Knowledge and attitudes among life scientists towards reproducibility within journal articles

Evanthia Kaimaklioti Samota[1,2], Robert P. Davey[1]
1. Earlham Institute, Norwich, UK; 2. University of East Anglia, Norwich, UK
Corresponding author: evanthia.samota@earlham.ac.uk

Abstract

We constructed a survey to understand how authors and scientists view the issues around reproducibility, and how solutions such as interactive figures could enable the reproducibility of experiments from within a research article. This manuscript reports the results of this survey on the views of 251 researchers, including authors who have published in *eLIFE* Sciences, and those who work at the Norwich Biosciences Institutes (NBI). The survey also outlines to what extent researchers are occupied with reproducing experiments themselves and what are their desirable features of an interactive figure. Respondents considered various features for an interactive figure within a research article that would allow for them to better understand and reproduce *in situ* the experiment presented in the figure. Respondents said that the most important element that would enable the better reproducibility of published research would be that authors describe methods and analyses in detail. The respondents believe that having interactive figures in published papers is a beneficial element. Whilst interactive figures are potential solutions for demonstrating technical reproducibility, we find that there are equally pressing cultural demands on researchers that need to be addressed to achieve greater success in reproducibility in the life sciences.

KEY WORDS: experiments reproducibility; computational experiments; interactive figure; reproducibility metrics; cultural reproducibility

Introduction

Reproducibility is a defining principle of scientific research, and refers to the ability of researchers to replicate the findings of a study using same or similar methods, materials and data as did the original researchers [1]. However, irreproducible experiments are common across all disciplines of life sciences [2]. A recent study showed that 88% of drug-discovery experiments could not be reproduced or replicated even by the original authors, in some cases forcing retraction of the original work [3]. Irreproducible genetic experiments with weak or wrong evidence can have negative implications on our healthcare [4]. For example, 27% of mutations linked to childhood genetic diseases cited in literature have later been discovered to be common polymorphisms or misannotations [5]. While irreproducibility is not confined to biology and medical sciences [6], irreproducible biomedical experiments pose a strong

financial burden on society; an estimated $28 billion was spent on irreproducible biomedical science in 2015 in the USA alone [7].

Reproducibility is an important element of robust science, relating to the way in which conclusions rely on specific analyses or procedures undertaken on experimental systems. It is important to differentiate the definition of reproducibility as a term from repeatability and replicability.

1. **Repeatability** The original researchers using the same data, running precisely the same workflow and getting the same results [8].

2. **Replicability** Different team performing the same experimental setup (in the same or different location) resulting in achieving the same result as the original researchers [8,9]. The replication of computational experiments is termed as ***recomputability*** [9,10]

3. **Reproducibility** Achieving the same precise result by a different team and experimental setup. This might mean running similar or the same data with the same or different workflow [8-10]. It is argued that in many science disciplines reproducibility is more desirable than replicability, as a result needs to be corroborated independently before it can be generally accepted by the scientific community [9].

Computational reproducibility has both technical and cultural aspects. Technical challenges to reproducibility include poorly written, incorrect, or unmaintained software, changes in software libraries on which tools are dependent, or incompatibility between older software and newer operating systems [11]. Cultural challenges include insufficient descriptions of methods, reluctance to publish original data and code under FAIR (Findable, Accessible, Interoperable, and Reusable) principles, and other social factors such as the favouring of high prestige or high impact science publications over performing rigorous and reproducible science.

Several projects have attempted to address some of the technical aspects of reproducibility by making it easier for authors to disseminate fully reproducible workflows and data, and for readers to perform computations. For example: F1000 Living Figure [12]; Whole Tale Project [13]; RetroZIP project (reprozip.org); Python compatible tools and widgets (IPython notebook interactive widgets, Jupyter Notebooks); FigShare (figshare.com) as an example of a scientific data repository; Galaxy [14]; CyVerse (formerly iPlant Collaborative) [15]; myExperiment [16]; UTOPIA [17, 18]; GigaScience Database [19]; Taverna [20-22]; workflow description efforts such as the Common Workflow Language [23]; and Docker (docker.com), Singularity (singularity.lbl.gov) [24], and other container systems.

Even though these tools are widely available, and seem to address many of the issues of technical and cultural reproducibility, they have not yet become a core part of the life sciences experimental and publication lifecycle. There is an apparent disconnection between the development of tools addressing reproducibility and their use by the wider scientific and publishing communities who might benefit from them. However, there have been notable efforts to make this connection. The *Living Figure* by Björn Brembs and Julien Colomb was the first prototype of a dynamic figure that allowed readers to change parameters of a statistical computation underlying a figure [12]. The first *eLIFE* computationally reproducible

article, formed by converting manuscripts created in a specific format (using the Stencila Desktop, stenci.la, and saved as a Document Archive file) into interactive documents, offers more interactivity at the publication level, allowing the reader to "play" with the article and its figures when viewed in a web browser [25].

While there are few incentives to promote cultural reproducibility [26, 27], efforts in most science domains are being made to establish a culture where an expectation to share data for all publications according to the FAIR principles is prioritised. It is widely accepted that better reproducibility will benefit the scientific community and the general public [28, 29]. Although studies have suggested that reproducibility in science is a serious issue [30, 31], with costly repercussions, fewer studies have investigated the attitudes and knowledge of researchers around reproducibility [32] and what would be the most desirable solutions and infrastructures to enable reproducibility. In particular, minimal research has been conducted into the frequency of difficulties experienced with reproducibility, the perception of its importance, and preferences with respect to potential solutions among the general life sciences community. This paper presents a survey that was, in part, designed to inform the design of interactive figures within a journal article by canvassing respondents' preferred features for these figures. We aimed to address this critical gap in reproducibility knowledge, in order to inform the development of tools that better meet the needs of producers and consumers of life science research. We constructed the survey in order to understand how the following are experienced by the respondents:

- *Computational reproducibility*: issues with accessing data, code and methodology parameters, and how solutions such as interactive figures could promote reproducibility from within an article.
- *Cultural reproducibility*: attitudes towards reproducibility, the social factors hindering reproducibility, and interest in interactive figures and their feature preferences.

Methods

Population and sample

Our sample populations were selected to include all life sciences communities across levels of seniority, discipline and level of experience with the issues we wished to survey. The first survey was conducted in November 2016 and sent out to 750 researchers working in the Norwich Biosciences Institutes (NBI) at post-doctoral level or above. The NBI is a partnership of four UK research institutions: the Earlham Institute (formerly known as The Genome Analysis Centre), the John Innes Centre, the Sainsbury Centre, and the Institute of Food Research (now Quadram Institute Bioscience). Invitations to participate were distributed via email, with a link to the survey. The second survey, similar to the first but with amendments and additions, was distributed in February 2017 to a random sample of 1662 active researchers who had published papers in the *eLIFE* journal. Invitations to participate were sent using email by *eLIFE* staff. We achieved an 15% (n = 112) response rate from the NBI researchers, and an 8% response rate from the *eLIFE* survey (n = 139). Table 1 shows the survey questions. Questions were designed to give qualitative and quantitative answers on technical and cultural aspects of reproducibility. Questions assessed the frequency in difficulties encountered in accessing data, the reasons for these difficulties, and how

respondents currently obtain data underlying published articles. They measured understanding of what constitutes reproducibility of experiments, interactive figures, and computationally reproducible data. Finally, we evaluated the perceived benefit of interactive figures and of reproducing computational experiments, and which features of interactive figures would be most desirable.

Table 1: Questions used to survey the knowledge of respondents about research reproducibility.

	Survey questions
1	**How often do you encounter difficulties with working with bioinformatic analysis tools (that are not your own)? (Problems such as: installing, configuring, running the software, working with command line software)?**
2	**How difficult is it to source (or access) the data presented in published papers?**
3	**What difficulties have you encountered in accessing the data described in published papers?**
4	**How are you currently sourcing the data (if applicable)? Select all that apply to you.**
5*	**What is your current understanding of reproducibility of experiments? Please select any that apply. Should you wish to add any additional information, please add it to the "Other" box.**
6*	**Have you ever tried reproducing any published results? Please select the answer that applies best for you.**
7*	**In your opinion, what could be done to make published research more reproducible?**
8	**When thinking about *interactive figures,* what comes to your mind? (please describe of what you understand of what an interactive figure to be, its features, and where you have seen such a feature before if applicable).**
9	**An *interactive figure* is a figure within a paper that is dynamic and becomes "live" when the user interacts with it and where the data displayed changes according to various parameter options. Which of the following features of an interactive figure tool would be good to have? Please rank them in the order of preference, where 1 is the most preferred feature, and 11 the least preferred feature.**
10	**What other features an *interactive figure* could have that were not mentioned in the previous question?**
11	**Do you perceive benefit in being able to publish *interactive figures?***
12	**Does the provision or option of an *interactive figure* in the paper affect your decision in choose the publishing journal or publisher?**
13	**Have you heard of the term *computationally reproducible data*, and do you understand what the term means? If answered yes or unsure, please explain what you understand from the term.**
14	**Would you benefit from being able to automatically reproduce computational experiments, or**

	other analyses (including statistical tests) described within a paper?
15	**How often do you work with *bioinformatic analysis tools* (e.g. assemblers, aligners, structure modelling)?**
16	**Have you received any of the following training? Training whether formal or informal (training through a colleague etc.).**
17	**Which of the following type(s) of data do you work with?**

*Questions indicated with an asterisk were only available to the *eLIFE* survey. Answer options to the questions are shown in Supplementary section 1.

Statistical analysis

Results are typically presented as proportions of those responding, stratified by the respondent's area of work, training received, and version of the survey as appropriate. Chi-square tests for independence were used to test for relationships between responses to specific questions, or whether responses varied between samples. Analysis was conducted using R (version 3.5.2; R Core Team, 2018) and Microsoft Excel. All supplementary figures and data are available on Figshare (see Data Availability).

We assessed if there was a significant difference in the ability and willingness to reproduce published results between the cohort of *eLIFE* respondents who understand the term "computationally reproducible data" and those who do not and whether training received, had an effect. We did not include those that replied "unsure" with regards to their understanding of the term "computationally reproducible data". The respondents who chose "yes tried reproducing results, but unsuccessfully", "have not tried to reproduce results" and "it is not important to reproduce results" were grouped together under "unsuccessfully".

Results

Characteristics of the sample

Figure 1 shows the distribution of areas of work of our respondents, stratified by survey sample. Genomics (proportion in whole sample = 22%), biochemistry (17%), and computational biology (15%) were the most common subject areas endorsed in both NBI and *eLIFE* samples. With regard to how often respondents use bioinformatics tools, 25% replied "never", 39% "rarely", and 36% "often". Many (43%) received statistical training, (31%) bioinformatic training, (20%) computer science training.

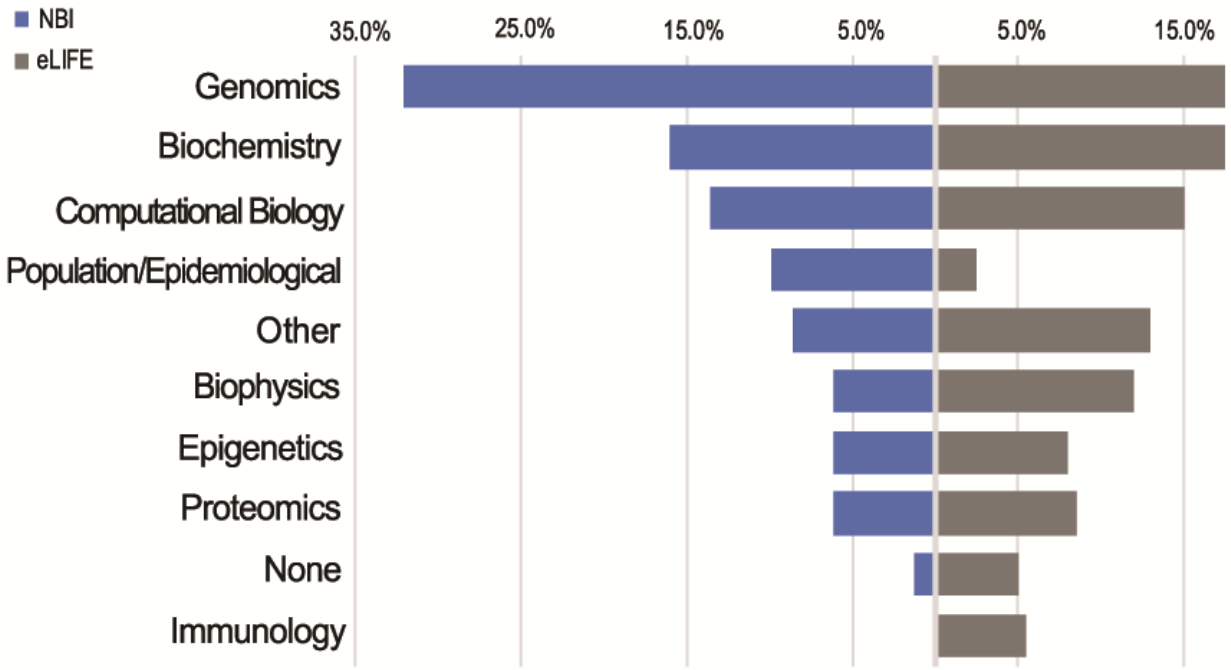

Figure 1: Data types used by NBI and *eLIFE* respondents. Responses were not mutually exclusive. Data type choices were the same as the article data types available in the *eLIFE* article categorisation system.

Access to data and bioinformatics tools

In both samples, 90% of those who responded reported having tried to access data underlying a published research article (Fig. 2). Of those who had tried, few had found this "easy" (14%) or "very easy" (2%) with 41% reporting that the process was "difficult" and 5% "very difficult". Reasons for difficulty were chiefly cultural (Fig. 2), in that the data was not made available alongside the publication (found by 75% of those who had tried to access data), or authors could not be contacted or did not respond to data requests (52%). Relatively few found data unavailable for technical reasons of data size (21%), commercial sensitivity (13%) or confidentiality (12%). With respect to data sources, 57% of the total sample have used open public databases, 48% reported data was available with a link in the paper, and 47% had needed to contact authors.

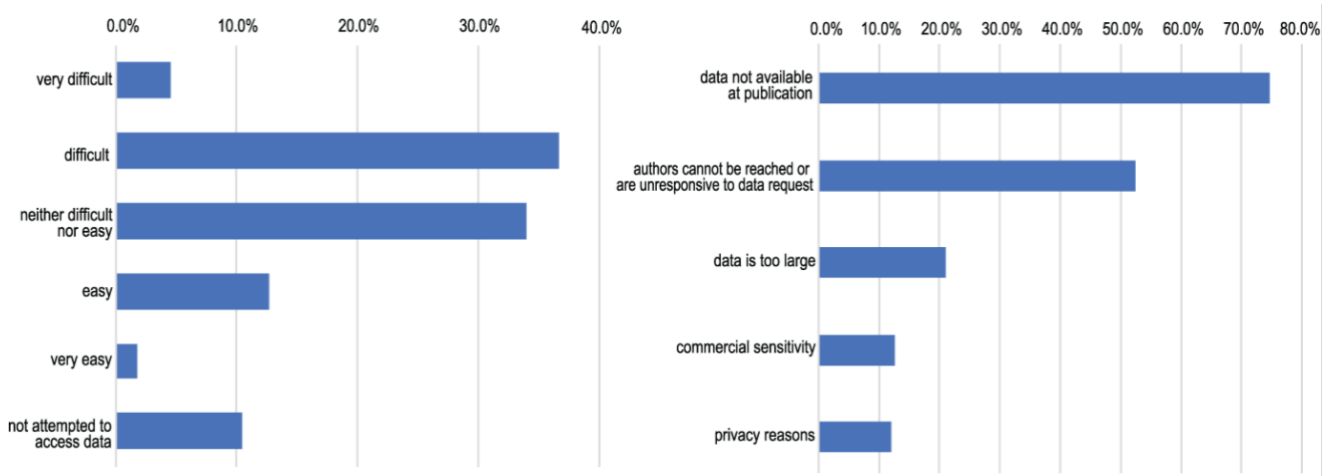

Figure 2. Left panel: Difficulty encountered accessing data underlying published research. Whether respondents have attempted to access data underlying previous publications and the level of difficulty typically encountered in doing so. **Right panel: Reasons given for difficulty accessing data.** The reasons given by respondents for being unable to access data (restricted to those who have attempted to access data).

Very few of the respondents either "never" (2%) or "rarely" (8%) had problems with running, installing, configuring bioinformatics software. Problems with software were encountered "often" (29%) or "very often" (15%) suggesting that nearly half of respondents regularly encountered technical barriers to computational reproducibility.

Understanding of reproducibility, training and successful replication

The majority of respondents reported that they understood the term "reproducibility of experiments" and selected the correct explanations for the term. However, there is still confusion between the terms: repeatability, replicability and reproducibility. Many (43%) of respondents chose the repeatability and replicability definitions: "the original authors or others running the same data with precisely the same workflow and getting the same results". In contrast, most (52%) participants did not know what the term "computationally reproducible data" means, while 26% did know and 22% were unsure. We received several explanations (free text responses) of the term "computationally reproducible data", some of which were more accurate than others (Supplementary section, free responses to question 13).

Some (18%) reported not attempting to reproduce or revalidate published research. Very few (n = 5; 6%) of the sample endorsed the option that "it is not important to reproduce other people's published results" (Supplementary figure 1). Even though the majority (60%) reported successfully reproducing published results, almost a quarter of the respondents found that their efforts to reproduce any results were unsuccessful (23%). Table 2 shows the ability of respondents in reproducing experiments stratified by the understanding of the term "computationally reproducible data" and the training received (bioinformatics, computer science, statistics). We found significant difference between the ability to reproduce published experiments and knowing the meaning of the term "computationally reproducible data". Among the 25 respondents who understood the term "computationally reproducible data", 18

(72%) had successfully reproduced previous work, compared to only 26 (52%) of the 50 who responded that they did not understand the term (Chi-square test for independence, p = 0.048). The training variable did not show any significant distribution. However, when testing with the responses "yes tried reproducing results, but unsuccessfully", "have not tried to reproduce results" and "it is not important to reproduce results" not grouped together under "unsuccessfully" in order to get an indication of how willingness and success together differed between the training groups, we found a significant distribution (see Supplementary Table 1). The distribution of the training variable with those who received computer science training and those without was significantly different (Fisher exact test for independence, p = 0.018). It appears that respondents with computer science training are less likely to have tried to reproduce an experiment but be more likely to succeed when they did try.

There was no evidence for a difference in the ability and willingness to reproduce published results between the respondents who use bioinformatics tools often, and those who use them rarely or never (data not shown). The majority of the respondents who use bioinformatics tools often were coming from the scientific backgrounds of Biophysics, Biochemistry, Computational Biology and Genomics. Most of the respondents who answered "reproducibility is not important" and "haven't tried reproducing experiments" were scientists coming from disciplines using computational or bioinformatics tools "rarely" or "never" (Supplementary Table 2).

Table 2: Success in reproducing any published results stratified by their knowledge of term "computationally reproducible data" and training received.

	Number (% of total sample)	**Success in reproducing any published results**		
Variable		**Successful (% within variable)**	**Not Successful* (% within variable)**	**P-value**
Knowledge of term *"computationally reproducible data"* *(n = 75)*				
Yes	25 (33.3)	18 (72)	7 (28)	0.048**
No	50 (66.7)	24 (48)	26 (33)	
Training *(n = 90)*				
Bioinformatics	42 (46.7)	26 (61.9)	16 (38.1)	0.73
Not trained in Bioinformatics	48 (53.3)	28 (58.3)	20 (41.7)	
Computer Science	33 (36.7)	21 (63.6)	12 (36.4)	0.59
Not trained in Computer Science	57 (63.3)	33 (57.9)	24 (42.1)	
Statistics	71 (78.9)	42 (59.2)	29 (40.8)	0.75
Not trained in Statistics	19 (21.1)	12 (63.2)	7 (36.8)	
No training	10 (11.1)	6 (60)	4 (40)	0.73***
All other training	80 (88.8)	48 (60)	32 (40)	

n is different for the two variables as not all participants answered all the questions

*Unsuccessful includes answers: “Yes, I have tried reproducing published results, but I have been unsuccessful in producing any results, or same results”, “No, I have never tried reproducing any published results” and “It is not important to reproduce other people’s published results”
**Statistically significant at the level of P<0.05
***Chi-square statistic with Yates correction, applied when expected frequencies were lower than 5

Improving Reproducibility of Published Research

The vast majority (91%) of respondents stated that authors describing all methodology steps in detail, including any formulae analysing the data, would be the most effective way to make published science more reproducible. Around half endorsed the view that “authors should provide the source code of any custom software used to analyse the data and that the software code is well documented” (53%), and that authors provide a link to the raw data (49%) (Supplementary figure 2). Two respondents suggested that achieving better science reproducibility would be easier if funding was more readily available for reproducing the results of others and if there were opportunities to publish the reproduced results (Supplementary section, free responses). Within the same context, some respondents recognised the current culture in science that there are not sufficient incentives in publishing reproducible (or indeed negative findings) papers, but rather being rewarded in publishing as many papers as possible in high Impact Factor journals (Supplementary section, free responses).

Interactive Figures

Participants ranked in terms of preference potential features for an interactive figure within an article. These included choices such as “easy to manipulate” as the most preferred, and have “easy to define parameters” (Fig. 3). Generally, the answers from both the *eLIFE* and NBI surveys followed similar trends. Furthermore, free text responses were collected, and most respondents stated that having further insights into the data presented in the figure would be beneficial (Supplementary section, free responses).

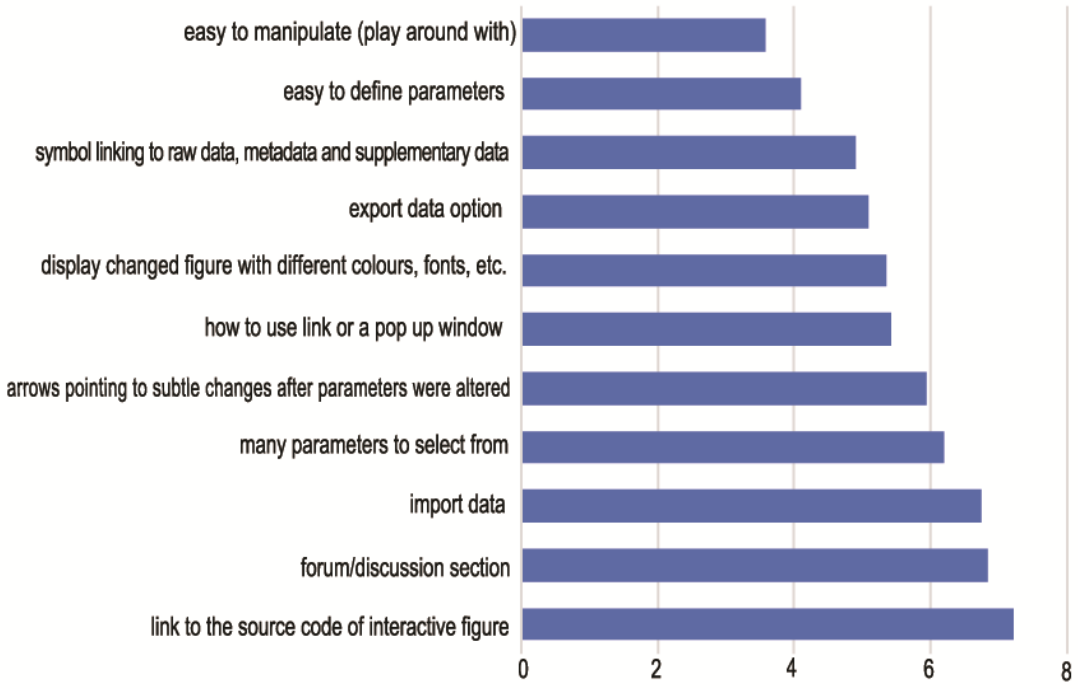

Figure 3. Preferred features for the interactive figure. Responses to question 9: Respondents were asked to rank in order of preference the above features, with 1 most preferred feature, to 11 the least preferred feature. The average score for each feature was calculated in order of preference as selected by the respondents from both NBI and *eLIFE* surveys. The lower the average score value (x-axis), the more preferred the feature (y-axis).

The majority of the respondents perceive a benefit in having interactive figures in published papers for both readers and authors (Fig. 4). Examples of insights included: the interactive figure would allow visualising further points on the plot from data in the supplementary section, as well as be able to alter the data that is presented in the figure; having an interactive figure as a movie, or to display protein 3D structures, would be beneficial to readers. The remaining responses we categorised as software related, which included suggestions of software that could be used to produce a figure that can be interactive, such as R Shiny (shiny.studio.com). A moderate proportion of *eLIFE* respondents (19%) and NBI (27%) stated that they had no opinion on the utility of interactive figures. Free text answers for this group suggested that they had never seen or interacted with such a figure before, and no indication was given that an interactive figure would help their work.

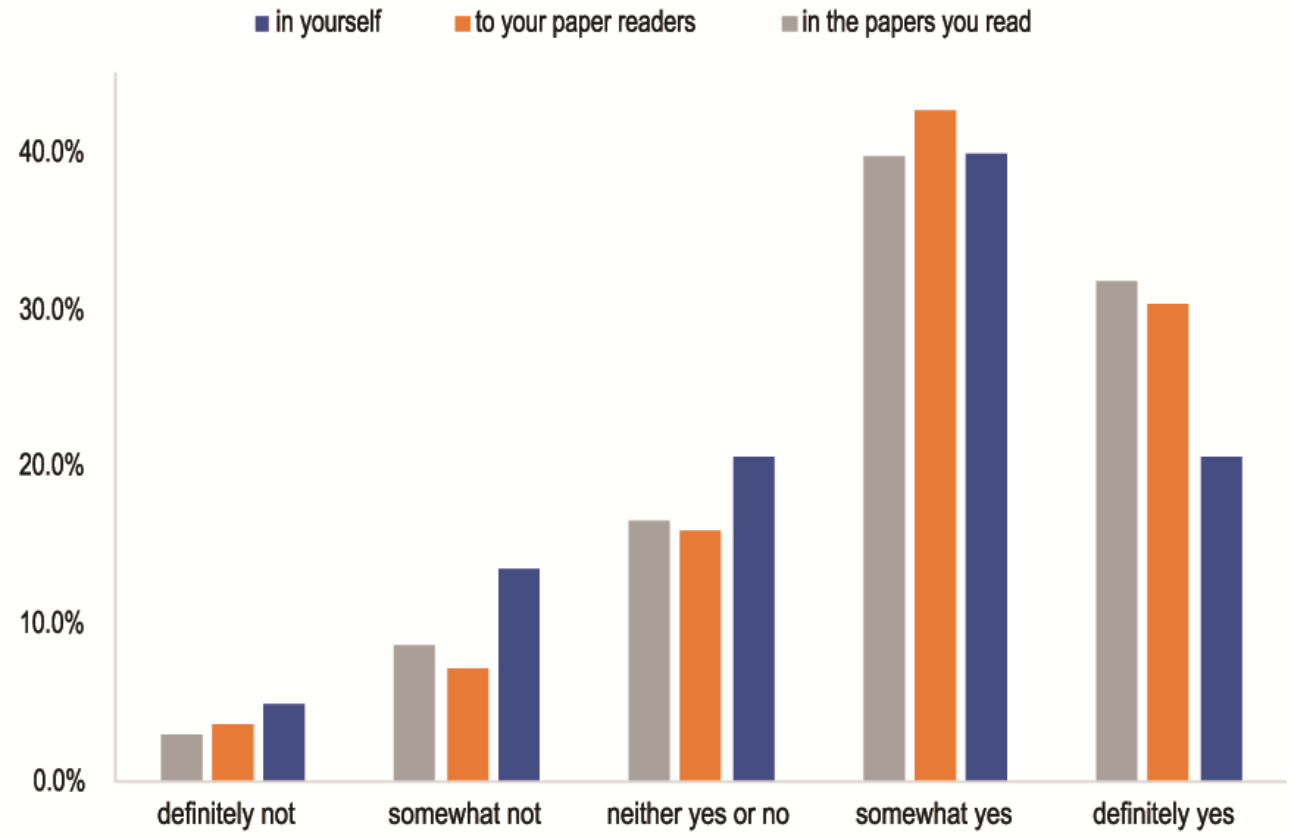

Figure 4 The level of perception of benefit to having the ability to publish papers with interactive figures. The benefit to the author, to the readers of the author's papers and to the papers the author reads. Answers include the responses from both NBI and *eLIFE* surveys for question 11.

The majority of the respondents also said that they see benefit in automatically reproducing computational experiments, and manipulating and interacting with parameters in computational analysis workflows. Equally favourable was to be able to computationally reproduce statistical analyses (Fig. 5). Despite this perceived benefit, most respondents (61%) indicated that the ability to include an interactive figure would not affect their choice of journal when seeking to publish their research.

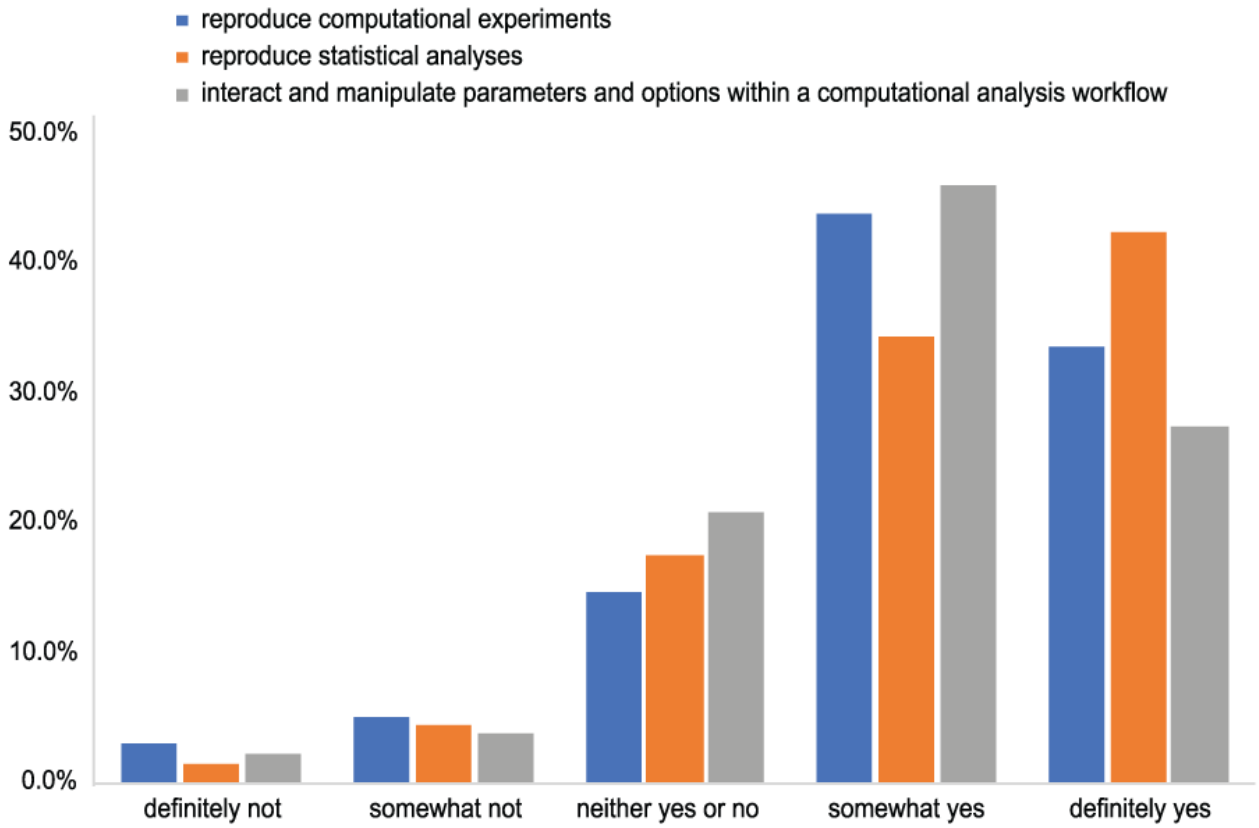

Figure 5. Assessment of perceived benefit for automatically reproducing computational experiments or other analyses (including statistical tests). Responses from both NBI and *eLIFE* for question 14.

Discussion

This study highlights the difficulties currently experienced in reproducing experiments, and expressed positive attitudes of scientists involved in the current publishing system towards enabling and promoting reproducibility of published experiments through interactive elements in online publications. All respondents of the survey were active life sciences researchers and therefore we believe the opinions collected are representative of researchers in life sciences who are routinely reading and publishing research. While progress has been made in publishing standards across all life science disciplines, the opinions of the respondents reflect previously published shortcomings of the publishing procedures [33-35]: lack of data and code provision; storage standards; not including or requiring detailed description of methods and code structure in the published papers. When data is difficult to

obtain the reproducibility problem is exacerbated. However, the level of interest and incentives in reproducing published research is at its infancy, or it is not the researchers' priority, something also mentioned extensively in previous literature [36- 37]. Responses to our surveys suggested that most life scientists understand that science becomes implicitly more reproducible if methods (including data, analysis, and code) are well-described and available, and perceive a potential benefit of tools that enable this. Respondents stated they could see the benefit in having interactive figures for their readers and being able as authors to present their data as interactive figures. However, the availability of this facility would not affect their decisions on where to publish. Despite technologies existing to aid reproducibility and authors know they are beneficial, many scientific publications do not meet basic standards of reproducibility. Respondents endorsed articles which include interactive elements, where access to the raw data, code, and detailed analysis steps in the form of an interactive figure would help article readers better understand the paper and the experimental design and methodology, and improve the reproducibility of the experiment presented in the interactive figure, especially computational experiments. This contradiction suggests that cultural factors play an underestimated role in reproducibility.

Retraction rates would suggest that the current publishing system is yet to provide a mechanism to reliably check whether a published study is reproducible [38]. There remains a perception that researchers do not get credit for reproducing the work of others or publishing negative results. Whilst some journals explicitly state that they welcome negative results articles (e.g. PLOS One "Missing Pieces" collection), this is by no means the norm in life science publishing as evidenced by low, and dropping, publication rates of negative findings [39, 40]. Ideally the publication system would enable checking of reproducibility by reviewers and editors at the peer-review stage, with authors providing all data (including raw data), a full description of methods including statistical analysis parameters, any negative findings based on previous work and open source software code [41]. Peer reviewers would then be better able to check for anomalies, and editors could perform the final check to ensure that the science paper to be published is presenting true, valid, and reproducible research. Some respondents have suggested that if reviewers and/or editors were monetarily compensated, spending time to reproduce or validate the computational experiments in manuscripts would become more feasible, and would aid the irreproducibility issue. However, paying reviewers does not necessarily ensure that they would be more diligent in checking or trying to reproduce results [42] and there must be optimal ways to ensure effective pressure is placed upon the authors and publishing journals to have better publication standards [43, 44]. The increasing adoption by journals of reporting standards for experimental design and results, provide a framework for harmonising the description of scientific processes to enable reproducibility. However, these standards are not universally enforced [45]. Similarly, concrete funding within research grants for implementing reproducibility itself, manifested as actionable Data Management Plans (dcc.ac.uk, 2019) rather than what is currently a by-product of the publishing process, could give a level of confidence to researchers who would want to reproduce previous work and incorporate that data in their own projects.

Our findings are in accordance with the current literature [30, 46] that highlight that the lack of data access at the publication stage is one of the major reasons leading to the irreproducibility of published studies. Even with current policies mandating data openness [28, 29], authors still fail to include their data alongside their publication. This is supported by our

findings that the majority of respondents replied that data is either not available upon publication (57%) or authors cannot be reached/are unresponsive to data provision requests (44%). This continues to be a cultural artefact of using a paper's methods section as a description of steps to reproduce analysis, rather than a fully reproducible solution involving public data repositories, open source code, and comprehensive documentation. Pre-print servers such as bioRxiv have been taken up rapidly [47], especially in the genomics and bioinformatics domains, and this has the potential to remove delays in publication whilst simultaneously providing a "line in the sand" with a Digital Object Identifier (DOI) and maintaining the requirements for FAIR data. In some cases, sensitivity of data might discourage authors from data sharing, [48, 49], but this reason was only reported by a small proportion of our respondents. Whilst there are efforts that attempt to apply the FAIR principles to clinical data, such as in the case of the OpenTrials database [50], they are by no means ubiquitous.

Reproducibility of experiments could be improved with better storage solutions for large data files and citing them within the publication document, especially those in the order of terabytes, for their proper reusability [51, 52]. Currently, there are several services that allow storing large data files and perform cloud analyses, such as CyVerse, Amazon Web Services [53, 54] and Google Genomics (cloud.google.com/genomics). Despite the potential advantage these services can provide for data *accessibility*, they do not implicitly solve the problem of data *reusability*. This is mostly apparent when data is too large to be stored locally or transferred via slow internet connections or there is no route to attach metadata that describes the datasets sufficiently for reuse or integration with other datasets. There is also the question of data repository *longevity* - who funds the repositories for decades into the future? Data within public repositories with specific deposition requirements (such as the EMBL-EBI European Nucleotide Archive, ebi.ac.uk/ena), might not be associated or annotated with standardised metadata that describes it accurately [55], rather the bare minimum for deposition. In addition, corresponding authors often move on from projects and institutions or the authors themselves can no longer access the data, meaning "data available on request" ceases to be a viable option to source data or explanations of methods.

In a 2016 survey of 3987 National Science Foundation Directorate of Biological Sciences principal investigators (BIO PIs), expressed their greatest unmet training needs by their institutions [56]. These were in the areas of integration of multiple data (89%), data management and metadata (78%) and scaling analysis to cloud/high performance computing (71%). The aforementioned data and computing elements are integral to the correct knowledge "how to" for research reproducibility. Our findings indicated that those who stated they had experience in informatics also stated they are better able to attempt and reproduce results. Practical bioinformatics and data management training, rather than in specific tools, may be an effective way of reinforcing the notion that researchers' contributions towards reproducibility are a responsibility that requires active planning and execution. This may be especially effective when considering the training requirements of wet-lab and field scientists, who are becoming increasingly responsible for larger and more complex computational datasets. Further research needs to be undertaken to better understand how researchers' competence in computational reproducibility may be linked to their level of informatics training.

Respondents mentioned that there are word count restrictions in papers, and journals often ask authors to shorten methods sections and perhaps move text to supplementary information placed many times in an unorganised fashion or having to remove it altogether. This is a legacy product of the hard-copy publishing era and, readability aside, word limits are not consequential for internet journals. Even so, if the word count limit was only applicable to the introduction, results and discussion sections, then the authors could describe methods in more detail within the paper, without having to move that valuable information in the supplementary section. When methods are citing methodology techniques as described in other papers, where those original references are hard to obtain, typically through closed access practices or by request mechanisms as noted above, then this can be an additional barrier to the reproducibility of the experiment. This suggests that there are benefits to describing the methods in detail and stating that they are similar to certain (cited) references as well as document the laboratory's expertise in a particular method. However, multi-institutional or consortium papers are becoming more common with ever-increasing numbers of authors on papers, which adds complexity to how authors should describe every previous method available that underpins their research [57]. There is no obvious solution to this issue. Highly specialised methods (e.g. electrophysiology expertise, requirements for large computational resources or knowledge of complex bioinformatics algorithms) and specific reagents (e.g. cell lines, antibodies) might not be readily available to other research groups. As stated by some respondents, in certain cases the effective reproducibility of experiments is obstructed by numerical issues with very small or very large matrices or datasets, or differing versions of analysis software used, perhaps to address bugs in analytical code, will cause a variation in the reproduced results.

Previous studies have provided strong evidence that there is a need for better technical systems and platforms to enable and promote the reproducibility of experiments. We provide additional evidence that that paper authors and readers perceive a benefit from having an interactive figure that would allow for the reproducibility of the experiment shown in the figure. The figure would give access to the raw data, code and detailed data analysis steps, allow for *in situ* reproducing computational experiments by re-running code including statistical analyses “live” within the paper. The findings of this survey including understanding what is desirable for interactive figures, helped the development of two prototypes of interactive figures (see Data and Code availability) and subsequently the creation of *eLIFE’s* first computationally reproducible document [25]. Despite the benefits that interactive documents and figures can provide to the publishing system, and that those benefits that are in demand by the scientific community, work is needed in order to promote and support their use. Given the diversity of biological datasets and ever-evolving methods for data generation and analysis, it is unlikely that a single interactive figure infrastructure type can support all types of data. More research into how different types of data can be supported and presented in papers with interactivity needs to be undertaken. Yet problems with data availability and data sizes will persist - many studies comprise datasets that are too large to upload and render within web browsers in a reasonable timescale. Even if the data are available through well-funded repositories with fast data transfers, e.g. the INSDC databases (insdc.org), are publishers ready to bear the extra costs of supporting the infrastructure and people required to develop or maintain such interactive systems in the long run? These are questions that

need to be further investigated, particularly when considering any form of industry standardisation of such interactivity in the publishing system.

We show that providing tools to scientists who are not computationally aware also requires a change in culture, as many aspects of computational reproducibility require a change in publishing behaviour and competence in the informatics domain. Encouraging and incentivising scientists to conduct transparent, reproducible and replicable research should be prioritised to help solve the irreproducibility issue, and implementing hiring practices with open science at the core of research roles [58] will encourage attitudes to change across faculty departments and institutions.

Another potential solution to the reproducibility crisis is to identify better (quantifiable) metrics of research reproducibility and its scientific impact. The current assessment of the impact of research articles are a set of quantifiable metrics that do not evaluate research reproducibility, but stakeholders are starting to request that checklists and tools are provided to improve these assessments [59]. It is harder to find a better approach that is based on a thoroughly informed analysis by unbiased experts in the field that would quantify the reproducibility level of the research article [60]. That said, top-down requirements from journals and funders to release reproducible data and code may go some way to improving computational reproducibility within the life sciences, but this will also rely on the availability of technical solutions that are accessible and useful to the majority of scientists.

Opinions are mixed regarding the extent and severity of the reproducibility crisis however our study and previous studies are stressing the need to find effective solutions towards solving the reproducibility issue. Steps towards modernising the publishing system by incorporating interactivity with interactive figures are deemed desirable. This may be a good starting point for improving research reproducibility by reproducing experiments *in situ*. From our findings, and given the ongoing release of tools and platforms for technical reproducibility, future efforts should be spent in tackling the cultural behaviour of scientists, especially when faced with the need to publish for career progression.

Acknowledgements

This project is funded by a BBSRC iCASE Studentship (project reference: BB/M017176/1). We would like to thank all the respondents of the surveys for their time. We would also like to thank George Savva from the Quadram Institute (QIB, UK) for comments and suggestions for this manuscript; *eLIFE* Sciences Publications Ltd, with whom the corresponding author collaborates as an iCASE student; as well as Ian Mulvany, former *eLIFE* Head of Development, for his help in developing the survey questionnaire.

Data and Code Availability

All data files are available via this url: https://doi.org/10.6084/m9.figshare.c.4436912.v6
Prototypes of interactives figures developed by the corresponding author are available via these GitHub repositories: https://github.com/code56/nodeServerSimpleFig and https://github.com/code56/prototype_article_interactive_figure

References

1. Goodman S, Fanelli D, Ioannidis J. What does research reproducibility mean?. Science Translational Medicine. 2016;8(341):341ps12.
2. Grant B. Science's Reproducibility Problem. The Scientist. 18 Dec 2012. Available from: https://www.the-scientist.com/news-opinion/sciences-reproducibility-problem-40031 Cited 9 April 2016.
3. Begley C, Ellis L. Raise standards for preclinical cancer research. Nature. 2012;483(7391):531-533.
4. Yong E. Reproducibility problems in genetics research may be jeopardizing lives. 2015 Dec 17 [cited 12 February 2015]. In: Genetic Literacy Project [Internet]. Pennsylvania: Genetic Literacy Project. [about 1 screen]. Available from: https://geneticliteracyproject.org/2015/12/17/reproducibility-problems-genetics-research-may-costing-lives/
5. Bell C, Dinwiddie D, Miller N, Hateley S, Ganusova E, Mudge J et al. Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing. Science Translational Medicine. 2011;3(65):65ra4.
6. Ioannidis J, Doucouliagos C. What's to know about the credibility of empirical economics?. Journal of Economic Surveys. 2013;27(5):997-1004. doi: 10.1111/joes.12032.
7. Freedman L, Cockburn I, Simcoe T. The Economics of Reproducibility in Preclinical Research. PLOS Biology. 2015;13(6):e1002165. doi: 10.1371/journal.pbio.1002165.
8. Plesser H. Reproducibility vs. Replicability: A brief History of a Confused Terminology. Frontiers in Neuroinformatics. 2018;11. doi: 10.3389/fninf.2017.00076.
9. Drummond C. Replicability is not Reproducibility: Nor is it Good Science. Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th ICML., 2009. Available from: http://cogprints.org/7691/7/ICMLws09.pdf.
10. Gent I. The recomputation manifesto; 2013. Preprint. Available from: arXiv: 1304.3674v1. Cited 22 April 2016.
11. Cataldo M, Mockus A, Roberts J, Herbsleb J. Software Dependencies, Work Dependencies, and Their Impact on Failures. IEEE Transactions on Software Engineering. 2009;35(6):864-878. doi: 10.1109/tse.2009.42.
12. Colomb J, Brembs B. Sub-strains of Drosophila Canton-S differ markedly in their locomotor behavior. F1000Research. 2015;3:176. doi: 10.12688/f1000research.4263.2.
13. Brinckman A, Chard K, Gaffney N, Hategan M, Jones M, Kowalik K et al. Computing environments for reproducibility: Capturing the "Whole Tale". Future Generation Computer Systems. 2018;94:854-867. doi: 10.1016/j.future.2017.12.029.
14. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research. 2018;46(W1):W537-W544. doi: 10.1093/nar/gky379.
15. Goff SA, Vaughn M, Mckay S, Lyons E, Stapleton AE, Gessler D, et al. The iPlant Collaborative: Cyberinfrastructure for Plant Biology. Frontiers in Plant Science. 2011;2. doi: 10.3389/fpls.2011.00034.

16. Goble C, Bhagat J, Aleksejevs S, Cruickshank D, Michaelides D, Newman D et al. myExperiment: a repository and social network for the sharing of bioinformatics workflows. Nucleic Acids Research. 2010;38(suppl_2):W677-W682. doi: 10.1093/nar/gkq429.
17. Pettifer S, Sinnott J, Attwood T. UTOPIA—User-Friendly Tools for Operating Informatics Applications. Comparative and Functional Genomics. 2004;5(1):56-60. doi: 10.1002/cfg.359.
18. Pettifer S, Thorne D, McDermott P, Marsh J, Villéger A, Kell D et al. Visualising biological data: a semantic approach to tool and database integration. BMC Bioinformatics. 2009;10(S6). doi: 10.1186/1471-2105-10-s6-s19.
19. Sneddon T, Li P, Edmunds S. GigaDB: announcing the GigaScience database. GigaScience. 2012;1(1). doi: 10.1186/2047-217x-1-11.
20. Wolstencroft K, Haines R, Fellows D, Williams A, Withers D, Owen S et al. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. Nucleic Acids Research. 2013;41(W1):W557-W561. doi: 10.1093/nar/gkt328.
21. Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M et al. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics. 2004;20(17):3045-3054. doi: 10.1093/bioinformatics/bth361.
22. Hull D, Wolstencroft K, Stevens R, Goble C, Pocock M, Li P et al. Taverna: a tool for building and running workflows of services. Nucleic Acids Research. 2006;34(Web Server):W729-W732. doi: 10.1093/nar/gkl320.
23. Peter Amstutz, Michael R. Crusoe, Nebojša Tijanić (editors), Brad Chapman, John Chilton, Michael Heuer et al. Common Workflow Language, v1.0**.** Specification, *Common Workflow Language working group. 2016*. Available from: https://w3id.org/cwl/v1.0/ doi:10.6084/m9.figshare.3115156.v2.
24. Kurtzer G, Sochat V, Bauer M. Singularity: Scientific containers for mobility of compute. PLOS ONE. 2017;12(5):e0177459. doi: 10.1371/journal.pone.0177459.
25. Maciocci G, Aufreiter M, Bentley N. Introducing *eLife*'s first computationally reproducible article. 2019 Feb 20 [cited 20 February 2019]. In: eLife Labs Blog [Internet]. Cambridge – [about 5 screens]. Available from: https://elifesciences.org/labs/ad58f08d/introducing-elife-s-first-computationally-reproducible-article.
26. Higginson A, Munafò M. Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. PLOS Biology. 2016;14(11):e2000995. doi: 10.1371/journal.pbio.2000995.
27. Pusztai L, Hatzis C, Andre F. Reproducibility of research and preclinical validation: problems and solutions. Nat Rev Clin Oncol [Internet]. Nature Publishing Group; 2013; 10(12):720–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24080600
28. National Institutes of Health Plan for increasing access to scientific publications and digital scientific data from NIH funded scientific research, [Internet]. NIH. 2015 [cited 5 May 2017]. Available from: https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf.
29. Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016; 3:160018. doi: 10.1038/sdata.2016.18.

30. Pulverer B. Reproducibility blues. The EMBO Journal. 2015;34(22):2721-2724. doi: 10.15252/embj.201570090.
31. Stodden V, Guo P, Ma Z. Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. PLOS ONE. 2013;8(6):e67111. doi: 10.1371/journal.pone.0067111.
32. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature.* 2016;*533(7604), 452–454.*doi:10.1038/533452a
33. Müller H, Naumann F, Freytag J. Data Quality in Genome Databases. Proc Conf Inf Qual (IQ 03). 2003;269–284. Available from: https://www.semanticscholar.org/paper/Data-Quality-in-Genome-Databases-Müller-Naumann/9dc6c5f4fe99caab9b4ba0b1eedc56f747c17194
34. Marx V. The big challenges of big data. Nature. 2013;498(7453):255-260. doi: 10.1038/498255a.
35. Stodden V. Reproducing Statistical Results. Annual Review of Statistics and Its Application. 2015;2(1):1-19. doi: 10.1146/annurev-statistics-010814-020127.
36. B. A. Nosek et al., Promoting an open research culture. Science. 2015; 348, 1422–1425. doi: 10.1126/science.aab2374.
37. Collins F, Tabak L. Policy: NIH plans to enhance reproducibility. Nature. 2014;505(7485):612-613. doi: 10.1038/505612a.
38. Cokol M, Ozbay F, Rodriguez-Esteban R. Retraction rates are on the rise. EMBO reports. 2008;9(1):2-2. doi: 10.1038/sj.embor.7401143.
39. Franco A, Malhotra N, Simonovits G. Publication bias in the social sciences: Unlocking the file drawer. Science. 2014;345(6203):1502-1505. doi: 10.1126/science.1255484.
40. Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics. 2011;90(3):891-904. doi: 10.1007/s11192-011-0494-7.
41. Iqbal S, Wallach J, Khoury M, Schully S, Ioannidis J. Reproducible Research Practices and Transparency across the Biomedical Literature. PLOS Biology. 2016;14(1):e1002333. doi: 10.1371/journal.pbio.1002333.
42. Hershey N. Compensation and Accountability: The Way to Improve Peer Review. Quality Assurance and Utilization Review. 1992;7(1):23-29. doi: 10.1177/106286069200700104.
43. Announcement: Reducing our irreproducibility. Nature. 2013;496(7446):398-398. doi: 10.1038/496398a.
44. Pusztai L, Hatzis C, Andre F. Reproducibility of research and preclinical validation: problems and solutions. Nature Reviews Clinical Oncology. 2013;10(12):720-724. doi: 10.1038/nrclinonc.2013.171.
45. Moher D. Reporting guidelines: doing better for readers. BMC Medicine. 2018;16(1). doi: 10.1186/s12916-018-1226-0.
46. Berg J. Progress on reproducibility. Science. 2018;359(6371):9-9. doi: 10.1126/science.aar8654.
47. Abdill R, Blekhman R. Tracking the popularity and outcomes of all bioRxiv preprints; 2019. Preprint. Available from BioRxiv: 10.1101/515643. Cited 3 April 2019.
48. Figueiredo AS. Data Sharing: Convert Challenges into Opportunities. Front Public Health. Frontiers Media SA; 2017; 5:327.doi:10.3389/fpubh.2017.00327.
49. Hollis KF. To Share or Not to Share: Ethical Acquisition and Use of Medical Data.

AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci [Internet]. 2016;2016:420–7.

50. Philip Chen CL, Zhang CY. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Inf Sci (Ny). Elsevier Inc.; 2014 Aug 10;275:314. doi: 10.1016/j.ins.2014.01.015
51. Poldrack RA, Gorgolewski KJ. Making big data open: Data sharing in neuroimaging. Nat Neurosci. 2014;17(11):1510–7. doi: 10.1038/nn.3818.
52. Faniel IM, Zimmerman A. Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse. Int J Digit Curation. 2011;6(1):58–69. doi: 10.2218/ijdc.v6i1.172.
53. Amazon Web Services (AWS) - Cloud Computing Services [Internet]. Amazon Web Services, Inc. 2019 [cited 11 April 2019]. Available from: https://aws.amazon.com/
54. Hazelhurst S. Scientific computing using virtual high-performance computing: a case study using the Amazon elastic computing cloud. Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries riding the wave of technology - SAICSIT '08.; 2008 Jan 1;94–103. doi: 10.1145/1456659.1456671
55. Attwood TK, Kell DB, McDermott P, Marsh J, Pettifer SR, Thorne D. Calling International Rescue: knowledge lost in literature and data landslide! Biochem J. 2009;424:317–33. doi: 10.1042/BJ20091474.
56. Barone L, Williams J, Micklos D. Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLOS Comput Biol. 2017;13(10). doi: 10.1371/journal.pcbi.1005755.
57. Gonsalves, A. Lessons learned on consortium-based research in climate change and development. CARIAA Working Paper no. 1. 2014 [cited 2019 Jan 1]. In: International Development Research Centre [Internet]. Ottawa, Canada and UK Aid, London, United Kingdom. Available from: www.idrc.ca/cariaa
58. Schönbrodt F. Changing hiring practices towards research transparency: The first open science statement in a professorship advertisement. 2016 Jan 6 [cited 2016 Mar 16]. In: Felix Schönbrodt's blog [Internet] [about 1 screen]. Available from: https://www.nicebread.de/open-science-hiring-practices/.
59. Wellcome Trust. Request for Information (RFI) A software tool to assess the FAIRness of research outputs against a structured checklist of requirements [FAIRWare] [Internet]. Wellcome Trust; 2018 [cited 5 March 2019]. Available from: https://wellcome.ac.uk/sites/default/files/FAIR-checking-software-request-for-information.pdf.
60. Flier JS. Irreproducibility of published bioscience research: Diagnosis, pathogenesis and therapy. Mol Metab. 2017 Nov 21;6(1):2–9. doi: 10.1016/j.molmet.2016.11.006.