



Open data in the life sciences: the 'Selfish Scientist Paradox'

D. Damalas^{1,*}, G. Kalyvioti¹, E. C. Sabatella², K. I. Stergiou^{1,3}

¹Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research, Aghios Kosmas, 16777 Athens, Greece

²NISEA Fisheries and Aquaculture Economic Research, Via Irno 11, 84135 Salerno, Italy

³Laboratory of Ichthyology, Department of Zoology, School of Biology, Aristotle University of Thessaloniki, UP Box 134, 541 24 Thessaloniki, Greece

ABSTRACT: Full and open access is promoted as the international norm for the exchange of scientific data by numerous scientific and political bodies. In the contemporary digital era, since scientists are both consumers and producers of data, they inevitably play a crucial role in defining the level of data accessibility. Yet, it is individual researchers usually who resist the release of their data. Through a global online questionnaire survey, the perception of 858 life scientists with respect to open data was investigated. Differences in scientists' perceptions were tested per major country, rank position and academic performance in order to identify partial and global preferences. The 'Selfish Scientist Paradox' was identified: although the majority of respondents were in favour of open access to life sciences data, and most acknowledged that data gathered by others is vital to their work, the same group of people were quite reluctant to share their own data; only a third of them were willing to make their data available unconditionally. Scientists with >10 yr professional experience were twice as likely to oppose open access, while almost half of junior researchers would rather not share their data prior to publishing. Senior scientists argued that although project funding in general was a significant incentive towards making their data available, at the same time certain confidentiality agreements in some projects become a main barrier to data sharing. Country of professional location largely affected most responses, revealing that southern Europeans had a 'conservative' attitude towards open access, being more unwilling to share their data. Analyses based on academic performance (publications and citations) indicated that established individuals were more dependent on data collected by others and more opposed to open access.

KEY WORDS: Open data · Life sciences · Questionnaire

INTRODUCTION

Scientists are driven by curiosity, constantly asking new questions and testing hypotheses. In doing so, they either have to produce their own data from carefully planned laboratory and/or field experiments or to extract data from 'libraries'. Their curiosity is generally supported by funding; e.g. US\$1.14 trillion was spent on research and development (R&D) in the countries of the Organisation for Economic Co-oper-

ation and Development (OECD) in 2015 (OECD 2017).

Conventionally, knowledge exchange in science is conducted through journal articles and conference proceedings. However, in the contemporary digital world, access to original data has become essential and is equally important to publishing for both individual scientists and their hosting institutes (Borgman 2007).

The level of the current global data supply is huge, reaching 2.8 trillion GB in 2012 (projected to reach

*Corresponding author: shark@hcmr.gr

40 trillion GB by 2020), of which only 0.5 % is used for analysis (Gantz & Reinsel 2012). There is a strong movement by the stakeholders related to scientific data (i.e. governments, research-funding agencies, universities and non-profit research institutes, professional societies, international scientific organizations, industry research institutions, and the general public; Uhler & Schroder 2007) promoting open-data policy (e.g. Editorial 2007a,b, Piwowar et al. 2008, Vision 2010, Molloy 2011, Warren 2016; <http://blogs.nature.com/scientificdata/>). For instance, after reviewing its existing policy on sharing research data, the US National Science Foundation (NSF) now requires that grant awardees make primary data available to others (NSF 2010). Following the same path, one of the pillars of research funding in Europe, the Horizon 2020 (H2020) financial instrument, has pledged to improve and maximize access to and reuse of research data generated by projects. To this end, all H2020-funded projects are currently required to deposit all data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible. Moreover, project partners will have to take measures so that third parties are able to access, mine, exploit, reproduce and disseminate project data—free of charge for any user (<https://www.openaire.eu/h2020-oa-data-pilot>). Also, 2900 journals from >500 publishers have signed up to the Transparency and Openness Promotion (TOP) guidelines for journals, along with 57 organizations (Gewin 2016). TOP advocates that transparency, open sharing, and reproducibility are core values of science, although not always part of daily practice, and has introduced a set of standards to encourage disclosure of data and increase the credibility of inferential research (Nosek et al. 2015).

There is a plethora of benefits arising from the openness of scientific data collected through research-funding instruments: e.g. the ability to address new questions and test new hypotheses, provide multiple perspectives, conduct large spatio-temporal scale analyses, develop new study and software, maximize data utility and avoid duplication of effort, increase visibility/transparency of scientific output, identify error, discourage scientific fraud, and increase opportunities for education and training (e.g. Piwowar et al. 2007, 2008, Boulton et al. 2011, Molloy 2011, Van den Eynden et al. 2011, Piwowar & Vision 2013, Warren 2016). Unquestionably, all these benefits accelerate scientific progress and eventually lead to better science for the common good.

Although scientists are considered the weak link in the open-data debate (Gewin 2016), the available

studies (Table 1) that have surveyed the opinions of scientists on open data suggest that in principle (as opposed to in practice), they are in favour. Herein, we present the results of a global online questionnaire survey, which took place between January and August 2017, concerning the perception of life scientists with respect to open data. We also tested for differences in scientists' perceptions per major country, rank position and citation index.

MATERIALS AND METHODS

Participants

This study involved an online questionnaire survey (<http://artemis2.ath.hcmr.gr/HcmrPolls/polls/questions/>), with the anonymity of respondents protected; no identifying questions were asked, and findings herein are reported in aggregation. The survey was open for responses from January 15 to August 21, 2017.

At the first stage, researchers were approached through a 'snowball' sampling method (Goodman 1961). An email cover letter was sent to a small group of selected members of scientific society, notably colleagues whom we have collaborated with. This group was requested to circulate the link to their colleagues in their department (faculty, lecturers, postdoctoral associates). At the second stage, the survey encompassed a wider spectrum of researchers, and the electronic survey targeted the mailing lists of faculty members of specific departments (biogeography, biophysics, botany, conservation biology, ecology, evolutionary biology, fisheries, physiology, systematics, zoology) from universities, research centres and institutions.

Research implementation

The survey implementation proceeded by partitioning the questionnaire into 3 sections: demographics, academic performance and relationship with data. The full version of the questionnaire is provided in the online Supplement at www.int-res.com/articles/suppl/e018p027_supp.pdf.

Scientists were asked to provide their country of employment, level of professional experience and rank in their organization. In order to evaluate differences in perception among established and emerging researchers, respondents were asked to provide a measure of academic performance, notably numbers of published papers and citations.

Table 1. Surveys investigating the opinion of scientists on open data in various disciplines

Source (year)	Discipline	Sample size	Perceptions of open access to data
Ceci (1988)	Academicians from 2 public state universities and 1 private	790	Although scientists endorsed the open-data norm enthusiastically with no disciplinary departures, they also reported cases of colleagues blatantly refusing to share data
Campbell et al. (2002)	Geneticists and other life scientists	1849	35% of geneticists said that sharing had decreased during the last decade, and 14 % that sharing had increased. Geneticists were as likely as other life scientists to deny others' requests and to have their own requests denied
Swan & Brown (2008)	Astronomy, chemical crystallography, classics, climate science, genomics, social and public health sciences, systems biology, rural economy	100	Many researchers wish to retain exclusive use of the data they have created until they have extracted all the publication value they can. When combined with the perceived lack of career rewards for data creation and sharing, this constitutes a major constraint on the publishing of data
Tenopir et al. (2011)	Environmental sciences, ecology, social sciences, biology, physical sciences, computer science, engineering, atmospheric science, medicine	1329	A majority of respondents to this international survey of data practices are willing to share at least some of their data and re-use others' data, pending certain conditions or restrictions on use
Enke et al. (2012)	Biodiversity science	799	Clear commitment to share biodiversity data, but also a reluctance to actually do so, due to a mixture of social and technical impediments, such as loss of control over data and lack of professional reward for sharing
European Commission (2012)	National governments, regional and local governments, research funding organizations, university/research institutes, libraries, publishers, international organizations, individual researchers, citizens, nongovernmental organizations (NGOs), industries, charities, learned societies, scientific and professional associations	1140	Strong support (90 % of responses) for research data that is publicly available and results from public funding to be, as a matter of principle, available for re-use and free of charge on the internet
Kim (2013)	Engineering, physical sciences, earth, atmospheric, and ocean sciences, mathematical sciences, computer science, agricultural sciences, biological sciences, psychology, social sciences, health fields	1317	Significant between-discipline variances as well as within-discipline variances in scientists' data-sharing behaviours, influenced by both institutional factors (i.e. regulative pressure by journals and normative pressure) and individual factors (i.e. perceived career benefit, perceived effort, and scholarly altruism)
Akers & Doty (2013)	Arts and humanities, social sciences, medical sciences, and basic sciences	330	Statistically significant differences among research domains, revealing important disciplinary distinctions in data-management actions and attitudes
DAMVAD (2014)	Humanities, agriculture and fishery, mathematics and natural science, medical science, social science, technology	1474	Around 80% of respondent researchers agreed that open access to research data enhances research, and that it is an ethical obligation of research to make research data available for validation
Tenopir et al. (2015)	Environmental sciences, ecology, social sciences, biology, physical sciences, computer science, engineering, atmospheric science, medicine	1015	Results point to increased acceptance of and willingness to engage in data sharing, as well as an increase in actual data-sharing behaviours. However, there is also increased perceived risk associated with data sharing, and specific barriers to data sharing persist
Kratz & Strasser (2015)	Biology, archaeology, social science, environmental science, physical sciences, earth science, computer science, mathematics	250	Researcher expectations of data publication centre on availability, generally through an open data-base or repository
Schmidt et al. (2016)	Environmental sciences, earth sciences, marine and polar sciences, biodiversity, social and economic sciences	1330	A wide range of good practice examples was pointed out by respondents, which demonstrates a substantial uptake of data sharing through e-infrastructures and a further need for enhancement and consolidation. Among all policy responses, funder policies seem to be the most important motivator. This supports the conclusion that stronger mandates will strengthen the case for data sharing
The present study (2017)	Life sciences	858	Although the majority were in favour of open access to life sciences data and most acknowledged that data gathered by others is vital to their work, the same group of people were quite reluctant to share their own data

To evaluate respondents' attitude to data openness, they were asked to what degree they agree or disagree with a series of statements. Each question or statement explored a different aspect: collection and use of research data; storage and maintenance of research data; views on data sharing and fair exchange of data and responsibility for their data.

Questions included both Likert scales (Likert 1932), based on a point system to assign levels of agreement or disagreement to a respondent's answer, as well as closed-ended questions expecting a 'yes' or 'no'. Likert-scaled questions provide quantitative information, allowing for data to be analysed with relative ease (see 'Likert scale' in the Supplement). In some cases, whenever the choices provided did not accurately represent the respondent's condition, the respondent was asked to describe the condition in a separate category ('other').

The present study focused on a subset of the questionnaire, namely questions Q1 to Q6 and Q11 to Q15.

RESULTS

Demographics of respondents

Between January 15 and August 21, 2017, we sent about 7500 emails, out of which a total of 858 researchers responded to the questionnaire (11%). Responses were almost equally distributed among academic ranking (Q1; Fig. 1a), with a slight preponderance of senior scientists (39%). Almost two-thirds of the replies came from scientists with >15 yr of professional experience (Q2; Fig. 1b).

Europe (63%) and North America (19%) dominated the responses, followed by Australia (8%). Overall, 15 out of 40 countries surveyed represented 85% of the responses (Q3; Fig. 2). The only 3 countries with no responses received were China, Tai-

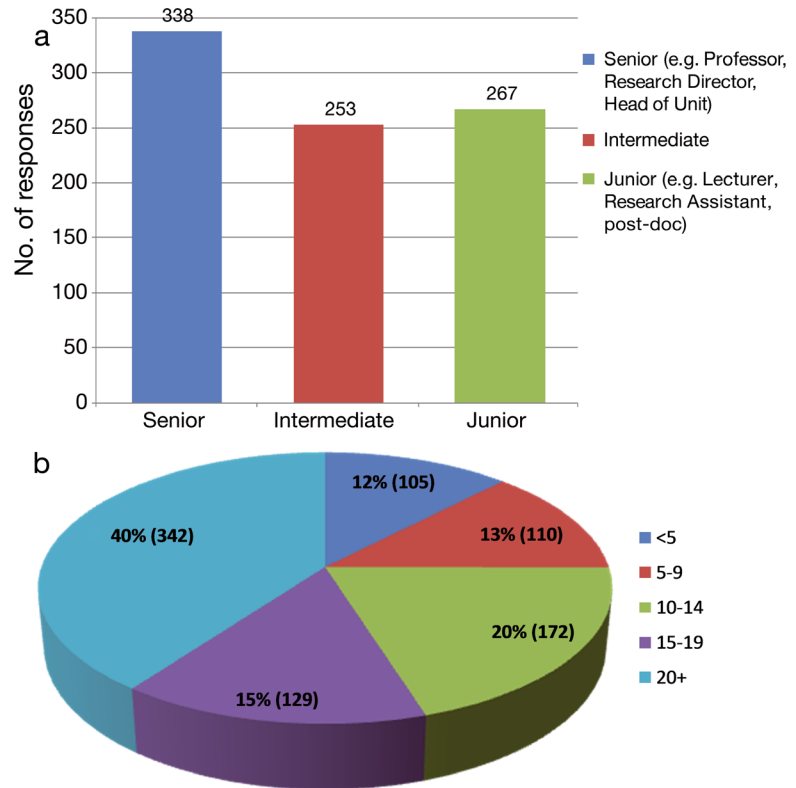


Fig. 1. Respondents by (a) academic ranking in their organization and (b) years of professional experience

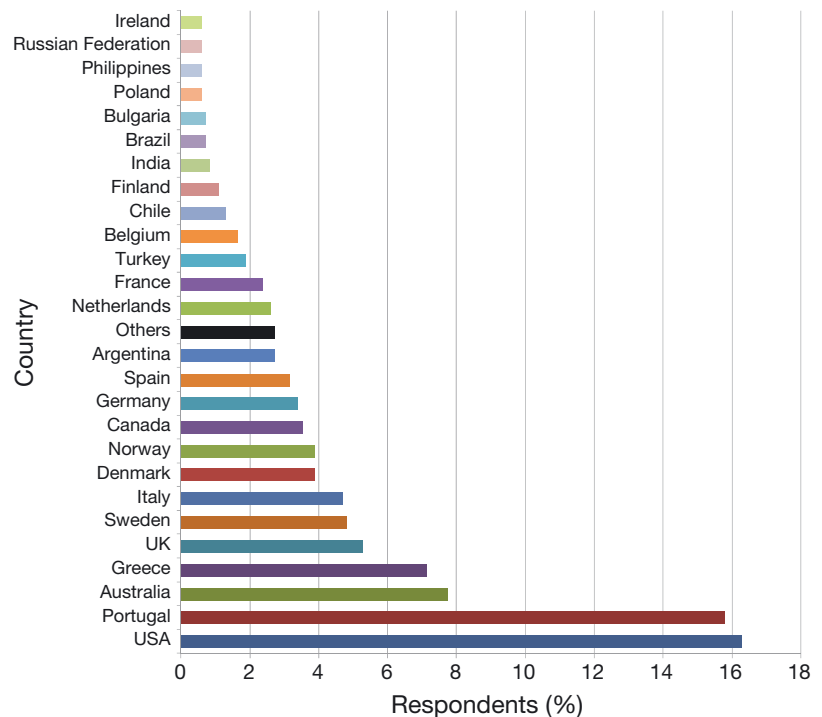


Fig. 2. Respondents by country of professional location (note: 1% represents approximately 9 respondents)

wan and Japan (although >300 requests were sent out). Almost two-thirds of the respondents had <50 peer-reviewed publications (median value: 40 publications; Q4; Fig. 3a), and 61% had <1000 citations (Fig. 3b), with those having >3000 citations making up 17%.

Global perceptions on data openness

The majority of the respondents (~80%) were in favour of open access to life sciences data (Q6;

Fig. 4a). Most researchers (64%) also acknowledged that data gathered by others is vital to their work (Q12; Fig. 4b), with >80% of them (Q13; Fig. 4c) asserting that during the past 5 yr, they have utilized data collected by other researchers (Q13). The source of these external data varied from local to international institutional databases (Q14; Fig. 5a). For those declaring other sources (35 responses), the majority (91%) identified publicly accessible repositories as their source of data (online databases: 37%, published manuscripts: 34%, official statistics: 20%).

Although the desire for accessing other scientists' data was evident, the same group of people were quite reluctant to share their own data (Q15). Less than one-third of respondents were willing to make their data available unconditionally, and the remaining two-thirds were willing to provide their data upon certain conditions (Q15; Fig. 5b). A subset of respondents (n = 106) provided comments on those certain conditions: e.g. availability only after publishing, gaining collaboration, co-authorship in a high-ranking journal, mutual exchange of data, and expertise swapping (Table 2).

The vast majority of respondents considered the rebuilding of data in case of catastrophic loss extremely difficult, signifying the issue of data storage, handling and access (Q11; Fig. 6).

Specific perceptions on data openness

The level of academic ranking (Q1) did not bias perceptions on data openness, working with data of others or reproducibility of research data (Fig. S1 in the Supplement). However, scientists with professional experience of >10 yr (Q2) were twice as likely to oppose open access, compared to their younger counterparts (with <10 yr of experience) (Fig. S2). Perceptions of one's own data sharing (Q15) were more or less comparable among all aca-

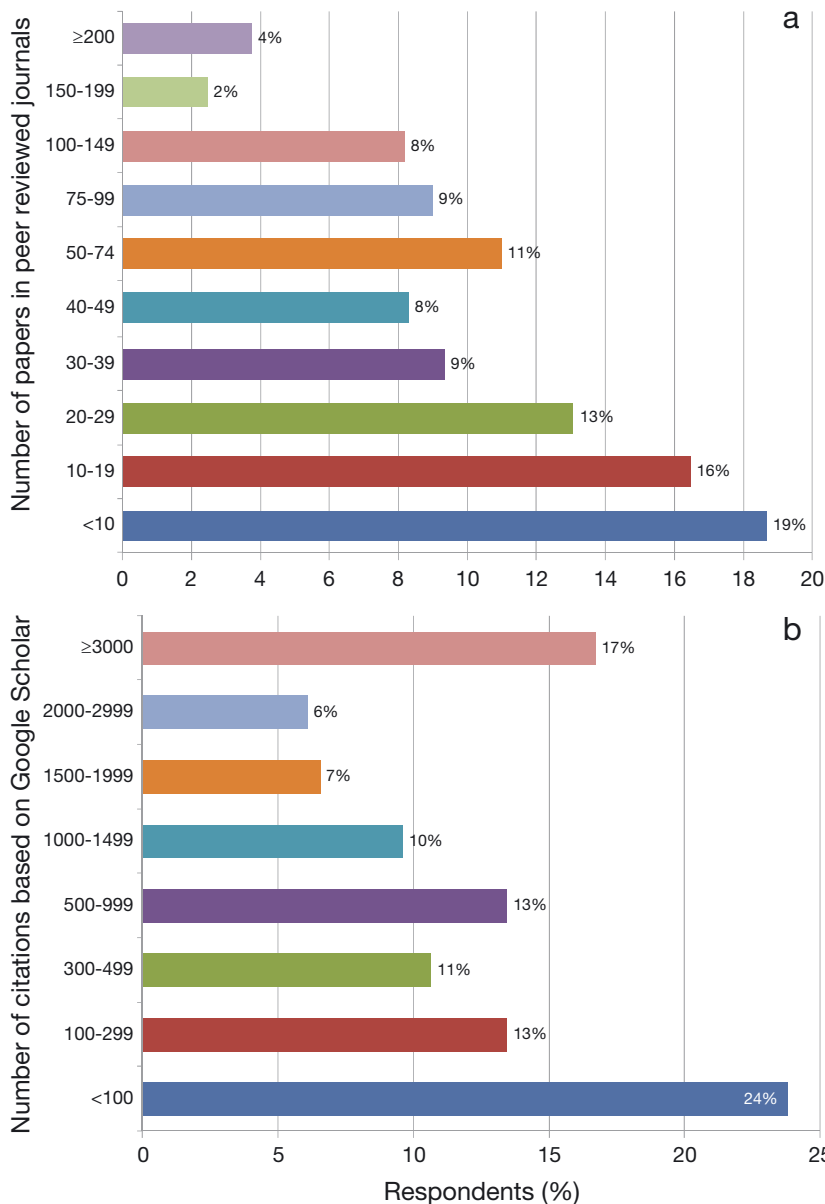


Fig. 3. Respondents by (a) number of papers published in peer-reviewed journals and (b) citations based on Google Scholar

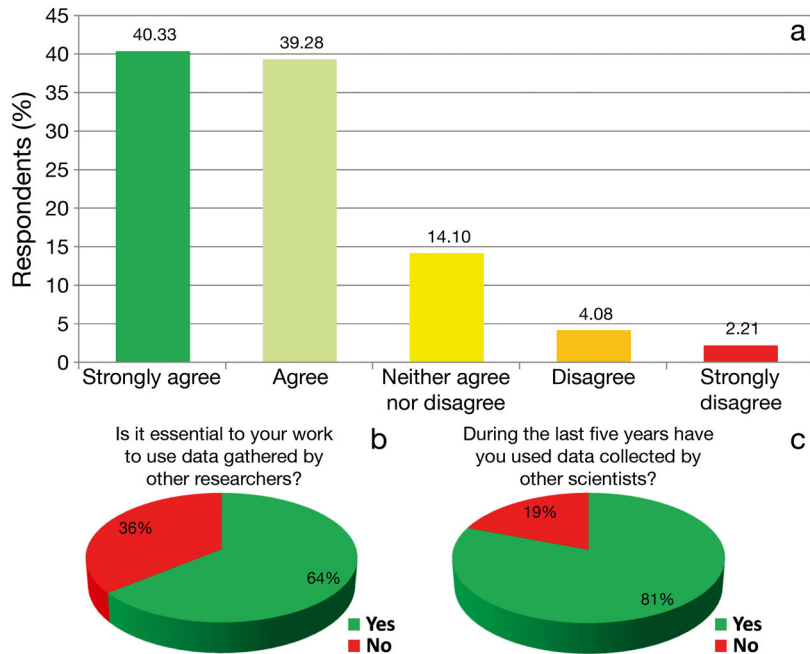
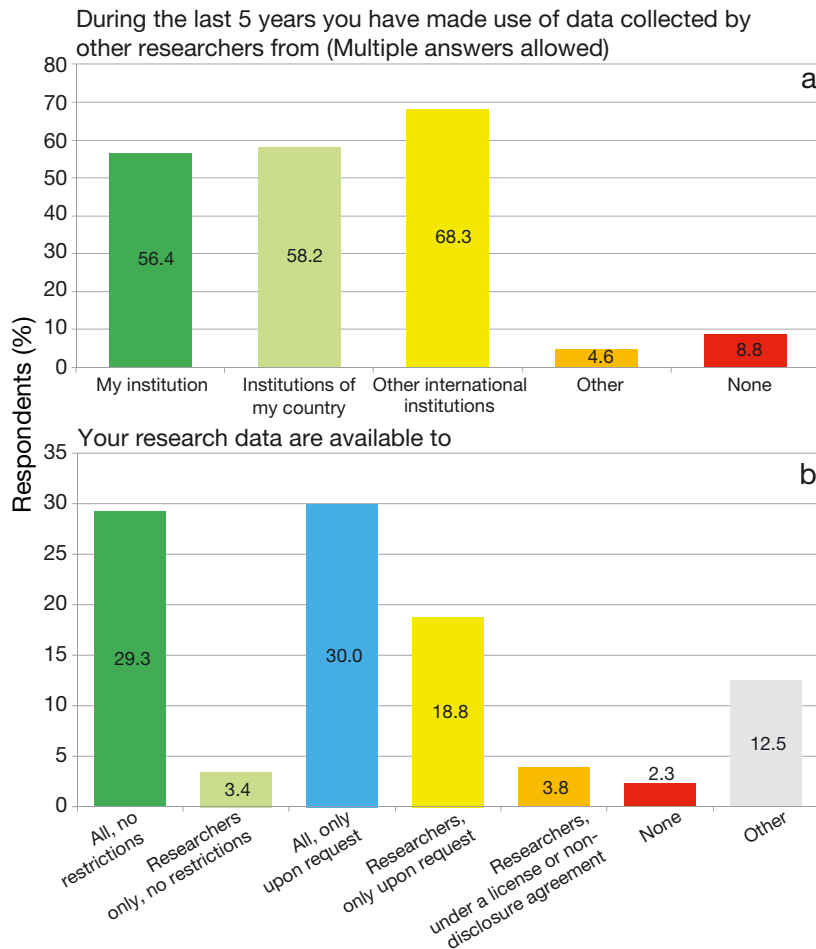


Fig. 4. Respondents' perception of (a) open access to life sciences data, (b) importance of data collected by others to their work and (c) use of others' data during the past 5 yr



demographic rankings (Table S1). Almost half of the junior researchers (46%) would rather not share their data prior to exploiting them to the advantage of their academic career (e.g. a publication or a conference presentation). In case they do so, they would prefer to share them with prominent researchers or to publish in high-ranking journals. Senior scientists argued that gaining project funding is also a significant incentive towards making their data available.

Country of professional location (Q3) largely affected most responses. Spanish, Greek, Portuguese and Australian scientists had the most negative opinion on open access (>10% against it), while scientists located in Asia (India, Philippines), Germany, 'other Americas' (Brazil, Argentina, Chile), France, Sweden, Denmark and Canada were at the other extreme (>80% in favour; Fig. S3, Table S2 in the Supplement). African scientists (Cameroon and South Africa) were predominantly (75%) of no opinion (Fig. S3). Scientists from the UK, Australia, the Netherlands and the USA considered it less essential to their work to use external data, whereas German, Danish, Norwegian and Canadian scientists were more likely to use other researchers' data (Fig. S3). During the past 5 yr (Q14–15), scientists from northern Europe (Denmark, Norway, Sweden) and Canada were more dependent on data collected by others, with Italians, Greeks and Germans being less prone to use external data sources (Fig. S3). Canadians, Italians and the British were the respondents most concerned about their data reproducibility, whereas scientists from the Netherlands, Sweden, Norway, Germany and the USA were the least concerned (Q11; Fig. S3).

Fig. 5. Researchers' response on (a) using data of others and (b) making their data available to others. (Due to multiple answers allowed, the top graph does not depict cumulative percentages, but % of total)

Table 2. Respondents' willingness to provide their research data upon certain conditions. Actual numbers are given in parentheses

Condition	Percentage of respondents (n)
After publishing	39% (42)
Only with other experts in my field	6% (7)
In case of offered collaboration	5% (5)
Only with other researchers	5% (5)
Only if to publish in a high-ranking journal	3% (3)
With engaged stakeholders	2% (2)
If proper credits are given	2% (2)
In case of mutual exchange of data	2% (2)
If co-authorship is offered	2% (2)
A mix of all the above	36% (38)

The perception towards data openness was inversely related to R&D expenditure (Fig. 7).

Analyses based on academic performance (publications and citations, Q4 and Q5) indicated that established individuals have made more frequent use of data collected by others (although this difference is rather marginal); this group of researchers was also more opposed to open access (Fig. S4 in the Supplement). Once again, junior–intermediate scientists, with a publishing record of <40 papers, were not willing to share their data before making the most out of them (publications), while researchers with 50+ published manuscripts were of the opinion that project funding and collaboration under confidentiality agreements are desirable conditions prior to sharing their data (Table S3).

In case of a catastrophic loss of your research data, how easy can you rebuild them?

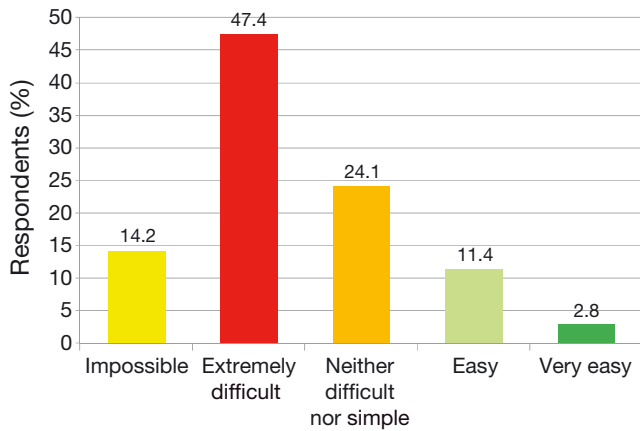


Fig. 6. Researchers' response on rebuilding research data

DISCUSSION

The debate about open access to research data is by no means new; it is an argument that is being increasingly raised. Open access is promoted as the international norm for the exchange of scientific data by numerous scientific and political bodies: the OECD (OECD 2007), European Commission (EC 2011), United Nations Educational, Scientific and Cultural Organization (UNESCO 2012), G8 science ministers (G8 Science Ministers 2013), US federal government (OMB 2013), and the International Council for Science (ICSU-WDS 2015). It has intensified in recent years due to the ever-increasing amount of data and the emergent possibilities offered

Investigating the country effect further, we identified an obvious association between country-specific percentage of gross domestic product (GDP) spending on research (OECD 2017) and data openness. For Greece, Italy and Spain, R&D expenditure fluctuated around 1% of GDP, while this figure was >3% in Finland, Sweden and Denmark.

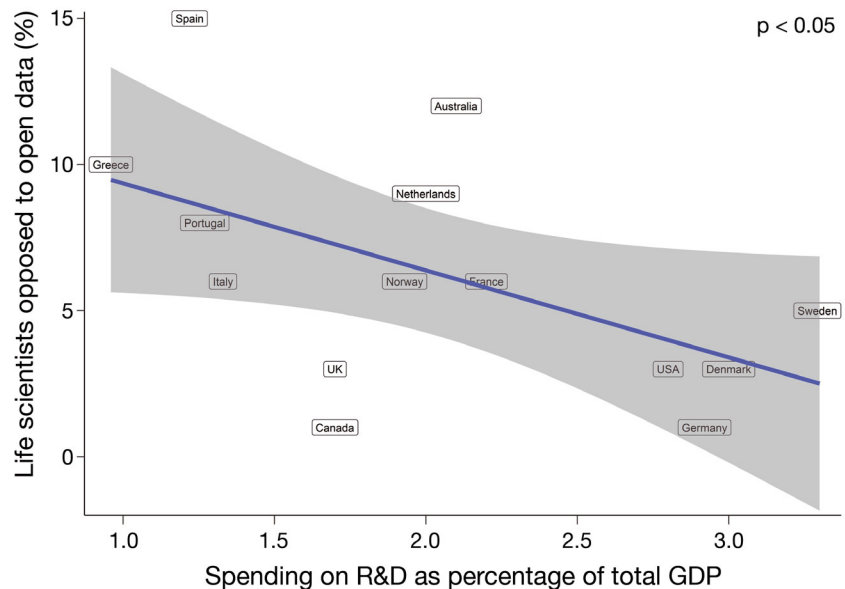


Fig. 7. Spending on research and development (R&D) as a percentage of total gross domestic product (GDP), against the percentage of life scientists opposed to open access. Linear trend is statistically significant ($p < 0.05$). Shaded area indicates 2 standard errors around the estimates

by information technology, along with growing recognition of the value of data.

In the contemporary scientific arena, data sharing through the emergence of e-science (Hey & Hey 2006) has revolutionized the way science is conducted. Technological developments, data repositories and collaborative efforts have brought to fruition a long-standing aspiration of researchers: access to data. However, everything comes at a cost, and in the academic realm, data has become an 'information currency' (Davis & Vickery 2007), shaping various institutional or individual data-sharing behaviours.

The present study strengthens the general belief that researchers are (in principle) in favour of open data, with most of them acknowledging that data gathered by others is vital to their work. This concept of knowledge sharing among scientists ('scholarly altruism') has been documented by numerous other studies (Constant et al. 1996, Davenport & Prusak 1998, Kankanhalli et al. 2005, Lin 2008, He & Wei 2009, Hung et al. 2011). However, career level (academic ranking and professional experience) somehow distorted this general pattern, with established researchers being more likely to oppose open access. Piwowar & Chapman (2010), investigating the biomedical sciences, arrived at the opposite conclusion by using bibliographic indices as a proxy for 'experience'. Perhaps the different culture of the biomedical discipline, compared to other life sciences, has shaped this contradictory perception. Kim (2013) also cites distinct data-sharing behaviours in diverse scientific disciplines; e.g. that in the field of life sciences, geneticists were more likely to deny others' requests than non-geneticists. Nevertheless, it must be acknowledged that the problem is far from straightforward. Each scientific discipline generates different types of data, and hence each discipline is expected to have different data-sharing requirements and expectations, which in turn shape distinctive scientific 'cultures'.

In this study we also identified differences in scientists' perceptions related to country/region of professional location. Such a notable difference was apparent between southern and northern Europe, probably related to the divergence in research funding and opportunities. Research is obviously not a priority in southern Europe, and it seems that the difficulties associated with fundraising for conducting research in this region probably manifests into a more 'conservative' view towards data openness. Piwowar & Chapman (2010) also showed that researchers with a corresponding address in the USA were twice as likely to embrace open data.

The most noticeable outcome of the present study was that although in principle, scientists were well disposed to open data and acknowledged that other scientists' data are essential and have been utilizing them, they were quite hesitant to share their own data. The withholding rate reported in this study is considerably high compared to other studies so far (Kim 2013 and references therein), where rates ranged from 3% to 75%, the majority being below 35%. Focusing specifically on the life sciences, these numbers varied from 8% to 32% (Kim 2013 and references therein). This is in fact the 'Selfish Scientist Paradox': scientists want to use other scientists' data but protect their own data.

Although many reasons were brought up to justify data withholding, the main reason was exploiting data to publish. Kim (2013) lists numerous reasons for data withholding: funding agency's policy, contract with industry sponsors, data sensitivity, privacy, losing publication opportunities, facing potential criticism and lack of data repositories. It seems that the current institutional setting and the academic reward system forces scientists to seek opportunities for more publications (and citations) as a way to build a reputation (Gewin 2016). This desire to make the most out of their data and the absolute necessity to publish has affected the viewpoints of even junior researchers in the present study. One would expect that at least emerging scientists would function under the norm of communalism (Merton 1968, Braxton 1986), rather than embracing particularism (Mitroff 1974). Yet, it seems that in the modern scientific arena, the need to capitalize on the data for future career benefits prevails over scholarly altruism.

In spite of the undeniable progress and change of mindset realized throughout human history, as this study also suggests, scholarly altruism is still not the norm, and numerous barriers are blocking the free exchange of scientific information: disciplinary traditions, institutional barriers, lack of technological infrastructure, intellectual property concerns, and individual perceptions (Kim & Stanton 2012). Although the researchers surveyed in the present study indicated confidentiality agreements or the desire to exploit data (by publishing) as their main reservations for not sharing data, other studies have concluded that scientific discipline, institutional environment and individual motivations are crucial factors influencing data-sharing behaviours (e.g. policies developed by funding agencies and journals; university tenure and promotion systems; Borgman 2010).

LITERATURE CITED

- ✦ Akers KG, Doty J (2013) Disciplinary differences in faculty research data management practices and perspectives. *Int J Digit Curation* 8:5–26
- Borgman CL (2007) *Scholarship in the digital age: information, infrastructure, and the internet*. MIT Press, Cambridge, MA
- ✦ Borgman CL (2010) Research data: Who will share what, with whom, when, and why? RatSWD Working Paper No. 161. Available at: <http://dx.doi.org/10.2139/ssrn.1714427>
- Boulton G, Rawlins M, Vallance P, Walport M (2011) Science as a public enterprise: the case for open data. *Lancet* 377: 1633–1635
- Braxton JM (1986) The normative structure of science: social control in the academic profession. In: Smart JC (ed) *Higher education: handbook of theory and research*. Agathon Press, New York, NY, p 309–357
- ✦ Campbell EG, Clarridge BR, Gokhale M, Birenbaum L, Hilgartner S, Holtzman NA, Blumenthal D (2002) Data withholding in academic genetics: evidence from a national survey. *JAMA* 287:473–480
- ✦ Ceci SJ (1988) Scientists' attitudes toward data sharing. *Sci Technol Human Values* 13:45–52
- ✦ Constant D, Sproull L, Kiesler S (1996) The kindness of strangers: the usefulness of electronic weak ties for technical advice. *Organ Sci* 7:119–135
- ✦ DAMVAD (2014) *Sharing and archiving of publicly funded research data*. Report to the Research Council of Norway. DAMVAD Analytics, Copenhagen. www.usit.uio.no/om/organisasjon/uav/itf/saker/forskningsdata/bakgrunn/sharing-and-archiving-research-data.pdf/ (accessed on 4 Sep 2017)
- Davenport TH, Prusak L (1998) *Working knowledge: how organizations manage what they know*. Harvard Business School Press, Boston, MA
- ✦ Davis HM, Vickery JN (2007) Datasets, a shift in the currency of scholarly communication: implications for library collections and acquisitions. *Ser Rev* 33:26–32
- ✦ Editorial (2007a) Got data? *Nat Neurosci* 10:931
- ✦ Editorial (2007b) Compete, collaborate, compel. *Nat Genet* 39:931
- ✦ Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B (2012) The user's view on biodiversity data sharing — investigating facts of acceptance and requirements to realize a sustainable use of research data. *Ecol Inform* 11:25–33
- ✦ European Commission (2011) *Open data: an engine for innovation, growth and transparent governance*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM (2011): 882 final. www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2011/0882/COM_COM%282011%290882_EN.pdf (accessed on 29 Aug 2017)
- ✦ European Commission (2012) *Online survey on scientific information in the digital age*. Office for Official Publications of the European Communities, Luxembourg
- ✦ G8 Science Ministers (2013) G8 Science Ministers statement. www.gov.uk/government/news/g8-science-ministers-statement (accessed on 29 Aug 2017)
- ✦ Gantz J, Reinsel D (2012) *The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the Far East*. IDC analyze the future. IDC, Framingham, MA. <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf> (accessed 12 July 2017)
- ✦ Gewin V (2016) Data sharing: an open mind on open data. *Nature* 529:117–119
- ✦ Goodman LA (1961) Snowball sampling. *Ann Math Stat* 32: 148–170
- ✦ He W, Wei KK (2009) What drives continued knowledge sharing? An investigation of knowledge-contribution and -seeking beliefs. *Decis Support Syst* 46:826–838
- ✦ Hey T, Hey J (2006) e-Science and its implications for the library community. *Libr Hi Tech* 24:515–528
- ✦ Hung SY, Durcikova A, Lai HM, Lin WM (2011) The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. *Int J Hum Comput Stud* 69: 415–427
- ✦ ICSU-WDS (International Council for Science-World Data System) (2015) *WDS data sharing principles*. ICSU-WDS, Tokyo. https://zenodo.org/record/34354/files/WDS_Data_Sharing_Principles_2015.pdf (accessed on 29 Aug 2017)
- ✦ Kankanhalli A, Tan BCY, Wei KK (2005) Contributing knowledge to electronic knowledge repositories: an empirical investigation. *Manage Inf Syst Q* 29:113–143
- Kim Y (2013) *Institutional and individual influences on scientists' data sharing behaviors*. PhD dissertation, School of Information Studies, Syracuse University
- ✦ Kim Y, Stanton JM (2012) *Institutional and individual influences on scientists' data sharing practices*. *J Comput Sci Educ* 3:47–56
- ✦ Kratz JE, Strasser C (2015) *Researcher perspectives on publication and peer review of data*. *PLOS ONE* 10:e0117619
- Likert R (1932) A technique for the measurement of attitudes. *Arch Psychol* 140:1–55
- ✦ Lin CP (2008) *Clarifying the relationship between organizational citizenship behaviors, gender, and knowledge sharing in workplace organizations in Taiwan*. *J Bus Psychol* 22:241–250
- Merton RK (1968) *Social theory and social structure*. Free Press, New York, NY
- ✦ Mitroff I (1974) Norms and counter-norms in a select group of the Apollo moon scientists: a case study of the ambivalence of scientists. *Am Sociol Rev* 39:579–595
- ✦ Mollooy JC (2011) *The Open Knowledge Foundation: open data means better science*. *PLOS Biol* 9:e1001195
- ✦ Nosek BA, Alter G, Banks GC, Borsboom D and others (2015) Promoting an open research culture. *Science* 348: 1422–1425
- ✦ NSF (National Science Foundation) (2010) *Scientists seeking NSF funding will soon be required to submit data management plans*. News release 10-077. NSF, Alexandria, VA. https://www.nsf.gov/news/news_summ.jsp?cntn_id=116928 (accessed on 31 Aug 2017)
- ✦ OECD (2007) *OECD principles and guidelines for access to research data from public funding*. No. 88180 2007. OECD Publications, Paris. www.oecd.org/science/scitech/38500813.pdf (accessed on 29 Aug 2017)
- OECD (2017) *Gross domestic spending on R&D (indicator)*. OECD iLibrary, Paris. doi:10.1787/d8b068b4-en (accessed on 6 Jul 2017)
- ✦ OMB (Executive Office of the President, Office of Management and Budget) (2013) *M-13-13: memorandum for the heads of executive departments and agencies. Open data policy — managing information as an asset*. OMB, Washington, DC. <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf> (accessed on 29 Aug 2017)

- ✦ Piwowar HA, Chapman WW (2010) Public sharing of research datasets: a pilot study of associations. *J Informetrics* 4:148–156
- ✦ Piwowar HA, Vision TJ (2013) Data reuse and the open data citation advantage. *PeerJ* 1:e175
- ✦ Piwowar HA, Day RS, Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. *PLOS ONE* 2:e308
- ✦ Piwowar HA, Becich MJ, Bilofsky H, Crowley RS (2008) Towards a data sharing culture: recommendations for leadership from academic health centers. *PLOS Med* 5:e183
- ✦ Schmidt B, Gemeinholzer B, Treloar A (2016) Open data in global environmental research: the Belmont Forum's open data survey. *PLOS ONE* 11:e0146695
- ✦ Swan A, Brown S (2008) To share or not to share: publication and quality assurance of research data outputs. A report commissioned by the Research Information Network, <https://eprints.soton.ac.uk/id/eprint/266742> (Last accessed: 04 May 2018)
- ✦ Tenopir C, Allard S, Douglass K, Aydinoglu AU and others (2011) Data sharing by scientists: practices and perceptions. *PLOS ONE* 6:e21101
- ✦ Tenopir C, Dalton ED, Allard S, Frame M and others (2015) Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE* 10:e0134826
- ✦ Uhler PF, Schröder P (2007) Open data for global science. *Data Sci J* 6:OD36–OD53
- ✦ UNESCO (2012) Minsk recommendation for open access to scientific information (report of the open access experts from the Eastern and Central Europe). Recommendation of regional consultation on 'Open access to scientific information and research—concept and policies', 05-07 September 2012. UNESCO, Paris. www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/recommendation_minsk_oa_sept2012_en.pdf (accessed on 29 Aug 2017)
- ✦ Van den Eynden V, Corti L, Woollard M, Bishop L, Horton L (2011) Managing and sharing data. UK Data Archive, University of Essex, Colchester
- ✦ Vision TJ (2010) Open data and the social contract of scientific publishing. *BioScience* 60:330
- ✦ Warren E (2016) Strengthening research through data sharing. *N Engl J Med* 375:401–403

*Editorial responsibility: Darryl Macer,
Scottsdale, Arizona, USA*

*Submitted: December 8, 2017; Accepted: February 23, 2018
Proofs received from author(s): May 4, 2018*