopting formal data management and appraisal processes designed to identify and maximise the potential of open data assets, the institution also reduces the risks of any inappropriate release of sensitive data

The value of Open Data – A European public health example
The treatment of a major outbreak of severe gastrointestinal infection, which originated in Hamburg in May 2011, illustrates the value of open data practices. The infection spread throughout Europe and the US, resulting in the deaths of over 50 people. Scientists in Hamburg and BGI-Shenzhen in China analysed samples of the rare toxic E. coli strain responsible and, after three days, openly published a draft genome licenced under a CC0 waiver.[xiii] Within a week, scientists on four continents had released around 25 reports whose analyses enabled health workers to treat the strain effectively and contain the infection. Open data publishing facilitated the crowd-sourcing of expert analysis and ultimately saved lives.[xiv]

4.   The risks of not providing support for RDM and Open Data

When weighing up options for enabling Open Data practice, allowing support to be provided solely at the research group level can be seen as a low-cost strategy. However, this approach carries with it significant risks to the institution's financial and reputational standing which should be considered. The capacity and capability of researchers to meet open data challenges varies considerably across disciplines, and in some research groups may be unacceptably low. By providing central support, the institution acts to set the minimum level of researcher understanding and ability, ensuring that its academics are at least capable of meeting their legal and funder requirements.

The University of Bristol recently produced a business case for the development of an RDM service, enabling researchers to engage in open data practices[xv]. The case identifies four key risks to the institution of not providing suitable support:

- The University will fail to comply with funder's research data principles, leading to a potential reduction in grants awarded.
- The University will fall behind its competitors who are developing such services, and may lose its competitive edge. This will result in unsuccessful applications, reduced publication in high-impact journals and a consequent reduction in comparative citation numbers.
- Individual researchers will fail to follow good research data management practice, thereby failing to comply with funder policy and jeopardising future funding opportunities.
- The University will fail to curate valuable data assets and will produce unfit-for-purpose data leading to a reduction in reputation.

Although these arguments were developed in the UK context, the issues they respond to are rapidly becoming universal concerns. The focus on open data from H2020 and other national-level EU funders will only become more intense over the coming years and these arguments can easily be applied to any European institution supporting a research programme.

5.   How do institutions support Open Data?

Simply relying on the science of others is not an option. The greater the strength of the home science base, the greater its capacity to absorb and benefit from science done elsewhere.

Royal Society (2011). Knowledge, Networks and Nations. Royal Society: London[xvi]

These benefits and risks have motivated institutions across Europe, who have responded by using policy to emphasise their commitment to open data and by developing infrastructure which enables their academics' open research practices. This typically includes the development of technical infrastructures supporting data storage, manipulation, publication, preservation and discovery. Alongside these, 'soft' services providing support, guidance and training in all aspects of data management raise awareness and enable engagement with other services.[xvii] This provision recognises that good data management is required throughout the lifecycle of research to ensure that open data outputs are genuinely discoverable and reusable.

The institutional context and its size and research focus will have an impact on the scale and sophistication of RDM services. However, certain key areas can be expected to be managed centrally and supported by administration:

RDM service model from DCC Research Infrastructure Self-Evaluation (RISE)[xviii]

It is essential to recognise that local services do not operate in isolation and that there will be other national and international-level services which will provide support to select groups of researchers – for example EUDAT's B2 services,[xix] training provided through the FOSTER portal[xx] and discipline -specific services such as ELIXIR for the life sciences.[xxi] As part of a wider open data ecosystem, institutions can benefit from drawing on the experiences of others and by complementing existing, specialist support services.

6.   Conclusion

The benefits of open data are driving change across society. Knowledge institutions are in a prime position to exploit the new opportunities arising from this seismic change, if they can foster communication between all open data stakeholders. Central services and researchers need to act together so that institutions can engage strategically developing policy and investing in appropriate and valuable support infrastructures. Fortunately, open data challenges are an international preoccupation and pioneer institutions have contributed to resources which can support organisations developing open research data support services across Europe.[xxii xxiii]

SPARC Europe would like to thank DCC for carrying out this work

www.dcc.ac.uk

i http://theodi.org/research-economic-value-open-paid-data

ii http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

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

iv https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/

v Report: A snapshot of Open Data and Open Science Policies in Europe. Retrieved March 2017 from: https://docs.google.com/document/d/1bC7EHsq6yplVKti6HMgKVhaR3T0qfRMwe2oSsej1xs0/edit?usp=sharing

vi http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

vii http://www.allianzinitiative.de/en/core-activities/research-data/principles.html

viii https://www.epsrc.ac.uk/about/standards/researchdata/expectations/

ix http://www.rcuk.ac.uk/research/datapolicy/

x Edwin Henneken, Alberto Accomazzi, (2011) Linking to Data - Effect on Citation Rates in Astronomy. http://arxiv.org/abs/1111.3618

xi Amy Pienta, George Alter, Jared Lyle, (2010) The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. http://hdl.handle.net/2027.42/78307

xii Piwowar H, Vision TJ. (2013) Data reuse & the open data citation advantage. PeerJ PrePrints 1:e1v1 http://dx.doi.org/10.7287/peerj.preprints.1v1

xiii https://creativecommons.org/choose/zero/

xiv Rohde H et al (2011). Open-Source Genomic Analysis of Shiga-Toxin–Producing E. coli O104:H4. New England Journal of Medicine, 365, 718-724. Available at: http://www.nejm.org/doi/full/10.1056/NEJMoa1107643#t=articleTop

xv University of Bristol, 2nd Business case for provision of an RDM service (internal document) [Personal correspondence with Stephen Gray, University of Bristol]

xvi https://royalsociety.org/topics-policy/projects/knowledge-networks-nations/report/

xvii https://www.jisc.ac.uk/guides/meeting-the-requirements-of-the-EPSRC-research-data-policy#casestudies

xviii http://www.dcc.ac.uk/resources/how-guides/RISE

xix https://www.eudat.eu/b2-service-suite

xx https://www.fosteropenscience.eu/

xxi https://www.elixir-europe.org/

xxii http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services

xxiii http://www.dcc.ac.uk/resources/how-guides/RISE