What factors do scientists perceive as promoting or hindering scientific data reuse?

Increased calls for data sharing have formed part of many governments' agendas to boost innovation and scientific development. Data openness for reuse also resonates with the recognised need for more transparent, reproducible science. But what are scientists' perceptions about data reuse? ***Renata Gonçalves Curty****,* ***Kevin Crowston****,* ***Alison Specht****,* ***Bruce W. Grant*** *and* ***Elizabeth D. Dalton*** *make use of existing survey data to analyse the attitudes and norms affecting scientists' data reuse. Perceived efficiency, efficacy, and trustworthiness are key; as is whether scientists believe data reuse is beneficial for scientific development, or perceive certain pressures contrary to the reuse of data. Looking ahead, synthesis centres can be important for supporting data-driven interdisciplinary collaborations, and leveraging new scientific discoveries based on pre-existing data.*

"If I have seen further, it was by standing upon the shoulders of giants." This quote, attributed to Sir Isaac Newton, expresses the cumulative and synergistic nature of the growth of science. Intellectual progress and major scientific achievements are built upon the contributions of previous thinkers and discoveries. Thus the scientific enterprise thrives upon openness and collaboration.

The unrestricted sharing of research outputs is increasingly seen as critical for scientific progress. The calls for data sharing in particular, aligned with investment in infrastructures for housing research data, have been part of many governments' agendas to boost innovation and scientific development, while optimising resources. The ability of researchers to access and build upon previous knowledge has thus evolved from elementary access to final published manuscripts and research reports, to the capability of accessing different outputs produced throughout the research lifecycle, including digital data files.

There have been a number of promising developments in funding bodies' policies promoting and requesting compliance with data sharing requirements to ensure preservation and access to scientific data for further reuse. In the US, the Data Observation Network for Earth (DataONE), supported by the National Science Foundation (NSF), is committed to broadening education on data-related issues (e.g. data documentation, data citation), as well as to provide standards/guidelines and sustainable cyberinfrastructure to secure openness, persistence, robustness, findability, and accessibility to environmental science data.

Image credit: Can hierarchy and sharing co-exist? by opensource.com. This work is licensed under a CC BY-SA 2.0 license.

Date originally posted: 2018-03-20

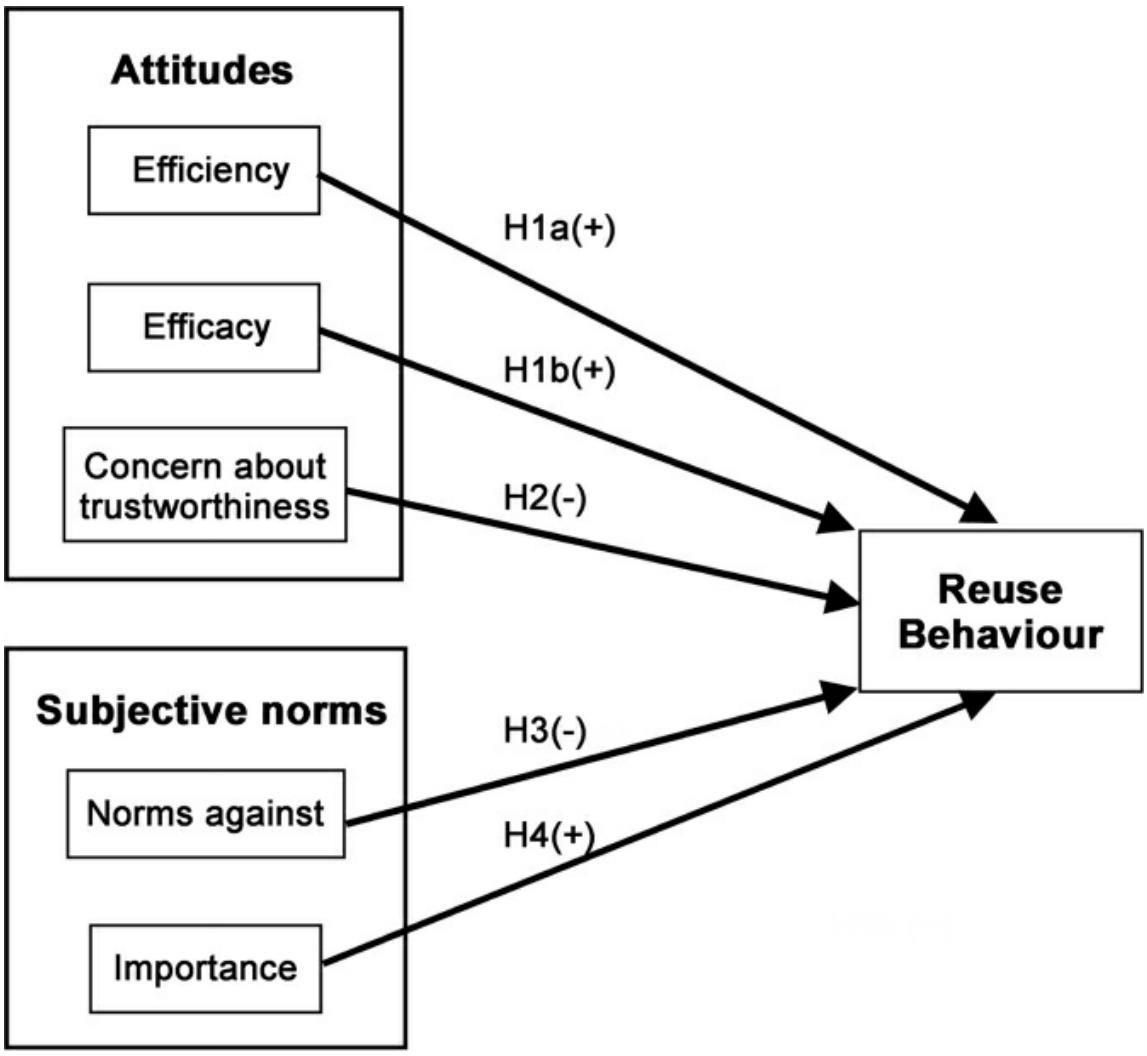
Permalink: http://blogs.lse.ac.uk/impactofsocialsciences/2018/03/20/what-factors-do-scientists-perceive-as-promoting-or-hindering-scientific-data-reuse/

Blog homepage: http://blogs.lse.ac.uk/impactofsocialsciences/

Data openness for reuse resonates well with the recognised need for more transparent and reproducible science. However, despite the growing availability of research data for potential reuse, key questions are often left unanswered. What are scientists' perceptions about the reuse of data? Do they want to use the data being shared? And, more specifically, what factors do scientists perceive as promoting or hindering scientific data reuse?

Aiming at exploring this fairly untapped subject, our recent research paper investigates these questions. We conducted an inferential analysis based on selected questions of a worldwide survey developed and administered by the DataONE Usability and Assessment Working Group, making our paper itself an example of data reuse.

We adopted the widely known Theory of Reasoned Action (TRA) to conceptualise the effect of attitudes and perceptions on self-reported reuse behaviour. Based on the proposition that scientists' attitudes (i.e. perceived benefits and risks) and subjective norms (i.e. perceived pressures) towards data reuse influence their own data reuse behaviour, we followed a two-step approach. First, we performed an exploratory factor analysis in order to develop scales for our two theoretical constructs (attitudes and norms). This stage allowed us to identify five factors which were later plugged to the theoretical constructs. The construct *attitudes* was comprised of three different factors: *perceived efficiency* (how effortless and not time-consuming the data reuse process is perceived to be), *perceived efficacy* (how effective the outcome of reusing data is perceived to be), and *concern about data trustworthiness* (the extent to which data produced by others are not reliable). Meanwhile, we examined *subjective norms* through two different factors: *perceived norms against data reuse behaviour* (the extent to which scientists perceive there are some pressures contrary to the reuse of data), and *perceived importance of data reuse* (the extent to which data reuse is considered beneficial for scientific development). Based on such factors and their corresponding constructs, we hypothesised the following correlations:



Date originally posted: 2018-03-20

Permalink: http://blogs.lse.ac.uk/impactofsocialsciences/2018/03/20/what-factors-do-scientists-perceive-as-promoting-or-hindering-scientific-data-reuse/

Blog homepage: http://blogs.lse.ac.uk/impactofsocialsciences/

Figure 1: Research model. This figure was originally published in the authors' article "Attitudes and norms affecting scientists' data reuse" and is published under a CC BY 4.0 license.

H1a: Perceived efficiency of data reuse will positively correlate with data reuse; H1b: Perceived efficacy of data reuse will positively correlate with data reuse; H2: Concerns about the trustworthiness of data will negatively correlate with data reuse; H3: Perceived norms against data reuse will negatively correlate with data reuse; and H4: Perceived importance of data reuse will positively correlate with data reuse.

All our hypotheses were supported by empirical data, except for the anticipated negative correlation between concerns about data trustworthiness and reuse behaviour (H2). To our surprise, we found that scientists who expressed concerns about the credibility of data produced by others did not necessarily refrain from reusing others' data.

This result did not change according to participants' data management experience. We assessed experience by splitting the sample into two groups: those knowledgeable about metadata vs. those without knowledge about metadata, which we initially assumed could influence their ability and rigour when judging data credibility.

We can highlight other findings revealed by the statistical analysis:

- Self-reported data sharing behaviour was only weakly correlated to data reuse behaviour, meaning data sharers are not necessarily reusers and vice-versa.
- Those who reported use of models and remote-sensed data are more keen to reuse data.
- For those with developed data management practices, the perceived subjective norms against data reuse were not an impediment for data reuse; this allowed us to infer that those with more ability for reusing data also feel that they are more capable of minimising or overcoming potential challenges associated with the reuse of data in their discipline.
- Perceived efficiency was linked with actual data reuse behaviour only for those who were knowledgeable about metadata. This indicated that the process is perceived as less time-consuming and laborious by those who are more capable of easily decoding and interpreting data produced and documented by others.

We can also add that the process of reusing the pre-existing survey data posed some important limitations to our study, as there were times we wished we had more questions and also that some of them had been worded differently in the original survey. We recognise the need for future studies to support more conclusive statements and to disentangle some of the remaining questions about scientists' data reuse behaviour. Because of that, we encourage future studies to disentangle the constructs of the subjective norms.

We suggest that future research could include attitudinal constructs in the model, as well as analyse data using a more sophisticated statistical analysis (e.g. structural modelling and path analyses) to see how our constructs correlate not only with the outcome variable (e.g. data reuse behaviour), but among themselves. Another opportunity is to see if and how these factors play different roles across disciplines. We did not have enough observations to run statistically reliable comparative tests across and between disciplines.

Both the literature and our own experience reveal the complexity of reusing data produced by others. The reuse of data becomes even more difficult if we consider situations where scientists combine two or more disparate data sources from different fields. In such cases the reuse process requires more sophisticated data-mashing capability as well as the ability for reusers to navigate into different knowledge domains to properly connect with and infer results from the data. Understanding the different issues surrounding the reuse of data will help to optimise the required supporting cyberinfrastructure. We argue that synthesis centres are of great importance for supporting data-driven interdisciplinary collaborations, and leveraging new scientific discoveries based on pre-existing data. These facilities can help to minimise frictions associated with data reuse, assisting scientists to look further from each others' shoulders.

This blog post is based on the authors' article, "Attitudes and norms affecting scientists' data reuse", published in PLoS ONE (DOI: 10.1371/journal.pone.0189288).

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy if you have any concerns on posting a comment below.

Date originally posted: 2018-03-20

Permalink: http://blogs.lse.ac.uk/impactofsocialsciences/2018/03/20/what-factors-do-scientists-perceive-as-promoting-or-hindering-scientific-data-reuse/

Blog homepage: http://blogs.lse.ac.uk/impactofsocialsciences/

About the authors

Renata Gonçalves Curty *is an Assistant Professor of the Information Science Department at the Universidade Estadual de Londrina (State University of Londrina), Brazil. She received her PhD in Information Science and Technology from the School of Information Studies, Syracuse University, in 2015. She is particularly interested in scientific data management, platforms and practices for collaborative data sharing and reuse, as well as other aspects surrounding the sociology of science and scholarly communication. Her ORCID ID is: 0000-0002-4615-6030.*

Kevin Crowston *is a Distinguished Professor of Information Science in the School of Information Studies at Syracuse University and is currently serving as Associate Dean for Research. He received his PhD (1991) in Information Technologies from the Sloan School of Management, Massachusetts Institute of Technology (MIT). His research examines new ways of organising made possible by the extensive use of information and communications technology. Specific research topics include science data management and work practices and technology support for citizen science research projects. His ORCID ID is: 0000-0003-1996-3600.*

Alison Specht *is the director of the French analysis and synthesis centre, CESAB, part of the Foundation for Research on Biodiversity. She spent over 25 years as a research academic, and her interests are the nature and causes of ecosystem biodiversity, in particular that of perennial and groundwater-dependent ecosystems. She is passionate about the conservation of data, communication, and facilitating trans-disciplinary research.*

Bruce W. Grant *is a Full Professor of Biology and Environmental Science at Widener University. Over the past 22 years, he has directed his scholarly activity to research on urban ecology, research on the pedagogy of academic service learning, and research on undergraduate ecological education.*

Elizabeth D. Dalton *is an Assistant Professor of Communication Studies at Middle Tennessee State University. She received her PhD in Communication and Information from the University of Tennessee, Knoxville in 2014. Her research is primarily in the areas of scholarly communication and health communication. Specifically within scholarly communication, she examines topics such as academics' attitudes toward open access publishing and data sharing. Within health communication, she examines disclosure processes and communication about pain in nurse-patient communication. Her ORCID ID is: 0000-0001-5525-361X.*

Date originally posted: 2018-03-20

Permalink: http://blogs.lse.ac.uk/impactofsocialsciences/2018/03/20/what-factors-do-scientists-perceive-as-promoting-or-hindering-scientific-data-reuse/

Blog homepage: http://blogs.lse.ac.uk/impactofsocialsciences/