

Costs and benefits of data provision

John Houghton¹ and Greg Laughlin²

¹Victoria University, Melbourne, Australia, John.Houghton@vu.edu.au

²Australian National Data Service, Canberra, Australia, Greg.Laughlin@ands.org.au

Over the last decade there has been increasing awareness of the potential benefits of more open access to Public Sector Information (PSI), research publications and research data both within Australia and around the world. That awareness is based on economic principles and evidence, and it finds expression in policy at organisational, national and international levels.

Government and research policies seek to optimise innovation by making publicly funded data available for use and re-use with minimal barriers in the form of cost or convenience. This confers three responsibilities on publicly funded agencies: (i) to arrange stewardship and curation of their data, (ii) to make their data readily discoverable and available for use and re-use with minimal restrictions, and (iii) to forgo fees wherever practical.

This paper reports on the findings of a study (currently underway) which: presents case studies examining the costs and benefits involved in making publicly funded data freely available for the agencies and their users; estimates the wider impacts of making publicly funded data available; and draws out lessons for the research sector regarding the curation and open sharing of research data.

It is always more difficult to identify and quantify benefits than costs. Benefits may accrue in a variety of ways, including cost savings, efficiency gains, and new opportunities to create value through doing things in new ways and doing new things. These are, successively, more difficult to quantify: not least because they often emerge over time and can only be realised in the future.

An obvious starting point is to begin with the most direct and directly measurable benefits, namely cost savings. Wider benefits are more difficult, and in some cases impossible, to measure. In this study we attempt to measure the benefits we can, realising that there many that we cannot capture, perhaps even foresee.

While there are a number of approaches to assessing the wider benefits of enhanced access to PSI, we focus on two. The first uses a simple microeconomic welfare approach and might be interpreted as indicative of the lower bound impact, and the second uses a macroeconomic approach in an attempt to include the multiplier and spillover impacts that are not captured in the first. While no more than exploratory, both have the merit of simplicity and being relatively undemanding in terms of data requirements.

We outline the issues to consider and data requirements for a cost-benefit analysis, exploring: PSI producer and users' activity costs before and after the transition to making data freely available; use impacts in the form of hits and downloads; and wider impacts in terms of welfare benefits and returns to investment. The approach is demonstrated in case studies, including the Australian Bureau of Statistics (national statistics) and Geoscience Australia (fundamental spatial data).

It is clear from the case studies presented that even the subset of benefits that can be measured outweigh the costs of making data more freely and openly available. It is also clear that it is not simply about access costs, but also about the transaction costs involved (i.e. the costs in handling the payments, accessing toll-secured data, understanding and complying with licensing requirements, etc.). Hence, standardised and unrestrictive licensing, such as Creative Commons, and data standards are crucial in enabling access that is free, immediate and unrestricted.

Attempting to draw out lessons for the research sector, we observe that the same basic economic principles apply to publicly funded research outputs and PSI, namely:

- Governments around the world fund research and the creation of PSI because it has public good characteristics (i.e. is more or less non-rivalrous in consumption and it is difficult to enforce exclusion);
- The Internet has drastically reduced the cost of distribution of information (i.e. the marginal cost of distribution of one additional copy is close to zero); and
- The efficient price for distribution of public goods is no more than the marginal cost of distribution (i.e. effectively zero): any other price leads to the inefficiency of excluding additional users who could have obtained access at the marginal cost of distribution (i.e. zero).

However, methods for measuring the impacts of enhanced access to PSI and research outputs often differ. Much more is known about the impacts of publicly funded research and returns to R&D expenditure than is known about the impacts of PSI. Conversely, some forms of PSI have been priced at cost recovery, whereas publicly funded research has not. This affects preferred approaches to valuing the information and/or valuing increases in access to it. While the use of case

studies and extrapolation is common to both, it is difficult to adopt a welfare economics approach where access has characteristically been free, and difficult to adopt a return to investment approach where returns have rarely been traced and measured.

The study is intended to produce both findings about the costs and benefits of data curation and sharing and be a how to guide for those considering the costs and benefits for particular public sector agencies and particular datasets. So, we outline practical approaches to simple cost-benefit analyses of both PSI and research data, to serve as a guide to those wishing to undertake such analyses. Such an analysis should include all quantifiable costs and benefits, but it must also include qualitative issues to help to prioritise PSI sharing and research data preservation and curation projects (*e.g.* incorporate a balanced scorecard approach to weighing the more intangible benefits).

What this study demonstrates is that the direct and measurable benefits of making PSI and research data available freely and unrestrictedly typically outweigh the costs. When one adds the longer-term benefits that we cannot measure, and cannot even foresee, the case for open access to public sector information and the outputs from publicly funded research appears to be very strong.

ABOUT THE AUTHORS

John Houghton is currently Professorial Fellow at Victoria University's Centre for Strategic Economic Studies (CSES) and Director of the Centre's Information Technologies and the Information Economy Program. He has had a number of years experience in information technology policy, science and technology policy and more general industry policy related economic research. He has published and spoken widely, and is a regular consultant to the Organization for Economic Cooperation and Development (OECD) in Paris. In 1998, John was awarded a National Australia Day Council, Australia Day Medal for his contribution to IT industry policy development. John's research is at the interface of theory and practice with a strong focus on the policy application of economic and social theory, and of leading-edge research in various relevant fields. Consequently, his contribution tends to be in bringing knowledge and research methods to bear on policy issues in an effort to raise the level of policy debate and improve policy outcomes.

Greg Laughlin is currently Principal Policy Adviser at The Australian National Data Service. He has a strong background in the physical sciences (formerly Senior Principal Scientist at the Bureau of Rural Sciences), risk management and statistics. Greg has significant experience and publications in applied modelling, ecology and climatology. He has held positions at the Australian National University, the Resource Assessment Commission, the Australian Bureau of Statistics and the Department of Agriculture, Fisheries and Forestry.