Reply to: Natasha Mauthner and Odette Parry's "Open access digital data sharing: principles, policies and practices"
Libby Bishop, University of Essex

Introduction

In their article, Mauthner and Parry (2012) extend their previous critiques of data reuse by focusing on open access policies. In this review, I assess their primary argument in light of recent evidence, then reflect critically on several of their secondary points, and close by finding common ground with one of their concluding points.

Main argument

Mauthner and Parry present a puzzle: a growing number of international, national, journal, and funder policies encourage the sharing of research data, and many researchers support data sharing in principle, yet the actual practice of formal sharing remains limited. This "limited compliance" does not, they argue, result from infrastructure problems, but from "the failure of these policies to recognise data sharing as a relational practice" (62). That is, researchers have a relational, not objective, relationship with data and it is this epistemological stance, Mauthner and Parry argue, that explains low compliance with institutionalised data sharing policies.

The available data do suggest data sharing rates remain low, despite findings that researchers support data sharing. (For a review of researchers' attitudes and data sharing policies, see Corti et al. 2013). However, the bulk of evidence calls into question whether the epistemology of a "relational stance" to data is the predominant explanation for low rates of sharing. There are many recent studies about data sharing practices (not cited by the authors), and these studies suggest factors other than researchers' attitudes account for limited data sharing (Borgman 2012; Borgman et al. 2012; Fry et al. 2008; Lyon et al. 2010; Open Data Exchange 2011; Piwowar 2011; Tenopir et al. 2011; Wallis et al. 2010; Whyte and Pryor 2011). I will report findings from one of the studies as an example. Neither this survey, nor any others, identified a relational stance to data as a primary cause for not sharing data.

Tenopir et al. (2011) surveyed 1329 scientists across many disciplines about data sharing. Over three-quarters (78%) are willing to share some or all of their data with no restrictions; for social scientists, the figure is still over 70%. However, only about one-third (36%) say others can access their data easily (Table 13; Table; 19; Table 11). These data are consistent with Mauthner and Parry's argument, however the reasons researchers give are quite different.

Table 12. Reasons for not making data electronically available.

	Responses	Percent
Insufficient Time	603	53.6%
Lack of Funding	445	39.6%
Do not Have Rights to Make Data Public	271	24.1%
No Place to Put Data	264	23.5%
Lack of Standards	222	19.8%
Sponsor does not Require	196	17.4%
Do not Need Data	169	15.0%
Other Reasons For Data Not Available	164	14.6%
Should not be Available	162	14.4%

doi:10.1371/journal.pone.0021101.t012

Mauthner and Parry argue that open data advocates misinterpret the problem:

> Resistance to data sharing has therefore been seen as a technical matter of improving data sharing infrastructures (including methodological, ethical, legal and technological elements) in order to increase compliance amongst researchers (48).

The evidence suggests it is researchers themselves who identify precisely these barriers of infrastructure: insufficient time, lack of funding, lack of IP rights, and lack of data facilities were each cited by one-quarter to one-half of all researchers. Only 14% oppose sharing and it is not plausible to assume all of them would give the epistemological argument that Mauthner and Parry offer. Even social scientists, who might be the most likely to cite epistemological barriers, are more concerned with infrastructure. For example, they are less satisfied than average (15% vs. 25-27%) with the "technical" problem of inadequate funding for long-term data management (Tenopir et al., 13).

In sum, the preponderance of evidence from many studies points to willingness to share, yet actual rates of sharing are modest. Moreover, data from surveys and interviews with international researchers across disciplines document a very wide range of explanations for low rates of sharing. In addition to those above could be added: lack of professional rewards for sharing, lack of formal recognition for data creation, and uncertainty about funders' willingness to pay costs. Indeed, Mauthner and Parry's primary source about resistance lists these reasons (Nelson 2009). I entirely agree that for some researchers, epistemology is the primary factor, but the data suggest it plays a more modest role overall than Mauthner and Parry claim. I concur with Borgman (2012, 28) that we are

only just starting to understand why researchers do not share, making it imperative to evaluate all available evidence when attempting to understand researchers' motivations.

A few particulars

In this section, I raise questions about several statements in the article and go on to suggest how the clarifications impinge on the overall argument.

"Underpinning this data sharing movement is the principle of open access" (48).

Precise definitions matter here, because while there is modest consensus that open access means free unrestricted access to publications, there is far less agreement about the meaning of open access to research data itself. Even ardent open access supporters are critical of free and unrestricted access to research data (Harnad 2010). A closer reading of the formal policies reveals numerous qualifications when considering access to data, rather than publications.

The OECD policy stresses the importance of equal, low-cost, timely access, but it is immediately qualified. "Access to, and use of, certain research data will necessarily be limited by various types of legal requirements" e.g., security, privacy, confidentiality, intellectual property, endangered species protection, and legal restrictions (OECD 2007). Similar qualifications are equally prominent in national and funder policies. "RCUK [Research Councils UK] recognises that there are legal, ethical and commercial constraints on release of research data" and qualifies its principle to be "as few restrictions as possible" (RCUK 2012). "The ESRC [Economic and Social Research Council] recognises the fact that under certain circumstances some data may not be suitable for re-use and/or archiving" (ESRC 2010).

The vast majority of policies, funders, and journals recognise and respect diverse qualifications on openness when it comes to data, including registration requirements, regulated access, and embargo periods to enable publishing. Mauthner and Parry acknowledge this in some places, yet at other points, their work equates open access with free and unrestricted, for example, by referring to data being held in "public digital archives" (54). Most data centres are publicly funded, but none provide unregulated public access to all holdings.

Such qualifications on openness matter a great deal. Both survey data and ethnographic studies of research communities of practice have found that objections to open access diminish (or disappear) when modifications to openness are permitted (Whyte and Pryor 2011). Whyte's exemplary ethnographic study proposes a nuanced framework by looking at how willingness to share varies across a continuum of openness (within team to public) and stage in the research cycle.

"widespread institutionalization of open access data sharing policies across research funding and other agencies" (48)

Mauthner and Parry suggest that policies have been in place long enough and applied consistently enough to be deemed institutionalized. This overstates the case and indeed, they acknowledge many of the changes as recent. RCUK unified its principles only in 2012, the National Science Foundation (NSF) in 2011. A meticulous study of US funders found only about half have any detailed guidance on providing access to data, with far fewer recommending that data be shared (Dietrich et al. 2012). The DCC website on funder policies shows an equally mixed picture (DCC 2012).

I do not dismiss the significance of formal policy statements, however, it is important to distinguish policy from practice, and funders do maintain significant respect for researchers' preferences and acknowledge wide variation across communities (Borgman 2012, 28). This is explicit in NSF guidelines:

> What constitutes reasonable data management and access will be determined by the community of interest through the process of peer review and program management (NSF 2010).

So diverse are policies, it is worth considering whether researchers' reluctance to share data may be confusion, rather than resistance. Confusion is most acute when projects involve multiple funders and countries with different requirements.

"...alternative policies may be preferred that institutionalize a more relational approach to data sharing" (62)

Many features of such an approach have great appeal. Here I suggest further considerations for such a model. I use the inelegant terms of 'reuser' and 'depositor' to clarify different roles.

1. What happens when a reuser requests data and a depositor denies the request, on grounds that the reuser will employ a different methodology?

2. What about depositors who don't want to vet requests for data?

3. What happens when the depositor cannot be contacted? If data are destroyed, much valuable work has been lost; if the data are released, there may be disclosure risks.

4. If the depositor retains custody of data, how are the data to be cared for over the long-term? Who will update physical media and file formats?

5. Who pays? Will the research community, or the public, fund bespoke data-sharing regimes? If this is done for only some data, who chooses?

I do not intend to imply Mauthner and Parry could have dealt with all of these questions within this article, but I do wish to point out that these are precisely the questions data

archivists, myself included, have struggled with for some years. We will continue to do so because there are no easy solutions.

"open access data sharing policies embody an instrumental view of data in which data are seen as free-floating public commodities, openly available to anyone wishing to tap into their inherent meanings" (62)

I have dealt with the "public" issue above, but here I focus on the pregnant term "inherent". I am familiar with the positivist/interpretivist debate at issue, and the belief by some that data can have no inherent meaning. This debate will not be resolved here. Elsewhere I have addressed the related discussions about providing context for data reuse, risks of damaging relationships with participants, possible misuse of data, and how the reputations of original researchers can be ethically respected (Bishop 2009; Bishop 2013).

I remain puzzled by Mauthner and Parry's implication that if a researcher holds a relational view of her data, she could not permit data her data to be reused in other ways. Some researchers do not see this as a source of conflict. There are studies held at the UK Data Archive where researchers hold this view, yet have willingly deposited (Henderson 2011). Let me put argument in a different way. I am currently working with scholars who are investigating "paradata", or handwritten replies and marginalia, found on surveys. They are bringing a methodology and stance toward their data quite different to the positivist inclinations of the primary researchers. Would we permit the survey team to refuse to share their data, using as their justification that the reusers are too relational, or insufficiently objective, in their use of data? It is one thing to say, I prefer to use my data in this [relational] way, but quite another to say, my data can be used only in this [relational] way.

At the end, much agreement

"Data sharing takes place within a social and political context marked by power differentials…" (60)

Finally, we reach some common ground, albeit still with some modest qualifications. I agree wholeheartedly that data sharing (and much else) is embedded in power relations. However, I see a less monolithic data-sharing regime than the authors do. (I attempt to be critically reflexive, as I work within that regime.) There are significant projects that engage researchers in consultation about data sharing and open access (Neale and Bishop 2012; RECODE 2013). Low rates of deposit could, in fact, reveal the limited powers of enforcement behind data sharing policies. Mauthner and Parry see the primary danger as loss of control of data; I see other dangers, ones that affect options all researchers should have to share data in optimal ways.

I am concerned that prolonged austerity will result in insufficient funding for infrastructures and support necessary to handle large volumes of diverse data. I teach workshops on managing research data and reuse of data. I am impressed with the large

number of post-graduate and early career researchers who are keen to share their data and use others' data. I worry that, having succeeded in awakening their commitment, future cutbacks will deny researchers their opportunity.

Second, I see a growing tension between the technical complexity needed to handle diverse data and limited resources. I fully concur when Mauthner and Parry point out that bespoke facilities require time and resources. To cite just one example of the kind of technical capacity I have in mind, I believe it is essential that data centres build and maintain capabilities for fine-grained and nuanced systems of regulating access to data that needs additional protection on grounds that it is sensitive or disclosive. As Tenopir et al. show, 65% of researchers would be more likely to share their own data with some way to moderate who could access their data (Tenopir et al., Table 13).

Conclusion

In conclusion, I believe most researchers want to share data, subject to some conditions, but infrastructures are inadequate for this despite continuing improvements. In my personal view, those who prefer not to share their data, for epistemological or other reasons, should not be coerced. But, as is currently the situation, they should be required to explain their preference to withhold data, at least when public funding (direct project or through employment in publicly funded institutions) has been essential for the generation of that data. Otherwise, researchers could withhold publicly funded outputs with no accountability. Such a position is not tenable at any time, and especially not during a period of fiscal austerity. The case of access to research data then, is neither open nor shut. "The challenges are to understand which data might be shared, by whom, with whom, under what conditions, why, and to what effects" (Borgman 2011, 25).

Contact details: ebishop@essex.ac.uk

References

Bishop, Libby. 2013 forthcoming. "The role of moral theory in addressing ethical issues of reusing qualitative data." In *Methodological Innovation Online* (Special Issue on "Ethics: Research Ethics in Challenging Contexts"), edited by Rose Wiles and Janet Boddy.

Bishop, Libby. 2009. "Ethical sharing and re-use of qualitative data." *Australian Journal of Social Issues* 44: 255-272. Accessed 10 July 2011. http://www.data-archive.ac.uk/media/249157/ajsi44bishop.pdf.

Borgman, Christine L. 2012. "The conundrum of sharing research data." *Journal of the American Society for Information Science and Technology* 63: 1059-1078. Accessed July 10, 2013. http://dx.doi.org/10.1002/asi.22634.

Borgman, Christine L., Laura A. Wynholds, Jillian C. Wallis, Ashley Sands, Sharon Traweek. 2012. "Data, data use, and scientific inquiry: Two case studies of data practices." *ACM Press* 19. Accessed July 10, 2013. http://works.bepress.com/borgman/264.

Corti, Lousie, Veerle van den Eynden, Matthew Woollard, Libby Bishop. 2013 forthcoming. *Managing and sharing research data: A guide to good practice.* London: Sage.

DCC. 2012. "Overview of funders' data policies." *Digital curation centre website.* Accessed July 10, 2013. http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies.

Dietrich, Dianne, Trisha Adamus, Alison Miner, Gail Steinhart. 2012. "De-Mystifying the data management requirements of research funders." *Issues in Science and Technology Librarianship.* Accessed July 10, 2013. doi:10.5062/F44M92G2.

ESRC. 2010. "Economic and social research council data policy." Accessed July 10, 2013. http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf.

Fry, Jenny, Suzanne Lockyer, Charles Oppenheim, John Houghton, and Bruce Rasmussen. 2008. "Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes." Loughborough University / Centre for Centre for Strategic Economic Studies. Accessed July 10, 2013. https://dspace.lboro.ac.uk/2134/4600.

Harnad, Steven. 2010. "On not conflating open data (OD) with open access (OA)." American Scientist Open Access Forum. Accessed July 10, 2013. http://openaccess.eprints.org/index.php?/archives/733-On-Not-Conflating-Open-Data-OD-With-Open-Access-OA.html.

Henderson, Sheila, Janet Holland, Sheena McGrellis, Sue Sharpe, Rachel Thomson. 2011. *Inventing Adulthoods, 1996-2006* [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor], July 2011. SN: 5777, http://dx.doi.org/10.5255/UKDA-SN-5777-1.

Lyon, Liz, Chris Rusbridge, Colin Neilson, Angus Whyte. 2010. "Disciplinary Approaches to Sharing, Curation, Reuse and Preservation: DCC SCARP Final Report to JISC." Edinburgh: Digital Curation Centre. Accessed July 10, 2013, http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf.

Mauthner, Natasha S. and Odette Parry. 2013. "Open access digital data sharing: Principles, policies and practices." *Social Epistemology* 27 (1): 47-67. http://dx.doi.org/10.1080/02691728.2012.760663.

Neale, Bren and Libby Bishop. 2012. "The Timescapes archive: A stakeholder approach to archiving qualitative longitudinal data." *Qualitative Research* 12:53-65. Accessed July 10, 2013. http://qrj.sagepub.com/content/12/1/53.

Nelson, Bryn. 2009. "Data sharing: Empty archives." *Nature* 461: 160-163. doi:10.1038/461160a.

NSF. 2010. "Data management & sharing frequently asked questions." National Science Foundation Website. Accessed July 10, 2013. http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp#3.

OECD. 2007. "Principles and guidelines for access to research data from public funding." Accessed July 11, 2013. http://www.oecd.org/science/sci-tech/38500813.pdf.

Open Data Exchange. 2011. "Ten tales of drivers and barriers in data sharing." Opportunities for Data Exchange Project Website. Accessed July 11, 2013.

http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/7836_ODE_brochure_final.pdf.

Piwowar, Heather, A. 2011. "Who shares? Who doesn't? Factors associated with openly archiving raw research data." PLoS ONE 6. Accessed July 12, 2013. http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018657.

RCUK. 2012. "Policy on access to research outputs." Research Councils UK. Website. Accessed July 12, 2013. http://www.rcuk.ac.uk/documents/documents/RCUK%20_Policy_on_Access_to_ Research_Outputs.pdf.

RECODE. 2013. "Policy recommendations for open access to research data project website." Accessed July 12, 2013. http://recodeproject.eu/.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, Mike Frame. 2011. "Data sharing by scientists: Practices and perceptions." PLoS ONE 6. Accessed July 12, 2013. http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101.

Wallis, Jillian. C, Matthew S. Mayernik, Christine L. Borgman, Alberto Pepe. 2010. "Digital libraries for scientific data discovery and reuse: From vision to practical reality." *Proceedings of the 10th Annual Joint Conference on Digital Libraries, Association for Computing Machinery* 333-340. doi10.1145/1816123.1816173.

Whyte, Angus and Graham Pryor. 2011. "Open science in practice: Researcher perspectives and participation." *The International Journal of Digital Curation* 6: 99-213.