








ARTICLE



<https://doi.org/10.1057/s41599-023-01831-7>

OPEN

Perceived benefits of open data are improving but scientists still lack resources, skills, and rewards

Joshua Borycz ^{1✉}, Robert Olendorf², Alison Specht ³, Bruce Grant⁴, Kevin Crowston ⁵, Carol Tenopir ⁶, Suzie Allard ⁶, Natalie M. Rice⁶, Rachael Hu⁷ & Robert J. Sandusky^{8,9}

Addressing global scientific challenges requires the widespread sharing of consistent and trustworthy research data. Identifying the factors that influence widespread data sharing will help us understand the limitations and potential leverage points. We used two well-known theoretical frameworks, the Theory of Planned Behavior and the Technology Acceptance Model, to analyze three DataONE surveys published in 2011, 2015, and 2020. These surveys aimed to identify individual, social, and organizational influences on data-sharing behavior. In this paper, we report on the application of multiple factor analysis (MFA) on this combined, longitudinal, survey data to determine how these attitudes may have changed over time. The first two dimensions of the MFA were named *willingness to share* and *satisfaction with resources* based on the contributing questions and answers. Our results indicated that both dimensions are strongly influenced by individual factors such as perceived benefit, risk, and effort. *Satisfaction with resources* was significantly influenced by social and organizational factors such as the availability of training and data repositories. Researchers that improved in *willingness to share* are shown to be operating in domains with a high reliance on shared resources, are reliant on funding from national or federal sources, work in sectors where internal practices are mandated, and live in regions with highly effective communication networks. Significantly, *satisfaction with resources* was inversely correlated with *willingness to share* across all regions. We posit that this relationship results from researchers learning what resources they actually need only after engaging with the tools and procedures extensively.

¹ Vanderbilt University, Nashville, TN, USA. ² North Carolina State University, Raleigh, NC, USA. ³ TERN, University of Queensland, Brisbane, Australia. ⁴ Widener University, Chester, PA, USA. ⁵ Syracuse University, Syracuse, NY, USA. ⁶ University of Tennessee, Knoxville, TN, USA. ⁷ California Digital Library, Oakland, CA, USA. ⁸ University of Illinois at Chicago, Chicago, IL, USA. ⁹ Deceased: Robert J. Sandusky. ✉email: joshua.borycz@vanderbilt.edu

Introduction

Open data practices have become increasingly important as highly collaborative projects are needed to resolve important issues in science, such as biodiversity loss, climate change, and infectious diseases. Data-centered science, with a focus on open data, is increasingly seen as fundamental to solving these massive, interdisciplinary challenges and to increase the return on investment of research (Data sharing and the future of science, 2018). This requires a cultural shift in scientific practices that cannot take place without the concerted effort of all researchers, their institutions, and funding sources. In the past decade, the impact of sharing data on research quality and scientific progress has been studied extensively (Milham et al., 2018; Perez-Riverol et al., 2019; Zhang et al., 2021). Many international initiatives have been instrumental in illuminating sound data practices, workflow reproducibility (Baker, 2016; Fidler et al., 2017; Open Science Collaboration, 2015), and developing tools and guidelines to enable data sharing, such as the Enabling FAIR Data project (COPDESS, 2018; Wilkinson et al., 2016), GO FAIR (COPDESS, 2018; David et al., 2020), and DataONE (Michener et al., 2012). There has also been increasing interest in probing the attitudes of scientists toward data sharing and reuse to determine the social and institutional barriers that prevent it. Lack of recognition and rewards for publishing research data has been found to be a key barrier to sharing data (David et al., 2020; Fecher et al., 2015), and there is a positive relationship between the behavior and attitudes of researchers on data reuse (Curty et al., 2017; Ludäscher, 2016). Scientists who reported that reusing others' data increased their own efficiency and saved them time, were more likely to reuse data shared by others. The availability of institutional resources supporting data sharing and reuse (e.g., education, technology, expert help) and more generally available resources such as repositories and metadata creation tools have been found to greatly increase rates of data reuse (Kim and Yoon, 2017).

Data management organizations and institutions must make assumptions about which of these factors can be used to influence researchers' open data practices. DataONE (the Data Observation Network for Earth) was created in 2009 with the idea that attitudes and behaviors could be changed by (a) providing easy-to-use data-sharing infrastructure and training, and (b) helping influential researchers within the natural sciences adopt open data practices (Michener et al., 2012). Its initial focus was on the biological and environmental sciences. DataONE was funded by a large federal grant to create cyberinfrastructure that would address barriers preventing more open, global, and reproducible research (Michener et al., 2012). Part of the charter of DataONE was to determine what these barriers are by surveying researchers in different fields and developing personas to help address their needs.

The Findable, Accessible, Interoperable, Reusable (FAIR) initiative uses a multi-faceted approach to assess data sharing and guide training at research institutions (COPDESS, 2018; Wilkinson et al., 2016). Large institutions like the Office of Science and Technology Policy (OSTP) and the National Institutes for Health (NIH) in the United States are now mandating open data behaviors, using the FAIR principles as a guideline for compliance (Holdren, 2013; NIH, 2023). Such mandates highlight the importance of knowing whether these initiatives have had an impact and which factors have been influential on the adoption of open data behaviors.

Surveying researcher attitudes towards data sharing and reuse are one method of determining the effect of these research initiatives on the scientific culture. Three large surveys were designed and implemented for DataONE by a research group at the University of Tennessee and the results were published in 2011, 2015, and 2020 (Tenopir et al., 2011, 2015, 2018, 2020). These surveys have been highly influential (2011 survey has 754

citations, 2015 has 258 citations, and 2020 has 65 citations according to Scopus (date accessed: 2023/04/27)) and continue to serve as a benchmark for surveys of researchers' data sharing attitudes and behaviors. The multidisciplinary team of researchers in the DataONE Usability and Assessment Working Group provided input to the surveys. The first two surveys were open from October 2009 to July 2010 and October 2013 to March 2014, respectively (Tenopir et al., 2011, 2015). The third survey was administered in two instances. This survey was first sent to members of the American Geophysical Union and was open from March 2017 to March 2018 (Tenopir et al., 2018). It was then sent to a global audience from December 2017 to May 2018 and the two instances were amalgamated in a final analysis (Tenopir et al., 2020). In each survey, scientists were asked about their general views on data sharing and to report the issues that prevented or encouraged them to share data. The first study found that lack of access to and preservation of data were significantly hindering research progress, but researchers would be willing to share data if they were to receive proper citations or could place some conditions on access. The second survey, as well as work by Curty et al. (2017), revealed an increase in the willingness to share data and also a perceived risk of sharing data. The last study demonstrated continuing favorable attitudes toward open data; however, it was found that sharing was impeded by a lack of resources or a lack of perceived benefits of data sharing (Tenopir et al., 2020). The degree of cross-comparison between surveys was mainly qualitative. We are working to connect them quantitatively.

To determine the importance of individual, social, and organizational factors on data sharing, as well as how attitudes to data sharing and reuse have changed over time, this paper describes an integrated analysis of the three surveys from Tenopir et al. (2011, 2015, 2018, 2020). We use two theoretical frameworks, the Theory of Planned Behavior (Ajzen, 1991) and the Technology Acceptance Model (Davis, 1989) to determine which of these three factors each question within the survey relates to. The impact that world region, research domain, work sector, and funding source have on data-sharing attitudes is also analyzed and compared within this framework.

Theory

To examine the changing behaviors of the survey respondents, we employed several conceptual frameworks that have been used to explain underlying influences on human attitudes and behaviors. The theory of reasoned action (TRA) links the perception of social norms to behavioral intention (Ajzen and Fishbein, 1980). The TRA has been used to design and analyze several surveys on data sharing and reuse behaviors (Kim and Zhang, 2015). An extension of TRA, the theory of planned behavior (TPB) (Ajzen, 1991), incorporates an individual's perception that a goal is achievable and distinguishes between career benefits and risks. TPB has been used in several multi-level frameworks to describe data sharing and reuse attitudes and behaviors after survey data revealed the importance of ease-of-use and time commitments (Kim and Stanton, 2016; Kim and Yoon, 2017). The technology acceptance model (TAM) is another extension of the TRA that was popularized in the fields of information systems and communication (Davis, 1989). The TAM adds the variable of perceived usefulness, assuming that individuals are more likely to be open to behaviors that are beneficial to them. Individual and social factors are often contextualized with facilitating conditions (e.g., institutional support, training repositories, technical services, mandates) as these have a strong impact on research behavior in many research fields (Kim and Zhang, 2015; Talukder, 2012;

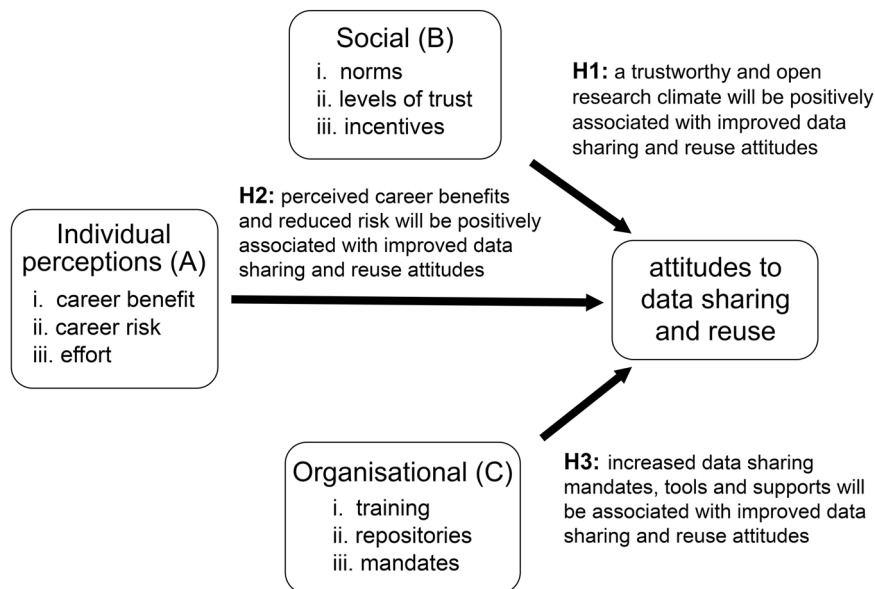


Fig. 1 Theoretical Framework. Framework for researchers' attitudes toward open data behavior showing the key elements of hypothesized influence, and hypotheses (H = hypothesis).

Venkatesh et al., 2003; Yoon and Kim, 2017). It is clear from the literature on data sharing and reuse that individual factors related to reputation and plausibility of success, social norms, and organizational commitment need to be considered (Kim and Zhang, 2015; Talukder, 2012; Venkatesh et al., 2003; Yoon and Kim, 2017).

It was clear to us, based on the work of Tenopir et al. (2011, 2015, 2018, 2020) that the three main influences on attitudes toward data sharing and reuse were individual motivations, social influences, and organizational support. There were several psychological and social theories that incorporated some of these influences, but capturing the complexities of all three required combining elements from multiple theories. As such, we have developed a theoretical framework which is a combination of TPB and TAM. The key elements we hypothesize will influence attitudes to data sharing and re-use are:

1. *Social influences (TAM)*, including the social norms around data sharing of a particular field, trust in colleagues' data and/or academic integrity, and incentives to share and reuse data.
2. *Organizational influences (TAM)*, including the availability of training, tools and repositories and funding mandates and guidelines provided by local or federal institutions.
3. *Individual influences (TAM and TPB)*, including effort associated with data sharing (TAM and TPB) and career benefits and risks (TPB).

We hypothesize that attitudes to data sharing and re-use will be positively related to:

H1: Increased perception that the research climate is trustworthy and open will be associated with improved data sharing and reuse attitudes.

H2: Perceived career benefits and reduced risk will be positively associated with improved data sharing and reuse attitudes.

H3: Increased sharing mandates, tools and support will be associated with improved data sharing and reuse attitudes. (Fig. 1).

Methods

Data cleaning. To detect attitudinal changes across time, questions that were comparable between the three Tenopir et al. surveys were selected and combined into a single longitudinal dataset. For the sake of simplicity in this paper, each survey is labeled by publication year. In several cases, there were slight changes in the wording of the question and/or choice offered. There were also questions removed and new ones added between surveys. For instance, as a result of comments from survey participants, skipped answers in previous surveys, and discussions with members of specific fields on their particular use case, some demographic questions were removed and new ones added between surveys (e.g., career stage, age, gender). We only included data for questions that were sufficiently consistent across the three surveys.

The data cleaning process for this project required making sure that the answer types and ranges of options were made consistent between surveys. When *yes* or *no* answers were changed to *yes*, *no*, or *not sure*, we changed *not sure* to blank responses (NA). In the cases where the Likert scales had both *neither agree nor disagree* and *not sure* as answers for some questions but not others, we combined them into just *neither agree nor disagree*. We also cleaned each dataset for non-response. Specifically, for each survey, respondents who did not answer more than 5 questions were excluded to prevent 'no answer' from dominating. Note that a higher proportion of people in the 2015 survey skipped questions in the survey (Table 1).

The cleaned dataset comprised 42 questions and responses from 3197 individuals (1214 in 2011; 539 in 2015; and 1444 in 2020; Table 1, raw data available on Zenodo (Olendorf et al., 2022)). Four of the 42 questions concerned the demographics of the respondents: region of the world, scientific domain, work sector, and funding agency. For these questions, answer types and scales were adjusted to ensure comparability across surveys. Countries were aggregated into six regions: Africa and the Middle East, Asia and Southeast Asia, Australia and New Zealand, Europe and Russia, Latin America, and the USA and Canada. The countries were clustered in the same way as they were in the surveys by Tenopir et al. (2011, 2015, 2018, 2020). These divisions were based on the International Union for Conservation of Nature (IUCN) regions (IUCN, 2022) (Table S1). This clustering

Table 1 Summary of survey responses according to respondent demographics.

Category	Sub-category	2011 Total (%)	2015 Total (%)	2020 Total (%)
Region	Africa & Middle East	25 (2.1)	44 (8.2)	99 (6.9)
	Asia & Southeast Asia	63 (5.2)	32 (5.9)	130 (9.0)
	Australia & New Zealand	22 (1.8)	11 (2.0)	72 (5.0)
	Europe & Russia	178 (14.7)	78 (14.5)	358 (24.8)
	Latin America	42 (3.5)	32 (5.9)	77 (5.3)
	USA & Canada	879 (72.4)	341 (63.3)	696 (48.2)
Domain	Natural Science	651 (53.6)	293 (54.4)	397 (27.5)
	Physical Science	298 (24.5)	115 (21.3)	802 (55.5)
	Information Science	49 (4.0)	54 (10)	54 (3.7)
	Social Science	161 (13.3)	60 (11.1)	31 (2.1)
	Other	55 (4.5)	17 (3.2)	159 (11.0)
	Work sector	Academic	975 (80.3)	411 (76.3)
	Commercial	30 (2.5)	15 (2.8)	45 (3.1)
	Government	155 (12.8)	89 (16.5)	252 (17.5)
	Non-Profit	32 (2.6)	18 (3.3)	58 (4.0)
	Other	20 (1.6)	6 (1.1)	36 (2.5)
Funding agency	Corporation	38 (3.1)	13 (2.4)	43 (3.0)
	Federal & national government	819 (67.5)	355 (65.9)	1039 (72)
	Private foundation	67 (5.5)	28 (5.2)	41 (2.8)
	State, regional & local government	195 (16.1)	57 (10.6)	127 (8.8)
	Other	89 (7.3)	85 (15.8)	193 (13.4)
Totals	Original Total	1329	1015	2184
	Total Used	1214	539	1444

The number of respondents to each demographic category and the percentage of the total for that category are shown.

allowed us to make quantitative comparisons. A similar clustering was made for the other categories, for example, the Physical Sciences included respondents in physics, chemistry, and engineering, and the Natural Sciences included those in biology, medicine, and zoology (Table S2). Work sectors were divided into Academic, Corporate, Government, Non-Profit, and Other. Funding agencies were divided into five categories: Corporation, Federal & national government, Private foundation, State, regional or local government, and Other. A corporation refers to for-profit companies, while a private foundation refers to a privately funded non-profit (e.g., the Bill & Melinda Gates Foundation).

The majority of the variables concern attitudes or beliefs about data sharing. The attitudinal questions required either a binary (*yes* or *no*) response or a selection on a Likert scale (Tables S3 and S4).

The variations in the portions of responses from each survey demographic for the 2011 and 2015 surveys were 2.9 ± 2.5 (mean \pm standard deviation), 7.4 ± 9.1 for the 2015 and 2020 surveys, and 7.9 ± 8.6 for the 2011 and 2020 surveys.

Data analysis. To determine the factors that explain the most variance in the survey responses we performed a multiple factor analysis (MFA) with the *Factominer* (Husson et al., 2020) and *Factoextra* (Kassambara and Mundt, 2020) packages in R. MFA reduces the dimensionality of a large set of variables to maximize the explanatory value with the fewest variables possible. The input for the MFA is the list of survey questions as columns and survey respondents as rows. The MFA outputs orthogonal dimensions in the order of how much statistical variance within the survey responses each dimension explains. Analysis of variance (ANOVA) using the R stats package was then performed to assess the effect of demographics on each factor, followed by Tukey's honest significant difference (TukeyHSD) test to determine the significance of the differences between the answers of the various demographic groups. All survey data and code for these analyses are available on Zenodo (Olendorf et al., 2022).

Degrees of freedom, the sum of squares, mean squares, *F*-values, and *p*-values for all ANOVAs performed in this experiment are

provided in Table S14. Bartlett's test for homogeneity of variance and residual tests were computed for each survey and each demographic variable (Table S15). Non-homogeneous variance could result from differences in the samples or the sampling methods across the three surveys (Tenopir et al., 2011, 2015, 2018, 2020). Sources of variation due to region and funding agency prevented analysis of these variables. One contributing factor was undoubtedly the fact that all the survey question responses were categorical variables with limited response ranges (*yes/no* or 1–5 Likert), which led to non-homogeneous variances. The standardized residual plots generally showed zero correlation with either dimensions one or two for each of the surveys and for all combined. The normal-QQ plots (Figs. S2–S33) generally showed linearity within the central four quantiles of responses, which suggests that these data plausibilities derive from the same distributions.

Results

Defining the dimensions. The multiple-factor analysis of the cleaned survey data from the three DataONE surveys in 2011, 2015, and 2020 (Tenopir et al., 2011, 2015, 2018, 2020) showed significant loadings on just two dimensions. These explained 4.0% and 2.9% of the variance, respectively, and were the only ones greater than the average percentage of explained variance of 2.7% (Fig. S1). The total mean and standard errors across all three surveys for all demographics for each dimension are shown in Table S5. The significant Tukey's HSD means and *p*-values for all demographics in the three surveys combined are shown in Tables S6–S9 and for each survey individually in Tables S10–S13. In the following material, Tukey's means and ANOVA significance levels are presented.

The questions clustered within Dimension 1 all related to openness to sharing data and the risks/benefits of sharing data (Table 2). The questions illustrated in Table 2, although significant, are not representative of the bulk of the responses (Table S3). The answers to these questions negatively correlate with dimension 1 when more openness to sharing is expressed

Table 2 Top 5 contributing questions for dimension 1 of the MFA.

Question	Answer type	Contribution (% variance)	Correlation	Framework element (Fig. 1)
I would be willing to place at least some of my data into a central data repository with no restrictions.	Likert (1-5)	7.94	-0.54***	Social (B ii)
The data provider is given and agrees to a statement of uses to which the data will be put.	Yes, no	7.83	0.63***	Individual Risk (A ii)
Results based (at least in part) on the data could not be disseminated in any format without the data provider's approval	Yes, no	7.57	0.63***	Individual benefit (A i)
Mutual agreement on reciprocal sharing of data	Yes, no	7.03	0.60***	Individual benefit (A i)
I would be willing to share data across a broad group of researchers who use data in different ways.	Likert (1-5)	6.90	-0.54***	Social (B ii)

This dimension was interpreted as *willingness to share* due to the questions highlighted. Likert scale, 1 = strongly agree, 2 = agree somewhat, 3 = neither agree nor disagree, 4 = disagree somewhat, and 5 = strongly disagree. ****p* < 0.001.

Table 3 Top 5 contributing questions for dimension 2 of the MFA.

Question	Answer type	Contribution (% variance)	Correlation	Framework elements
I am satisfied with the process for storing my data beyond the life of the project (long-term)	Likert (1-5)	6.85	-0.46***	Organizational (C ii)
I would be willing to place at least some of my data into a central data repository with no restrictions.	Likert (1-5)	6.20	0.16***	Organizational (C ii)
I would be willing to share data across a broad group of researchers who use data in different ways.	Likert (1-5)	5.64	0.06**	Social (B ii)
Lack of access to data generated by other researchers or institutions is a major impediment to progress in science.	Likert (1-5)	5.42	0.11***	Social (B ii)
Data may be misinterpreted due to complexity of the data.	Likert (1-5)	5.04	0.14***	Individual risk (A ii)

This dimension was interpreted as *satisfaction with resources*. For the Likert scale, 1 = strongly agree, 2 = agree somewhat, 3 = neither agree nor disagree, 4 = disagree somewhat, and 5 = strongly disagree. ***p* < 0.01, ****p* < 0.001.

and positively correlate when concern over risks of data sharing is expressed which led us to interpret dimension 1 as *willingness to share*.

The top five questions clustered with Dimension 2 related to processes, databases, data availability, and risk (Table 3). The questions illustrated in Table 3, although significant, are not representative of the bulk of the responses (Table S4). The answers negatively correlate with dimension 2, indicating more agreement, when the questions relate to satisfaction with processes, databases, and data availability, and positively correlate with dimension 2, indicating more disagreement, when the questions relate to risks and conditions on data. Hence, we interpreted dimension 2 as *satisfaction with resources*.

As illustrated in Tables 2 and 3, we used the data sharing and reuse framework (Fig. 1) to categorize all the questions included in the MFA (Table 4). Individual benefits and risks are the strongest elements for each of the first two dimensions, but social and organizational elements play a more important role in researchers' satisfaction with resources. Organizational influences have a much greater impact on satisfaction with resources.

Impact of demographics on willingness to share and satisfaction with resources. The effect of world region on *willingness to share* scientific data varied with the region and with the survey. Overall, respondents from Australia & New Zealand (0.288), USA & Canada (0.238), and Europe & Russia (-0.037) were more willing to share than those from Asia & Southeast Asia (-1.020) and Africa & Middle East (-0.939). Those from USA & Canada were more willing to share than respondents from Latin America

Table 4 Contributions to the first two dimensions of the MFA derived by assigning each question to an element from the framework in Fig. 1.

	Dimension 1 (Willingness to share)	Dimension 2 (Satisfaction with resources)
Individual	68.2	44.1
Social	27.5	25.0
Organizational	4.3	30.9

Percent contributions of each framework element to each dimension are shown.

(-0.468) (Table 5). Respondents coming from the USA and Canada showed a significant increase in *willingness to share* between 2011 and 2015 (-0.446 to 0.635, *p* < 0.001) and across all three surveys (-0.446 to 0.908, *p* < 0.001) as did those in Europe & Russia (2011-2020: -0.644 to 0.264, *p* < 0.001) (Table S10 and Fig. 2A). No other significant differences in *willingness to share* by region were observed between surveys.

Satisfaction with resources was significantly greater for respondents from Africa & Middle East (0.770), Asia & Southeast Asia (0.587), and Latin America (0.510) than those from USA & Canada (-0.163) (Table 6 and Fig. 2b). The same suite of countries were more satisfied with resources than Europe and Russia (-0.086). There were no significant differences observed between regions for individual surveys or between survey events.

Across regions and surveys, mean *willingness to share* was inversely correlated with satisfaction with resources (*R*² = 0.8, *p*-value = 0.017, *F*-value = 15.67). There were also significant

Table 5 Significant pairwise differences between willingness to share for world regions computed with ANOVA and TukeyHSD.

Dimension 1 (Willingness to share)	Mean	Africa & Middle East	Asia & Southeast Asia	Latin America
Mean	-	-0.940	-1.020	-0.468
Australia & New Zealand	0.288	-1.227***	-1.308***	-
Europe & Russia	-0.037	-0.902***	-0.983***	-
USA & Canada	0.238	-1.178***	-1.258***	-0.707**

p < 0.01, *p < 0.001.

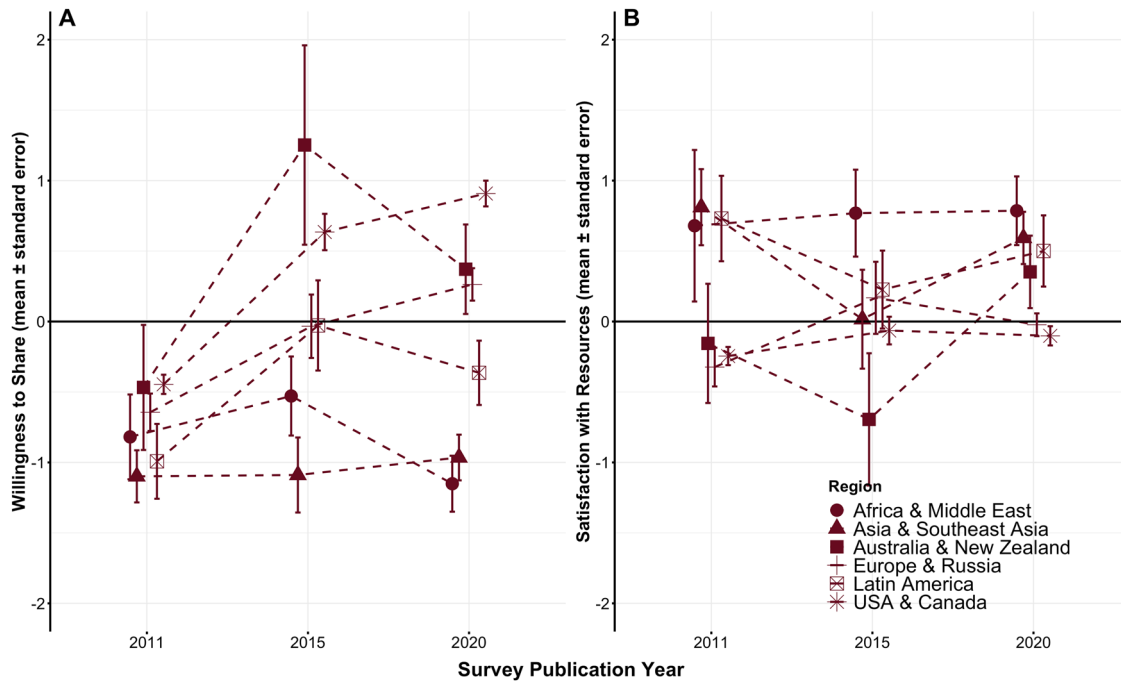


Fig. 2 Region MFA. Willingness to share (A) and satisfaction with resources (B) by region (Erway and Rinehart, 2016; European Union, 2013; Evans, 2010a). The world is divided into six regions (Africa & Middle East, Asia & Southeast Asia, Australia & New Zealand, Europe & Russia, Latin America, USA & Canada) based on previous work.

Tables 6 Significant pairwise differences between satisfaction with resources for world regions computed with ANOVA and TukeyHSD.

Dimension 2 (Satisfaction with resources)	Mean	Europe & Russia	Latin America	USA & Canada
Mean	-	-0.086	0.510	-0.163
Africa & Middle East	0.770	-0.856***	-	-0.933***
Asia & Southeast Asia	0.587	-0.673***	-	-0.750***
Europe & Russia	-0.086	-	0.596**	-
Latin America	0.510	-	-	-0.674***

p < 0.01, *p < 0.001.

correlations for the 2011 ($R^2 = 0.82$, p -value = 0.012, F -value = 18.68) and 2020 ($R^2 = 0.84$, p -value = 0.011, F -value = 20.44) papers, but the ANOVA results for satisfaction with resources for these individual papers were not significant, so these results are not included.

For domains of study, physical science (0.291) and information scientists (0.660) were more willing to share than natural scientists (-0.128) (Table 7). Social scientists (-0.592) were far less willing to share than researchers in the physical sciences, natural sciences, and information sciences. *Willingness to share* increased substantially across time in the physical sciences (e.g., physics, chemistry, engineering) (2011–2015: -0.566 to 0.462, $p < 0.01$; 2011–2020: -0.566 to 0.585, $p < 0.001$) and natural

sciences (e.g., biology, medicine, zoology) (2011–2015: -0.485 to 0.179, $p < 0.01$; 2011–2020: /0.485 to 0.216, $p < 0.001$; Fig. 3A). Satisfaction with resources, however, only increased significantly for the social scientists (2011–2020: -0.259 to 1.026, $p < 0.05$; Fig. 3B) (Table S11). There was no correlation across domains between *willingness to share* and satisfaction with resources.

For *work sector*, people working for the government (0.756) were much more likely to share their data than people working in the academic (-0.046) or commercial (-0.264) sectors (Table 8). The effect of work sector on respondents' *willingness to share* showed a significant increase over time in the government (2011–2015: -0.197 to 1.020, $p < 0.01$; 2011–2020: -0.197 to 1.249, $p < 0.001$) and academic sectors (2011–2015: -0.604 to

Table 7 Significant pairwise differences between willingness to share for domains of study computed with ANOVA and TukeyHSD.

Dimension 1 (Willingness to share)	Mean	Natural science	Social science	Other
Mean	-	-0.128	-0.592	-0.531
Physical science	0.291	-0.420***	-0.883***	0.821***
Information science	0.660	-0.791***	-1.252***	1.191***
Social science	-0.592	0.461*	-	-
Other	-0.531	0.400***	-	-

* $p < 0.05$, *** $p < 0.001$.

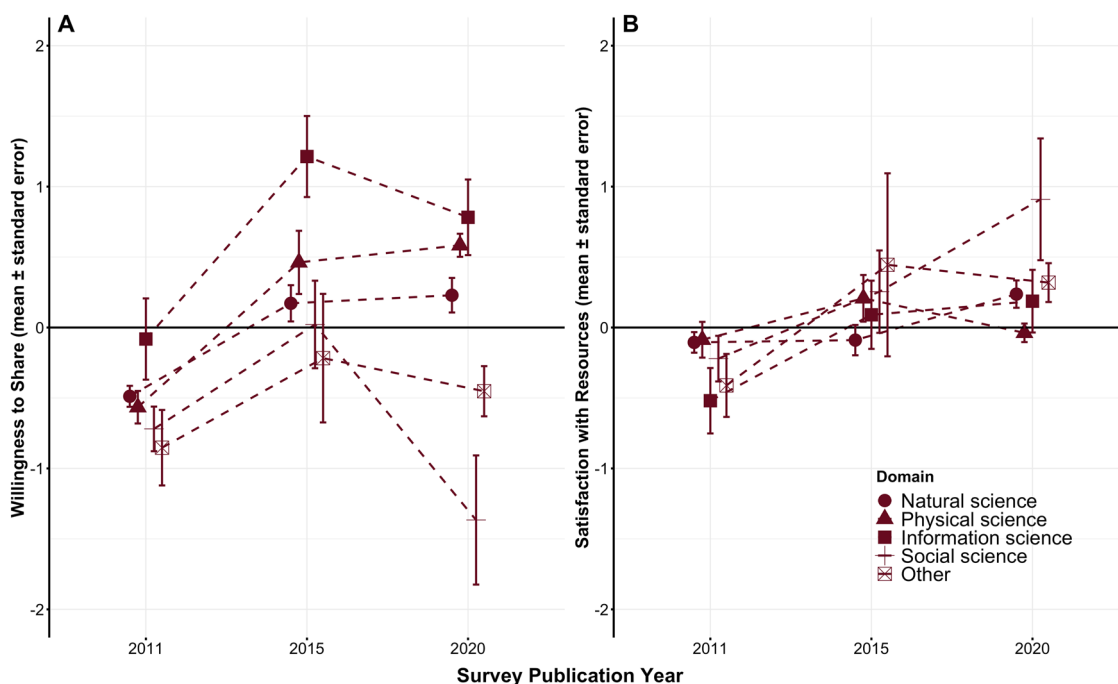


Fig. 3 Domain MFA. Willingness to share (A) and satisfaction with resources (B) by domain. The domains are divided into six categories (Natural Science, Physical Science, Information Science, Social Science, Other).

0.161, $p < 0.001$; 2011–2020: -0.604 to 0.155 , $p < 0.001$) (Fig. 4A and Table S12). In 2015, respondents from the government sector (1.020) became more willing to share data than those in academia (0.155, $p < 0.05$) and the difference between them was greater by 2020 (1.249, 0.155, $p < 0.001$) (Table S12). Satisfaction with resources was substantially larger in the commercial (0.645) sector than the government (-0.015) and academic (-0.046) sectors (Table 9). There was no trend observable over time (Fig. 4B). There was no correlation across work sectors between *willingness to share* and *satisfaction with resources*.

The effect of funding agency was the last variable analyzed (Table 10 and Fig. 5). Researchers with federal & national funding (0.257) were far more willing to share data than researchers with state, regional & local government (-0.676), corporate (-0.555), or private funding (-0.312). There was a substantial increase in *willingness to share* for researchers receiving funding from federal & national governments between 2011 and 2015 (-0.323 to 0.605 , $p < 0.001$) and 2011–2020 (-0.323 to 0.596 , $p < 0.001$) (Fig. 5A and Table S13). Researchers with funding from corporations, however, showed a marginal, non-significant increase in their *willingness to share* between 2011 and 2020 (-1.305 to 0.318 , $p = 0.055$). Satisfaction with resources showed no change overall or between years (Fig. 5B). There was no correlation across funding agencies between *willingness to share* and *satisfaction with resources*.

Table 8 Significant pairwise differences between willingness to share according to work sector computed with ANOVA and TukeyHSD.

Dimension 1 (Willingness to share)	Mean	Government	Non-profit
Mean	-	0.756	0.010
Academic	-0.046	0.904***	-
Commercial	-0.264	1.020***	-
Government	0.756	-	-0.746^*

* $p < 0.05$, *** $p < 0.001$.

Discussion

Our data indicate that researchers’ analysis of the costs, benefits, and effort associated with sharing and reusing data has the greatest impact on their *willingness to share* and *satisfaction with resources* (Table 4). The top-scoring responses associated with *willingness to share* relate to a general sense that sharing data is a good idea in general (Table 2). Organizational and social influences have a significant impact on *satisfaction with resources*, while individual elements relate to satisfaction with current practices or knowledge about the existence of data-sharing tools

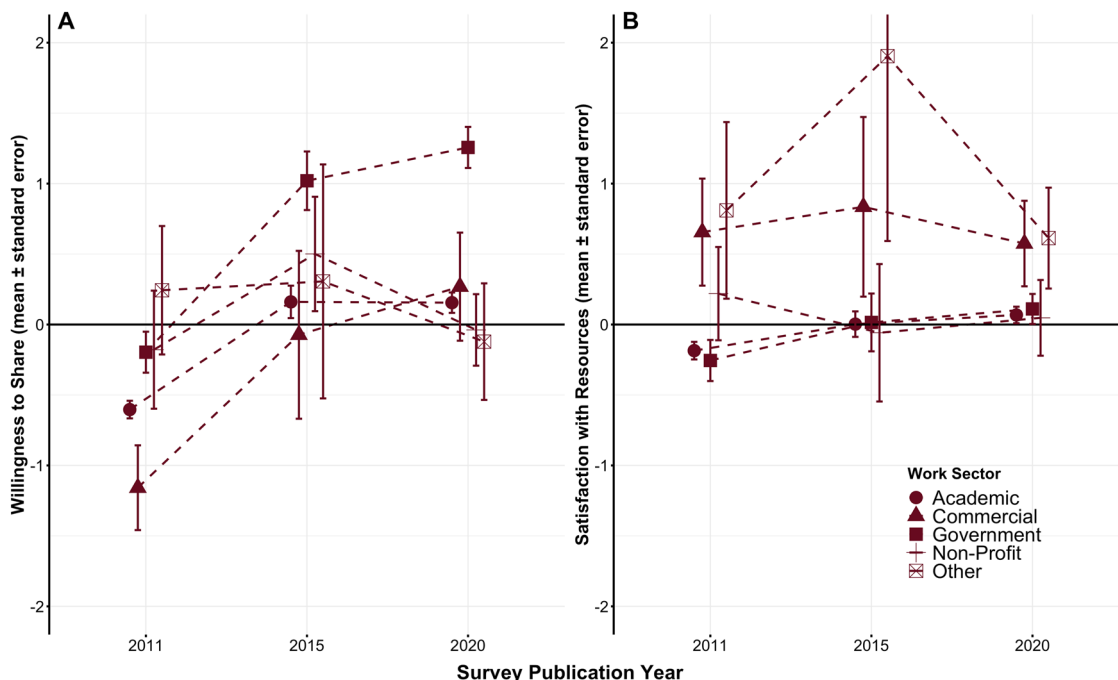


Fig. 4 Work sector MFA. Willingness to share (A) and satisfaction with resources (B) by work sector. The work sectors are divided into five categories (Academic, Commercial, Government, Non-Profit, Other).

Table 9 Significant pairwise differences between satisfaction with resources according to work sector computed with ANOVA and TukeyHSD.

Dimension 2 (Satisfaction with resources)	Mean	Commercial	Government	Other
Mean	-	0.645	-0.015	0.792
Academic	-0.046	0.691**	-	0.838**
Government	-0.015	0.661*	-	-
Other	0.792	-	-0.808*	-

*p < 0.05, **p < 0.01.

Table 10 Significant pairwise differences between willingness to share according to sources of funding computed with ANOVA and TukeyHSD.

Dimension 1 (Willingness to share)	Mean	Federal & national government
Mean	-	0.257
Corporation	-0.555	0.812**
Private foundation	-0.312	0.569*
State, regional & local government	-0.676	0.933***
Other	-0.590	0.847***

*p < 0.05, **p < 0.01, ***p < 0.001.

(Tables 2 and 3). This indicates that individual researchers are happy with their current data management practices, but tend not to be satisfied with the tools and data provided by other researchers (Table 3).

The hypotheses derived from our framework (Fig. 1) are supported by our analysis of dimensions 1 and 2. The correlations between the survey responses to individual questions and *willingness to share* (dimension 1) (Table 2) indicate that researchers who are less concerned with the risks of sharing and reusing data tend to see it as beneficial to their careers and the research community in general (H2). The strong impact of social influences on *willingness to share* (Tables 2 and 4) relates to the concept that a trustworthy and open research community is one that ensures credit is given to the researchers that deserve it (H1). The correlations between question responses in Table 3 and the *satisfaction with resources* dimension confirm hypothesis H3, that providing training and tools to researchers improves their data sharing and reuse attitudes. It is also clear that researchers who are satisfied with their data resources tend to be less concerned with the risks of sharing and reusing data based on the correlations in Table 3.

There was a substantial increase in researchers' *willingness to share* data in most parts of the world between 2011 and 2020

(Fig. 2). The USA & Canada, Australia & New Zealand, and Europe & Russia showed similar patterns, consistent with research on existing collaboration networks between member countries (Leydesdorff et al., 2013). In contrast, respondents from China and many Middle Eastern countries were not as willing to share their research data, which is possibly the result of a complex combination of historical, cultural, and governmental factors, although this situation is changing (Barrios et al., 2019; He, 2009) (Fig. 2 and Table S10).

The results by research domain (Fig. 3 and Table S11) are consistent with existing research on levels of collaboration across different fields (Barrios et al., 2019; Ding, 2011; Iglíč et al., 2017; Kyvik and Reymert, 2017), which showed that researchers in the social sciences are more likely to have single-authored papers and are less likely to share their data than are researchers in the physical and natural sciences (Iglíč et al., 2017; Kyvik and Reymert, 2017). This may result from the differences in data type and restrictions associated with human subjects data (Fecher et al., 2015), lower levels of interdisciplinary research in the humanities and social sciences (Bishop et al., 2014; Uddin et al., 2021), or relate to the more fundamental issue of funding. An

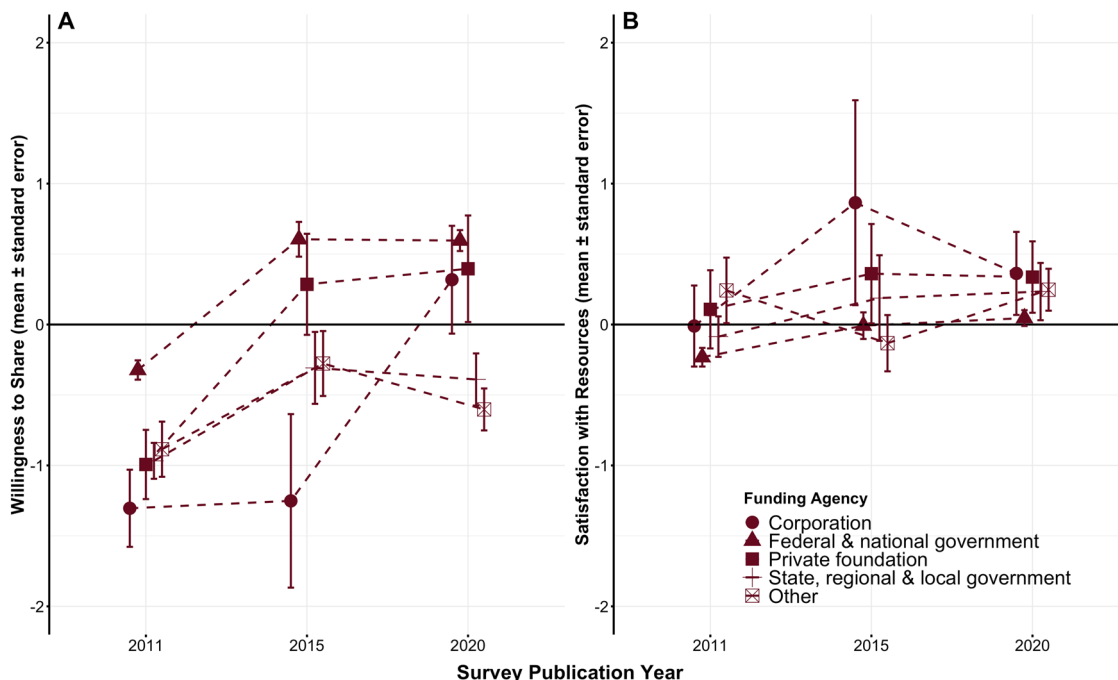


Fig. 5 Funding agency MFA. Willingness to share (A) and satisfaction with resources (B) by funding agency. The funding agencies are divided into five categories (Corporation, federal & national government, Private foundation, State, regional & local government, Other).

increasing overlap between commercial and federal funding is increasingly making research more profit rather than theory-focused. This trend favors science, technology, engineering, and mathematics research as these produce highly cited research and bring in a great deal of grant money (Münch, 2016). Social science was the only domain in our study that increased *satisfaction with resources* (2011–2020: -0.259 to 1.026 , $p < 0.05$). This change may have occurred due to the small number of respondents to the 2020 survey; however, it does align with other findings that researchers who are less likely to share their data, such as social scientists, feel satisfied with the resources they need for finding and sharing data (Tenopir et al., 2011, 2015, 2018, 2020).

Work sector has some notable impacts on *willingness to share*. Our observations of low *willingness to share* within the commercial sector in contrast to a high *willingness to share* in the government sector are consistent with results from the literature. Such results indicate that commercial involvement makes researchers more cautious about sharing data, while government involvement motivates open data practices through mandates and strong recommendations (Fig. 1C iii) (Evans, 2010b, 2010a). The sharp increase in the *willingness to share* data in the U.S. government sector (Fig. 4 and Table S12) may have been influenced by the Open Data Policy outlined by the U.S. Obama administration in 2013 (Holdren, 2013; Office of the Press Secretary, 2013). These data policies during the Obama administration resulted in a 400% increase in the number of datasets available on data.gov between 2012 and 2016 (Chief Information Officer’s Council, 2016). This is supported by the fact that the government sector of USA-Canada was the only sector to see a significant increase in *willingness to share* between 2011 and 2015 (-0.001 to 1.542 , $p < 0.05$). The European Union has also devoted a great deal of funding and effort to research data openness with the creation of the Public Sector Information (PSI) Group in 2002 (European Union, 2013). It is likely that the academic trends in *willingness to share* will follow those within the government sector because academic research depends heavily on funding from the federal government, and thus governmental policies will eventually apply.

Indeed, consistent with this idea, researchers receiving federal and national government funding (Fig. 5A and Table S13) were the most willing to share research data. Our observed increase in *willingness to share* research data involving commercial and federal funding may result from the large proportion of individuals in the commercial sector (31%) reported receiving the majority of their funding from a government source. These answers do not address the possibility of partial funding from the government, which means that this survey likely underestimates the proportion of those in the commercial sector that must abide by open data policies to receive government funding. It is difficult to interpret all the factors that drive data sharing within science, but generally speaking, the trends reported in this work indicate an increasing *willingness to share* research data without a substantial change in the level of *satisfaction with resources* for sharing research. This finding aligns with previously reported trends (Fane et al., 2019).

There were some substantial sampling differences between each of the three surveys, which may have affected the results (Table 1). The proportional response among groups for 2011 and 2015 was quite similar where the most substantial change in *willingness to share* and *satisfaction with resources* occurred. The 2020 survey had the most substantial differences in sampling proportions between the demographic categories. Despite this, *willingness to share* and *satisfaction with resources* did not change significantly from the results in 2015. This indicates that sampling differences are unlikely to be the cause of the observed differences between the 2011 and 2015 surveys.

The insights offered by looking at regional differences over time show that there is an overall negative relationship between *willingness to share* and *satisfaction with resources* depending on the region of the world (Fig. 6). Despite a high *willingness to share* data in Europe & Russia, USA & Canada, and Australia & New Zealand, well-documented (European Union, 2013; Fane et al., 2019; Holdren, 2013; Mason et al., 2020) advances in the establishment of mandates, training and the availability of resources were not reflected by a correspondingly high *satisfaction with resources*. This contrasts with other regions, where *willingness to*

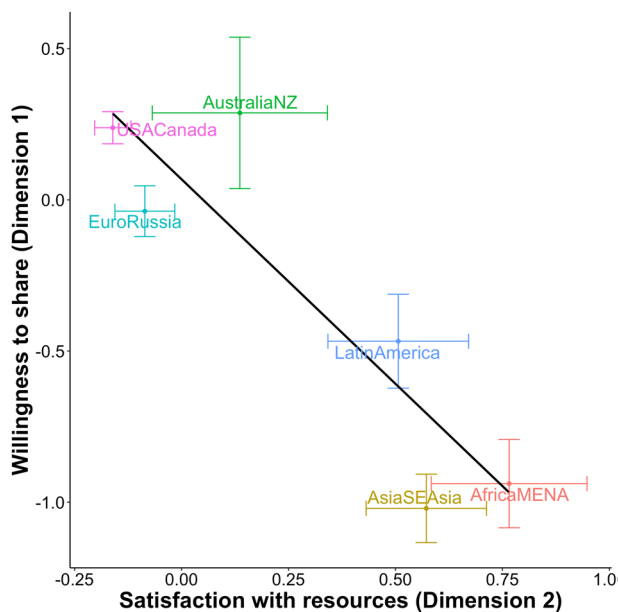


Fig. 6 Region MFA Correlation. Mean *willingness to share* (Dimension 1) versus mean *satisfaction with resources* (Dimension 2) by world region across all three surveys (All). The regression line is shown along with the standard errors of each point in each dimension.

share data may be low, but researchers are satisfied with the resources available. These results might appear to contradict H3, but really they indicate that data-sharing infrastructure, incentives, and social norms must be in place before data-sharing training and tools will have a meaningful impact. Mandates, tools, and training do improve data-sharing attitudes within organizations and regions (Fig. 2 and Table 3).

On the whole, scientists in the Western world (United States, Canada, Australia, New Zealand, Latin America, Europe) who work in academia, government, or corporations and who focus on the natural and physical sciences show an increased *willingness to share* research. There are many factors that could explain these trends. For example, there has been a substantial increase in international collaboration within Western countries over the past decade, but this scientific network has largely excluded African, Middle Eastern, and Asian countries (Leydesdorff et al. 2013). This outcome may be due to funding efforts that focus on collaboration across borders, for example, the work the Centers of Excellence for Europe (Bloch et al. 2016) and the International Panel on Climate Change. There have also been efforts to encourage collaboration and openness by federal agencies in the United States of America such as the Department of Energy with its Environmental Research Centers (Boardman and Ponomariov, 2011). Funding for collaborations between industry, government, and academia (Boardman and Gray, 2010; Gray et al., 2001; Lin and Bozeman, 2006) have also been prioritized with the growth of multidisciplinary, multipurpose university research centers. These collaborations have resulted in exponential growth in the funding and output of scientific research for the past decade (Clark and Llorens, 2012). This focus may partially explain why *satisfaction with resources* did not change for most of these categories throughout the three surveys. The increase in the need for collaboration and sharing has outpaced the investment in technologies and resources for sharing data (Bijsterbosch et al. 2016; Erway and Rinehart, 2016; NASEM, 2017). A survey of European research institutions showed that 56% devoted no funds to research data infrastructure and most respondents either did not know anything about said funding or reported it as insufficient

(Bijsterbosch et al. 2016). Data management is typically viewed as an indirect cost of research, rather than fundamental to the research process (Erway and Rinehart, 2016).

Finally, we want to mention that this work is a case study in data reuse, in particular its strengths and challenges. One of the primary objectives of open science is to ensure the reproducibility of results and hence reuse of data (Peng, 2011). Methodological consistency is a critical component of any type of data integration and sharing, and our experiences demonstrate this point. The results described from the combined data presented in this work demonstrate the importance and impact of good data stewardship. We, therefore, recommend that the final, combined data set we prepared for this work should be utilized for any future longitudinal research that involves these data. It is always challenging to balance a desire for long-term consistency and an ability to capture new trends. We believe that more robust findings could have been drawn between surveys had there been greater design consistency across surveys and their deployment. Despite these challenges, the trends we extracted from these data clearly indicate improvement in *willingness to share* data over time. We hope that future surveys on scientific attitudes toward data sharing and reuse will benefit from the lessons on the consistency that we have learned in this work.

Conclusions

We have shown that a trustworthy and open research climate and a perception of individual benefit increase data-sharing behaviors. The fact that *willingness to share* has increased since the 2010s in government and academic circles suggests that mandates and societal norms work. This is emphasized by the increase in *willingness to share* by researchers that receive federal or national funding. Such funding sources are dominated by research funding agencies such as the National Science Foundation and the Australian Research Council, all of whom have increasingly set an expectation of open data and open science for their grantees in alignment with recommendations from the International Science Council and the World Data System.

The difference in domain uptake of data-sharing attitudes suggests, as others have found (Specht and Crowston, 2022), that research that uses instruments or shared infrastructures such as astronomical telescopes, satellites or large marine vessels leads to an expectation and perhaps requirement that the research emanating from use of such facilities will be shared. Here the importance of social or community norms is paramount. For social scientists and other similar domain specialties work is more individual, resulting in them being late adopters of the open data concept.

We have confirmed that researchers in domains that have an existing high reliance on instruments or shared resources, researchers reliant on funding from national or federal sources, work sectors where internal practices are mandated, and regions where communication networks are highly effective have all improved in their *willingness to share* data. It is only when you move to action that you begin to test the tools and procedures necessary to engage in the practice. We encourage, based on this observation, increased collaboration with countries with fewer research resources to improve their access to data-sharing tools and expose them to open data norms that could foster a substantial increase in *willingness to share* data. Also, we encourage increased investment in data-sharing resources throughout the world as it will translate a *willingness to share* into actual sharing of data.

This work is limited by the fact that these data management surveys were not designed to be longitudinal or with our theoretical framework in mind. The questions were designed to

capture a wide array of behaviors and causes rather than identify specific ones. As a result, dimensions 1 and 2 did not contribute as strongly to the MFA as they would with a survey designed for this purpose. However, the results from this analysis are still significant and show interesting trends in the attitudes of researchers. In the future, we recommend such surveys should have a consistent theoretical framework to enable more precise diagnoses and the questions should be standardized over time to ensure the results are comparable.

Data availability

The datasets generated during and/or analyzed during the current study are available in the Zenodo repository, <https://zenodo.org/record/5932694>.

Received: 7 February 2022; Accepted: 7 June 2023;

Published online: 20 June 2023

References

- Ajzen I (1991) The theory of planned behavior. *Organ Behav Hum Decis Process* 50(2):179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Ajzen I, Fishbein M (1980) *Understanding attitudes and predicting social behavior*. Prentice-Hall
- Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature* 533(7604):452–454. <https://doi.org/10.1038/533452a>
- Barrios C, Flores E, Martínez MÁ, Ruiz-Martínez M (2019) Is there convergence in international research collaboration? An exploration at the country level in the basic and applied science fields. *Scientometrics* 120(2):631–659. <https://doi.org/10.1007/s11192-019-03133-9>
- Bijsterbosch M, Cordewener B, Duca D et al (2016) Funding research data management and related infrastructures. Zenodo. <https://doi.org/10.5281/ZENODO.5060104>
- Bishop PR, Huck SW, Ownley BH et al (2014) Impacts of an interdisciplinary research center on participant publication and collaboration patterns: a case study of the National Institute for Mathematical and Biological Synthesis. *Res Eval* 23(4):327–340. <https://doi.org/10.1093/reseval/rvu019>
- Bloch C, Schneider JW, Sinkjær T (2016) Size, accumulation and performance for research grants: examining the role of size for centres of excellence. *PLoS ONE* 11(2):e0147726. <https://doi.org/10.1371/journal.pone.0147726>
- Boardman C, Gray D (2010) The new science and engineering management: cooperative research centers as government policies, industry strategies, and organizations. *J Technol Transf* 35(5):445–459. <https://doi.org/10.1007/s10961-010-9162-y>
- Boardman C, Ponomariov B (2011) A preliminary assessment of the potential for “team science” in DOE Energy Innovation Hubs and Energy Frontier Research Centers. *Energy Policy* 39(6):3033–3035. <https://doi.org/10.1016/j.enpol.2011.03.066>
- Chief Information Officer’s Council (2016) Open data and open government. In: State of federal IT report. Chief Information Officer’s Council. <https://www.cio.gov/assets/resources/soft/02.03.soft.open.govt.open.data.pdf>. Accessed 6 Jun 2023
- Clark BY, Llorens JJ (2012) Investments in scientific research: examining the funding threshold effects on scientific collaboration and variation by academic discipline. *Policy Stud J* 40(4):698–729. <https://doi.org/10.1111/j.1541-0072.2012.00470.x>
- COPDESS (2018) Enabling FAIR data project. <https://copdess.org/enabling-fair-data-project>. Accessed 6 Jun 2023
- Curtly RG, Crowston K, Specht A et al (2017) Attitudes and norms affecting scientists’ data reuse. *PLoS ONE* 12(2):e0189288. <https://doi.org/10.1371/journal.pone.0189288>
- Data sharing and the future of science (2018) *Nat Commun* 9(1):2817. <https://doi.org/10.1038/s41467-018-05227-z>
- David R, Mabile L, Specht A et al (2020) FAIRness literacy: the Achilles’ heel of applying FAIR principles. *Data Sci*. <https://doi.org/10.5334/dsj-2020-032>
- Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MISQ* 13(3):319–340. <https://doi.org/10.2307/249008>
- Ding Y (2011) Scientific collaboration and endorsement: network analysis of coauthorship and citation networks. *J Informetr* 5(1):187–203. <https://doi.org/10.1016/j.joi.2010.10.008>
- Erway R, Rinehart A (2016) If you build it, will they fund? Making research data management sustainable. OCLC Research. <http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management-sustainable-2016.pdf>. Accessed 6 Jun 2023
- European Union (2013) European legislation on open data and the re-use of public sector information. In: *Shaping Europe’s digital future*. European Commission. <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>. Accessed 6 Jun 2023
- Evans JA (2010a) Industry collaboration, scientific sharing, and the dissemination of knowledge. *Soc Stud Sci* 40(5):757–791. <https://doi.org/10.1177/0306312710379931>
- Evans JA (2010b) Industry induces academic science to know less about more. *Am J Sociol* 116(2):389–452. <https://doi.org/10.1086/653834>
- Fane B, Ayris P, Hahnel M et al (2019) The state of open data report 2019. Digit Sci. <https://doi.org/10.6084/m9.figshare.9980783.v2>
- Fecher B, Friesike S, Hebing M (2015) What drives academic data sharing? *PLoS ONE* 10(2):e0118053. <https://doi.org/10.1371/journal.pone.0118053>
- Fidler F, Chee YE, Wintle BC et al (2017) Metaresearch for evaluating reproducibility in ecology and evolution. *BioScience* 67(3):282–289. <https://doi.org/10.1093/biosci/biw159>
- Gray DO, Lindblad M, Rudolph J (2001) Industry–university research centers: a multivariate analysis of member retention. *J Technol Transf* 26(3):247–254. <https://doi.org/10.1023/A:1011158123815>
- He T (2009) International scientific collaboration of China with the G7 countries. *Scientometrics* 80(3):571–582. <https://doi.org/10.1007/s11192-007-2043-y>
- Holdren JP (2013) Increasing access to the results of federally funded scientific research. Memorandum for the heads of executive departments and agencies. Office of Science and Technology Policy. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf. Accessed 6 Jun 2023
- Husson F, Josse J, Le S et al (2020) FactoMineR: Multivariate exploratory data analysis and data mining (2.3). <https://CRAN.R-project.org/package=FactoMineR>. Accessed 6 Jun 2023
- Iglić H, Doreian P, Kronegger L et al (2017) With whom do researchers collaborate and why? *Scientometrics* 112(1):153–174. <https://doi.org/10.1007/s11192-017-2386-y>
- IUCN (2021) IUCN regions. <https://www.iucn.org/regions>. Accessed 6 Jun 2023
- Kassambara A, Mundt F (2020) Factoextra: extract and visualize the results of multivariate data analyses (1.0.7). <https://CRAN.R-project.org/package=factoextra>. Accessed 6 Jun 2023
- Kim Y, Stanton JM (2016) Institutional and individual factors affecting scientists’ data-sharing behaviors: a multilevel analysis. *J Assoc Inform Sci Technol* 67(4):776–799. <https://doi.org/10.1002/asi.23424>
- Kim Y, Yoon A (2017) Scientists’ data reuse behaviors: a multilevel analysis. *J Assoc Inform Sci Technol* 68(12):2709–2719. <https://doi.org/10.1002/asi.23892>
- Kim Y, Zhang P (2015) Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Lib Inf Sci Res* 37(3):189–200. <https://doi.org/10.1016/j.lisr.2015.04.006>
- Kyvik S, Reymert I (2017) Research collaboration in groups and networks: differences across academic fields. *Scientometrics* 113(2):951–967. <https://doi.org/10.1007/s11192-017-2497-5>
- Leydesdorff L, Wagner C, Park HW et al (2013) International collaboration in science: the global map and the network. *ArXiv* 1301.0801. <https://doi.org/10.48550/arXiv.1301.0801>
- Lin MW, Bozeman B (2006) Researchers’ industry experience and productivity in university–industry research centers: a “scientific and technical human capital” explanation. *J Technol Transf* 31(2):269–290. <https://doi.org/10.1007/s10961-005-6111-2>
- Ludäscher B (2016) A brief tour through provenance in scientific workflows and databases. University of Illinois Research and Tech Reports—Computer Science. University of Illinois. <http://hdl.handle.net/2142/89717>. Accessed 6 Jun 2023
- Mason CM, Box PJ, Burns SM (2020) Research data sharing in the Australian national science agency: understanding the relative importance of organisational, disciplinary and domain-specific influences. *PLoS ONE* 15(8):e0238071. <https://doi.org/10.1371/journal.pone.0238071>
- Michener WK, Allard S, Budden A et al (2012) Participatory design of DataONE—enabling cyberinfrastructure for the biological and environmental sciences. *Ecol Inform* 11:5–15. <https://doi.org/10.1016/j.ecoinf.2011.08.007>
- Milham MP, Craddock RC, Son JJ et al (2018) Assessment of the impact of shared brain imaging data on the scientific literature. *Nat Commun* 9(1):2818. <https://doi.org/10.1038/s41467-018-04976-1>
- Münch R (2016) Academic capitalism. In: *Oxford research encyclopedia of politics*. <https://doi.org/10.1093/acrefore/9780190228637.013.15>. Accessed 6 Jun 2023
- NASEM (2017) Fostering integrity in research. <https://doi.org/10.17226/21896>. Accessed 6 Jun 2023
- NIH (2023). NIH policy for data management and sharing. NIH grants and funding. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>. Accessed 6 Jun 2023
- Office of the Press Secretary (2013) Executive order: making open and machine readable the new default for government information. Executive orders. <https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive>

- order-making-open-and-machine-readable-new-default-government-. Accessed 6 Jun 2023
- Olendorf R, Borycz J, Grant B (2022) Scientists attitudes toward sharing data over time. Zenodo. <https://doi.org/10.5281/zenodo.5932694>
- Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science* 349(6251). <https://doi.org/10.1126/science.aac4716>
- Peng RD (2011) Reproducible research in computational science. *Science* 334(6060):1226–1227. <https://doi.org/10.1126/science.1213847>
- Perez-Riverol Y, Zorin A, Dass G et al. (2019) Quantifying the impact of public omics data. *Nat Commun* 10(1):3512. <https://doi.org/10.1038/s41467-019-11461-w>
- Specht A, Crowston K (2022) Interdisciplinary collaboration from diverse science teams can produce significant outcomes. *PLoS ONE* 17(11):e0278043. <https://doi.org/10.1371/journal.pone.0278043>
- Talukder M (2012) Factors affecting the adoption of technological innovation by individual employees: an Australian study. *Procedia Soc Behav Sci* 40:52–57. <https://doi.org/10.1016/j.sbspro.2012.03.160>
- Tenopir C, Allard S, Douglass K et al. (2011) Data sharing by scientists: practices and perceptions. *PLoS ONE* 6(6):e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir C, Christian L, Allard S et al. (2018) Research data sharing: practices and attitudes of geophysicists. *Earth Space Sci* 5(12):891–902. <https://doi.org/10.1029/2018EA000461>
- Tenopir C, Dalton ED, Allard S (2015) Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE* 10(8):e0134826. <https://doi.org/10.1371/journal.pone.0134826>
- Tenopir C, Rice NM, Allard S (2020) Data sharing, management, use, and reuse: practices and perceptions of scientists worldwide. *PLoS ONE* 15(3):e0229003. <https://doi.org/10.1371/journal.pone.0229003>
- Uddin S, Imam T, Mozumdar M (2021) Research interdisciplinarity: STEM versus non-STEM. *Scientometrics* 126(1):603–618. <https://doi.org/10.1007/s11192-020-03750-9>
- Venkatesh V, Morris MG, Davis GB et al. (2003) User acceptance of information technology: toward a unified view. *MISQ* 27(3):425–478. <https://doi.org/10.2307/30036540>
- Wilkinson MD, Dumontie M, Aalbersberg IJ et al (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. <https://doi.org/10.1038/sdata.2016.18>
- Yoon A, Kim Y (2017) Social scientists' data reuse behaviors: exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. *Lib Inf Sci Res* 39(3):224–233. <https://doi.org/10.1016/j.lisr.2017.07.008>
- Zhang Z, Hernandez K, Savage J et al. (2021) Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat Commun* 12(1):1226. <https://doi.org/10.1038/s41467-021-21254-9>

Acknowledgements

This project did not receive funding, but the authors would like to acknowledge DataONE for facilitating our participation in the Usability and Assessment Working

Group and for their consistent support for this project. We also would like to thank the several reviewers who provided invaluable advice on drafts of the article.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors. The data used in this work was extracted from four published research papers that each underwent the proper Institutional Review Board (IRB) procedures for human subjects' data (see below). Informed consent was given by all survey participants in accordance with the requirements of the University of Tennessee IRB: Tenopir (2011), Tenopir (2015), Tenopir (2018) and Tenopir (2020).

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-023-01831-7>.

Correspondence and requests for materials should be addressed to Joshua Borycz.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023