2019

A SOCIO-TECHNICAL PERSPECTIVE ON REPRODUCIBILITY IN RESEARCH DATA MANAGEMENT

Armel Lefebvre
Utrecht University, a.e.j.lefebvre@uu.nl

Marco Spruit
Utrecht University, Utrecht, Netherlands, m.r.spruit@uu.nl

A SOCIO-TECHNICAL PERSPECTIVE ON REPRODUCIBILITY IN RESEARCH DATA MANAGEMENT

Research full-length paper

Track N°11

Lefebvre, Armel, Utrecht University, Utrecht, The Netherlands, a.e.j.lefebvre@uu.nl

Spruit, Marco, Utrecht University, Utrecht, The Netherlands, m.r.spruit@uu.nl

Abstract

The Open Science paradigm has brought the dissemination of experimental artifacts on the agenda of funding agencies, research institutions, and academic publishers. Managing research data is a crucial part of guaranteeing the reusability and reproducibility of published results. In this research, we suggest a conceptualization of reproducibility based on threats, risks, and vulnerabilities identified in current research data management (RDM) practices. By doing so, we can describe a range of threats to reproducibility and pinpoint areas where current RDM practices and the scholarly communication infrastructure insufficiently address these threats. Further, we elaborate on a socio-technical approach to reproducibility in RDM by collecting evidence from researchers and scientific publications. We show that the STS approach complements current IS research on RDM by offering a holistic view of reproducibility challenges in RDM.

Keywords: Research data management, Reproducibility, Open Science.

1 Introduction

Since the last decade, the reproducibility issue of scientific research becomes apparent and calls for attention. As reported by Laine, Goodman, Griswold, and Sox (2007), the amount of errors or misinterpretations of statistical analyses and reproduction failures of peer-reviewed academic work is surging (Donoho, 2010). As a consequence, a number of academic communities starts to promote a better scrutiny of reported results and encourage replication studies in different fields, which include also information systems (Laine et al., 2007; Casadevall and Fang, 2010; Sandve, Nekrutenko, Taylor and Hovig, 2013; Dennis and Valacich, 2014). Although reproducibility is a notoriously ill-defined term in the literature (Plesser, 2018; Schloss, 2018), reproducibility is defined in the Oxford English Dictionary as "the extent to which consistent results are obtained when an experiment is repeated" (OED Online., 2019).

Moreover, reproduction (or replication) issues have also been discussed in information systems (IS) research in conference panels at the International Conference on Information Systems (ICIS) and the European Conference on Information Systems (ECIS) by Brown et al. (2016) and Olbrich et al. (2017). More, Dennis and Valacich (2014) launched the AIS Transactions on Replication Research, where IS scholars submit replication studies. In their replication, manifesto, Dennis, and Valacich (2014) state that replication falls into three categories: exact replications, methodological replications and, conceptual replications. The distinction made by Dennis and Valacich made is a starting point for our study. We observed that what fundamentally distinguishes exact, methodological, and conceptual replicability mentioned in the manifesto are that these categories are variations of who (i.e., same or other authors), what (i.e., theory, tasks, results) , how (i.e., same or different methods) and, where (i.e., same or different environment) studies are repeated. These categories also apply to scientific experimentation when we opt for a holistic view of the actors, tasks, technology, and structures participating in scientific experiments.

Besides, some communities of researchers took the initiative to underline the necessity of leveraging the scholarly communication infrastructure for managing and making research data findable, accessible, interoperable to be reusable (FAIR) by human and machine consumers (Wilkinson et al. 2016). Although the concept of FAIR data reached research data management policies at international and national levels (European Commission, 2016), there is no joint agreement on what FAIR data is, nor what reproducible and reusable data entail.

Thus, our work is guided by the following research question: "What reproducibility threats occurring in experimental systems stem from vulnerabilities in research data management practices?". By answering this question, we seek to contribute to the topics of research data management and reproducible research by (1) characterizing and identifying threats to reproducibility related to challenges encountered in research data management (RDM) (2) articulate reproducibility threats and risks according to a socio-technical perspective on scientific experimentation. By doing so, this paper extends risk management approaches applied previously on digital preservation (Miksa et al., 2014), which is one of the critical tasks of RDM.

The present paper is structured as follows: in the related work section, we make a parallel between experimental systems and socio-technical systems. Next, we introduce research data management activities and concepts. In Section 4, we present the outcomes of a mixed methods (i.e., quantitative and qualitative) approach to acquire evidence from practitioners, institutions, funders in publishers. Finally, the analysis of the evidence led to the development of an STS reproducibility framework introduced in Section 5.

2 Related work

2.1 Socio-technical perspective on experimental systems

Most of the literature dedicated to scientific experimentation belongs to the area of logic, epistemology, and statistics. Studies dedicated to scientific experimentation from a working scientist's perspective are scarcer than the studies on the logic, validity, and methodology of scientific experimentation. Nonetheless, academic work using a socio-technical view on scientific experimentation emerged in the philosophy of science (Rheinberger, 1997; Radder, 2012) and sociology of science (Latour and Woolgar, 1986; Stevens, 2013).

Therefore, we first need to introduce the experimental system perspective of scientific experimentation developed by Radder (2012). Radder's framework depicts scientific experiments as a system consisting of theory, materialization, and results. Hans Radder defines a closed experimental system S as a "complex of object and equipment within a specified spatial area and during a fixed interval of time" (Radder, 2012). Radder defines the instantiation of S as a theoretical description (i.e., formal experimental process and theory), results (i.e., the outcomes of experimental events), and human intervention (i.e., operationalization, the translation of theory to experimental procedures). First, a theoretical description (TD) delineates the episodes (i.e., events and activities) occurring inside S. Radder adds that some episodes have a specific role which is to determine the relative closure of S. In short, S is qualified as being a closed system if non-experimental episodes do not interfere with the episodes and results.

Radder's view on experimental systems echoes more generic socio-technical systems (STS). As explained earlier, experimental systems can be decomposed into the production or manipulation of (IT) artifacts by human intervention. According to socio-technical models (Leavitt, 1965; Ahmad, Lyytinen, and Newman, 2011; Silver and Markus, 2013), variables composing ST systems are structure, tasks, technology, and actors. Experimental systems implicitly refer to similar variables. Such a correspondence enables the analysis of scientific experiments as a socio-technical system. The scholarly communication infrastructure (Wallis, Rolando, and Borgman, 2013) corresponds to the structure variable is STS. Further, the variable tasks correspond to the operationalization of experimental de-

signs, as tasks are defined as being the artifacts and rules used by the actors in an STS (Lyytinen and Newman, 2008).

2.2 Research data management

Academia is facing similar challenges as other sectors such as business and industry to extract valuable knowledge from the increasing amount of data produced worldwide (Borgman, 2012). In experimental science, where sophisticated machines produce large quantities of measures and meta-data about phenomena under investigation in laboratories, the consumption, processing, management, and diffusion of these data are notorious challenges (Borgman, 2012; Baesens et al., 2016). Nevertheless, besides researchers, research data management governance has responsibilities distributed among multiple stakeholders.

Further, external laboratories, stakeholders such as academic funders, are progressively governing the production, preservation, diffusion of scientific data created by (publicly) funded research (OECD, 2007). As a result, researchers are facing new regulations, procedures, and technological challenges for managing data at each step of scientific experimentation.

First, public research funders posit some prerequisites for managing research data generated with public resources. For instance, grant applicants need to describe the (future) data sets, storage systems, and anonymization techniques, among other items, in data management plans (DMPs). Funders pursue societal ambitions of opening data and disseminating scientific knowledge. This ambition can be seen from the evolving National and European regulation, which encourages the dissemination of scientific artifacts in novel ways (European Commission, 2015).

Second, publishers are critical stakeholders of the scholarly communication infrastructure, which is an essential structure of communication in science. In recent years, scholars have investigated the challenges of current scholarship practices to deal with data sharing and reproducible research (Borgman, 2008; Reilly, Schallier, and Schrimpf, 2011). Research data management seeks to transform the scholarly communication infrastructure to push academic data sharing and preservation forward. RDM is a collective enterprise, for which efforts are shared between research funders, academic publishers, research institutions, and researchers to achieve reusability and reproducibility of scientific output (Lefebvre, Schermerhorn and Spruit, 2018).

More, academic publishers explicitly integrate research artifacts produced by researchers in their editorial processes. In recent years, publishers introduced data sharing, preservation, and dissemination guidelines and policies (Editorial, 2014). These policies are aimed at grant applicants who have to present data management strategies early in the application process. Experimental artifacts such as datasets, materials, and software must be precisely documented and, possibly, disseminated according to the publisher's and journals' guidelines.

Finally, research institutions reorganize their IT services to support researchers in managing research data at their host institutions. Research institutions deploy institutional repositories and technology for managing research data to secure funding opportunities. This fact led to new managerial and support roles in academia appearing in academia, such as data stewards and research data managers.

3 Mixed methods approach

We follow a mixed-methods approach (Bergman, 2008); thus, we apply quantitative and qualitative data collection techniques. We conducted semi-structured interviews to gather evidence from laboratory workers and acquired open data to analyze the scholarly communication infrastructure. We divided the data collection and analysis into two periods. During the first period, we gathered information about the management of scientific data by interviewing seven researchers in the bioinformatics community of one University in the Netherlands. There, we collected experiences of researchers in laboratories about data management practices. Throughout the interviews, we identified several challenges

related to the preservation, interpretation, and dissemination of scientific artifacts. To increase the contextualization of our interviews, we first obtained a dataset from the administration of the University and analyzed an anonymized version of the data. This survey was submitted to the academic staff of our university in August 2014. For this survey, 829 researchers out of 3197academic staff members (source: annual report of the institution) answered, which is a response rate of 26%. After removing incomplete cases, 489 records were retained for further analysis. As we focus our analysis on reproducibility in experimental systems, we filtered the respondents on faculties that are using scientific experimentation. Removing non-experimental disciplines further narrowed the sample to 289 respondents.

Next, after the interviews, we screened 323 publications in the domain of Biological Science (i.e., BIOC category on Scopus) as the focus of our study is on experimental work, 252 full-text publications were retained for further analysis (78%). The reason for removing 22% of the articles is that these articles did not report on experimental work (e.g., literature review) or did not produce research data with laboratory work (e.g., computer simulation using open data). Table 1 presents the characteristics of the sample of selected publications.

Finally, we sampled author guidelines and policies from seven publishers: Elsevier (ELS), Public Library of Science (PLOS), Cell Press (CP), American Chemical Society (ACS), Nature Publishing Group (NGP), Oxford University Press (OUP), eLife Sciences Publications (ELIFE). The guidelines can vary extensively from publisher to publisher. The rationale is that one single publisher might host several dozens of journals and decide that research data management matters are to be elaborated by each journal. Also, to cover funding agencies, we added documents from public funders: National Institute of Health (NIH), National Science Foundation (NSF), European Commission (EC), and The Netherlands Organization for Scientific Research (NWO, in Dutch).

Publisher	**Average SJR**	**Average Number of Authors**	**Average number of files in SI**	**Average Distinct Formats in SI**	**Percentage of Availability Statements**	**N**
Cell Press	10.7	24.3	3.5	1.61	69%	42
Elsevier	2.3	10.3	1	0.9	5%	20
Nature Publishing Group	14.2	31.0	4.4	1.88	94%	17
Other	3.3	14.2	2.5	1.26	25%	119
Public Library of Science	1.5	10.2	3.6	1.28	100%	35
eLife Sciences Publications	7.1	11.1	2.31	0.89	31%	19

Table 1 Characteristics of the scientific publications screened in this study. SI means Supplemental Information, which are files hosted on the publisher's website. Availability statements are paragraphs where authors describe how to retrieve the underlying data.

4 Results

4.1 People, technology and tasks

The first part of the results section presents the outcomes of the survey and interviews. Next, in Section 4.2, the results of the screening of publications are shown.

4.1.1 Survey

As can be seen from Table 2, RDM practices, as reported by researchers to the IT services of their host institution, reveal that reaching the ambitions set by RDM stakeholders is an ongoing effort. Overall, it

appears from the survey results that researchers seem reluctant to comply with RDM tasks set by funders, publishers, and research institutions. Also, there seems to be only a minority of respondents who would agree to depend on central IT services of their institutions, except data preservation. Data management planning (around 20%), assistance with lab notebook systems (and assistance with dissemination seem to rank less high than preservation, i.e., assistance and technology to back up research data in the long term.

Task	Statement	Response	N
Report reproducible results	"You are interested in digital Laboratory notebook systems."	Yes - 60 (20.7%)	289
	"You need to record who accesses and modifies datasets."	No – 150 (46%)	250
Conduct data management planning	"You are interested in expertise in writing data management plans."	Yes – 69 (23.9%)	289
	"You created a data management plan at the start of the project."	Yes – 30 (10.4%)	289
Elaborate preservation strategy	"You want assistance in organizing long term preservation of data."	Yes – 113 (39%)	289
	"You are interested in long-term backup facilities."	Yes – 139 (48%)	289
Elaborate dissemination strategy	"You plan to make data publicly available"	Yes – 79 (27.3%)	289
	"You are interested in expertise for publishing data in a public repository."	Yes – 70 (24.2%)	289

Table 2 We retained 289 responses from researchers in the faculties of science, geosciences, and veterinary medicine as these faculties mostly relied upon experimental systems and reported to work with experimental data.

4.1.2 Interviews

This section summarizes information about data management in distinct research laboratories. Seven researchers in the fields of biology and bioinformatics were interviewed. All interviews show different data preservation and dissemination practices as well as technology in place in laboratories. All interviewees are labeled by their laboratory, followed by their position: Principal Investigator (PI) or Postdoc (PD). We purposely discussed with interviewees who had proven experience in their respective domains, see Table 3 for an overview of the interviewees.

For analyzing data in **Computational Structural Biology, CSB/PI** has an advanced computation infrastructure (grid computing) and maintain self-developed analysis software utilized internationally. CSB/PI say that assessing the quality of the data needs specific expertise. The files generated have different structures that are specific to the application that generated them. There is also no permanent storage of intermediate processing products. Data sharing is sometimes not done by transferring data but by giving access to where the data is located as its size would be too resource consuming. CSB/PI recommends the use of meta-data to validate the format of the files, but meta-data is more challenging for evaluating the quality of the data itself.

In **Biomedical Genetics**, BG/PI explains that scientific data is reused, but analysis workflows are not as they should be better described. There are intrinsic quality measures in the sequencing files that are used to assess the quality of the sequence reads. BG/PD reuses datasets from different publications and merges them to answer his/her research questions. A lot of this storage is done on a shared network disk and processed on a local desktop. BG/PD says that there are no standards and no description of the data that s/he downloads, which imply to guess the meaning. BG/PD says that for this reason, it is needed to contact the authors who will generally provide the requested information. There is a high turnover of undergraduates and graduates, which means that there is sometimes no follow-up of pro-

jects. BG/PD suggests that information should be provided about available code or workflows, indicating that the material works and has been verified by peers (e.g., a stamp).

In a laboratory of **Stem Cell Biology**, the interviewee SCB/PD is a bioinformatician. Although the bioinformatics unit is shared between different groups, there is no appropriate structure or organization of the scientific data. Sequencing data processed by this unit is generated externally. According to SCB/PD, a central repository should be developed to structure this sequence data and identify its location from the start of the data generation process. The available meta-data annotation is considered as weak. Also, bioinformaticians in this group are perceived as being hesitant to make their code available as its quality might be judged as not being up to standards by peers.

Domain	Role	Identifier	Reproducibility challenges
Computational Structural Biology	PI	CSB/PI	Expertise required to evaluate data quality
Biomedical Genetics	PI, Post-doc	BG/PI, BG/PD	Absence of standard descriptions of remote data
Stem Cell Biology	Post-doc	SCB/PD	No shared infrastructure between laboratories, Weak meta-data annotations
Pediatric Oncology	PI	PO/PI	Absence of data preservation strategy
Medical Microbiology	Post-doc	MM/PD	Moving data between (legacy) systems. Conservative attitude towards data sharing
Metagenomics	PI	MG/PI	Data is depending upon a range of (online) databases which might not be adequately documented

Table 3 Overview of the interviewees' domain, role and a summary of reproducibility challenges

In the Department of **Pediatric Oncology**, the respondent PO/PI took the lead of a newly created research group, which leaves any data management issues open at the time of the interview. A custom-made laboratory information system (LIMS) manages micro-array data, and they seek to develop the same system for sequencing data. The identification of what data can be stored or dismissed is still an open question for which our interviewee believes that better and automated meta-data collection is critical. Regarding data reuse, one scenario is to re-purpose data that was used initially as quality controls.

The role of the interviewee in **Medical Microbiology** (MM), MM/PD, is to establish a bioinformatics pipeline and a private repository to make these datasets findable. Keeping the data consistent is an issue, as illustrated by a legacy issue that occurred when laboratory members in MM moved old files without any identifier assigned to the new repository. According to MM/PD, MM has a conservative attitude regarding data sharing that might evolve with the younger generation.

In **Metagenomics**, the constituents of a biological sample are unknown, and the goal of the analysis is to identify from which organisms the sequenced genomes are originating. A single sample might, therefore, be processed by calling different genomic reference databases to annotate this material. Still, MG/PI found that the available raw data has poor meta-data description, which evaluates data quality and further processing laborious. MG/PI explained that data sharing is widespread in his field and that data reuse is common for answering new research questions, but not for verification purposes. Intermediate processing products (e.g., files) are not preserved.

4.2 Structure: The scholarly communication infrastructure

The scholarly communication infrastructure has, besides researchers, stakeholders governing and managing the communication of scholarly work. Funders posit requirements to researchers before and after a project, mostly via data management planning. Publishers act as the central governing bodies of

science communication. We first introduce tasks as they ought to be conducted (i.e., policy view) and the screening of publications, in Section 4.2.2, to show how things are done (i.e., "real world" view).

4.2.1 Funders and Publishers

The documents we consulted from funders and publishers listed in Section 3 resulted in a classification of several RDM tasks that researchers are expected to complete for getting funding granted on the one side and publishing in journals on the other side. The main tasks which are reported by funders and publishers are shown in Table 4. We operated a division between the two main objectives of RDM: facilitating efficient preservation and dissemination. Besides, remaining RDM tasks support data management planning tasks, such as sending data management plans to funders and, reporting reproducible results.

Task	Origin of Policy	Exemplary Quotes from Policies
Report reproducible results	Publisher	"Authors of research articles in the life sciences, behavioral & social sciences and ecology, evolution & environmental sciences are required to provide details about elements of experimental and analytical design that are frequently poorly reported in a reporting summary" – Nature publishing group "Data, methods used in the analysis, and materials used to conduct the research must be clearly and precisely documented and be maximally available to any researcher for purposes of reproducing the results or replicating the procedure." - eLife
Conduct data management planning	Funder	"A data management plan that must be submitted after the proposal has been awarded funding. The approval of this plan is a prerequisite for NWO disbursing the grant." – Netherlands Organisation for Scientific Research (NWO)
Elaborate preservation strategy	Funder	"Which facilities (ICT, (secure) archive, refrigerators, or legal expertise) do you expect will be needed for the storage of data during the research and after the research? Are these available?" – Netherlands Organisation for Scientific Research (NWO)
Elaborate dissemination strategy	Publisher, Funder	"Before manuscript submission, the Authors must deposit the underlying data to an appropriate public repository for public release scheduled no later than the publication date of the article." – Oxford University Press (OUP) "Such applicants are expected to contact IC program staff prior to submission and are also expected to include a data-sharing plan in their application stating how they will share the data or, if they cannot share the data, why not" – National Institutes of Health (NIH)

Table 4 Activities extracted from policies of funding agencies and academic publishers

4.2.2 Publications

In addition to publishers and funders' policy, we seek to collect evidence about data sharing practices from screening scientific publications. In other words, we dive specifically into the tasks of reporting reproducible results and observe the consequences of dissemination strategies deployed by publishers.

In Figure 2, the results of the screening are shown. The analysis illustrates shortcomings in dissemination strategies in terms of the use of digital repositories (Fig. 1.A), modes of availability (B), available file formats (C) and, type of organizations maintaining repositories.

In Fig. 1.A. that most of the screened articles do not refer to any deposited material. The low percentage of deposits is surprising considering that all sampled articles report on experimental work,

thus with data acquired or produced. Moreover, Figure 1A shows that few publishers can invert this trend. Cell Press and Nature publishing groups host more prestigious outlets with a more extended history of attempts to improve reporting and data availability, which might explain why these publishers are more successful at convincing authors to deposit data.

In Fig. 1.B, supplemental information is preferred as an alternative to repository deposits. Supplemental information (SI) files are hosted on the publisher's servers. A limitation of this mode of availability is shown in 1.C, where most of the information available in SI is not in the original formats. Documents (i.e., PDFs, word documents) and spreadsheets (i.e., excel workbooks) are popular file formats. Original file formats that might prove useful for reproduction purposes, such as computer code, are seldom made available.

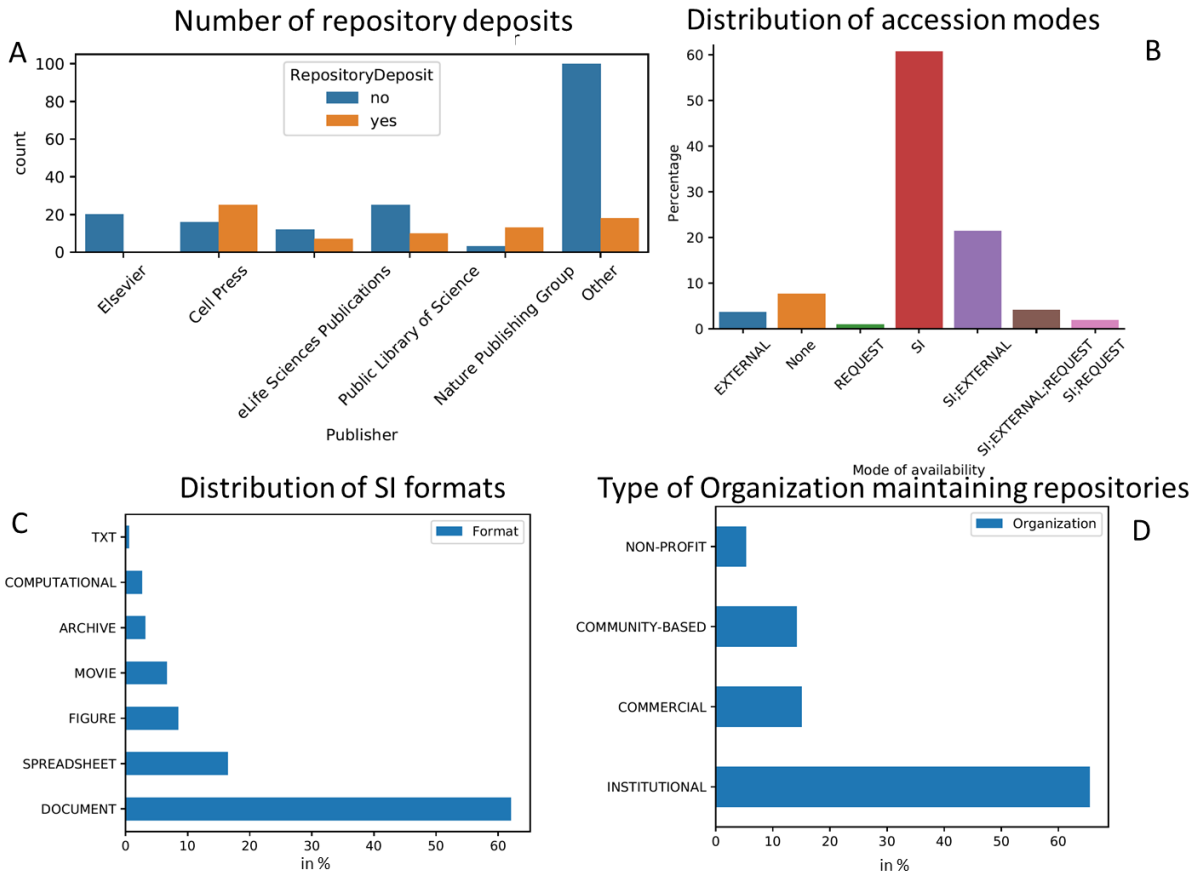

Figure 1 Results of the screening of scientific articles published in 2017 in the category Biochemistry, Genetics, and Molecular Biology (BIOC). Data source exported from Scopus.

Last, Fig. 1.D shows that authors privilege repositories that are established in their communities and hosted by renowned organizations such as EMBL. This choice might also be guided by the RDM policies of publishers, which mandate authors to deposit material in these types of repositories.

5 A Socio-Technical Framework of Reproducibility Threats

In this section, we introduce relevant concepts to decipher the implications on the reproducibility of RDM practices reported in Section 4. The threats are summarized in Section 5.2.

5.1 Dimensions of reproducibility

A shared understanding of the concept of reproducibility in scientific research is that reproducibility refers to the capability of re-enacting previous studies. When experimental systems acquire results, the complexity of re-enacting the objects, procedures, digital analysis, and theoretical descriptions calls for a division in terms of "what is reproduced?" Radder (1992), divides types of reproducibility in terms of who is reproducing and what is reproduced. Here, we opt for a slightly different division to consider the many levels at which reproducibility issues might occur in experimental science where laboratory and computer work are combined. In the end, five dimensions are retained:

First, **Phenomenal and technical reproducibility** apply to local laboratory work (Tabb et al., 2010). There are fundamental experimental techniques which ensure that results obtained from the instruments are accurate and biologically sound. Therefore, biological and technical reproducibility involves several defense mechanisms such as producing data in du/triplicates and comparing measurements on an object of study to positive and negative controls. BTR ensures that the experimental conditions at one location are well set. For instance, those instruments are calibrated and that observations did not occur by chance.

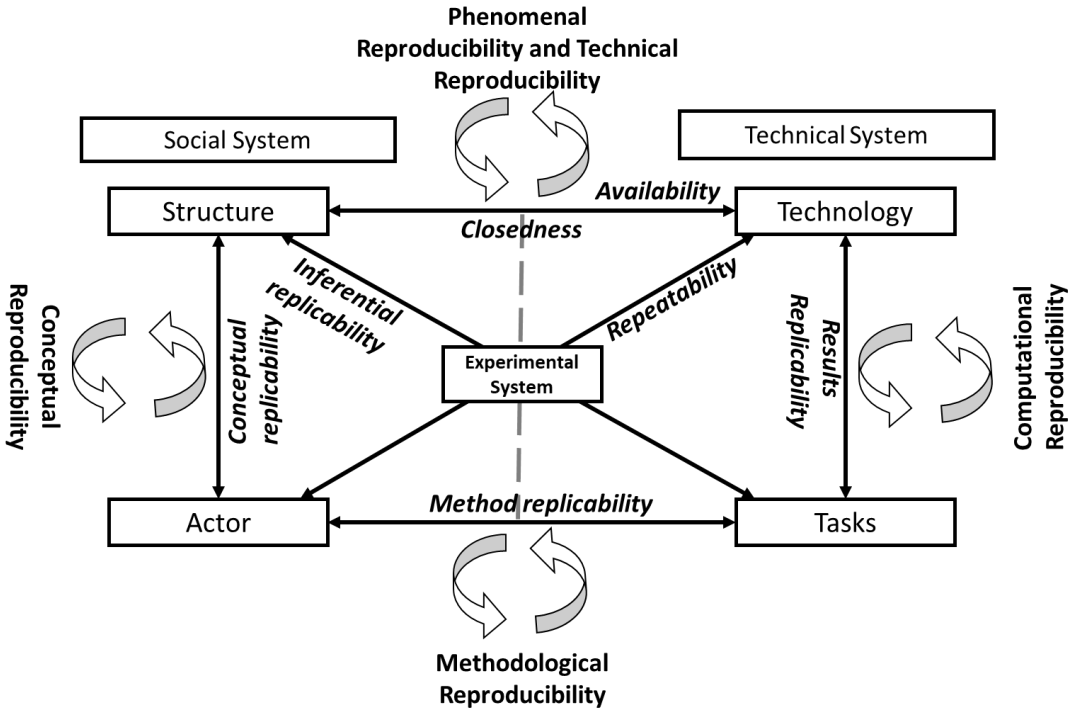

Figure 2 Results of the screening of scientific articles published in 2017 in the category Biochemistry, Genetics, and Molecular Biology (BIOC). Data source exported from Scopus.

Next, **Computational reproducibility** is a concern when computer software and hardware are used to generate, process, and analyze scientific data (Peng, 2011; Freire, Bonnet, and Shasha, 2012). As we exposed earlier, computing technology is pervasive in modern scientific experimentation. CR is quite diverse in scope. Among many requirements for achieving CR, we can name a few here: the availability of data and code (Peng, 2011), versioning and logging (Sandve et al., 2013), and the use of literate programming (Knuth, 1984). In short, biological, technical, and computational reproducibility guarantee the robustness of experimental operationalizations. They do not contribute to evaluating if the same results hold under different (experimental) conditions at different locations; this is the role of methodological and conceptual reproducibility (see below). Although CR is crucial for verification

and reuse of computational work, it is not enough for evaluating the stability of results as CR focuses on making computations repeatable and reusable.

Finally, **Methodological and conceptual reproducibility** are forms of reproducibility that apply (also) outside the boundaries of a laboratory. At that level, reproducibility is an integral part of the scientific method (Andersen and Hepburn, 2016). Methodological reproducibility (MR) aims at testing the rigor of experimental designs or the stability of experimental outcomes at different points in time and space. Some authors make a clear distinction between methodological and conceptual reproducibility by stating that one can assess the stability of results by applying identical methods on new data or test a similar theoretical framework using new methods (Niederman and March, 2015). Thus, these two types of reproducibility are employed to challenge published theories and results. This is differing from the first three types, which can only say something about the rigor and robustness of data analysis pipelines and laboratory procedures.

5.2 Threats to reproducibility

Previous work by Schloss, (2018) and Goodman et al., (2016) show that reproducibility corresponds to a diversity of threats that occur when an independent team wants to reproduce results. Risks are too often that material is not available nor preserved in optimal conditions. To increase the understanding of the relations between reproducibility and its related threats, risks, and RDM vulnerabilities, we comment hereunder on the relations depicted in Figure 2.

First, we can represent **inferential replicability** as a **Structure ↔ Task** relation. Inferential replicability belongs to the conceptual reproducibility dimension as it relies on the scholarly communication infrastructure on the one hand and the capability to derive the (experimental) tasks that are conducted by the experiments to reach similar conclusions. **Conceptual Replicability** differs from inferential replicability; it is represented as the **Structure ↔ Actor** since it depends on the capacity of the same or alternative (team of) experimenters to reach similar conclusions based on the description of experimental work. Last, the extent to which artifacts are made available by the authors of a study, which we labeled **availability**, is represented the relation **Structure ↔ Technology.** In short, these relations all rely upon the availability of artifacts and detailed reporting of results.

Second, the dimension of computational reproducibility involves the results of replicability and repeatability. The difference between these terms can be explained by referring to the relations depicted in Figure 2. **Results Replicability** is ensured when **Technology ↔ Task** yields consistent results at each run, independently of the fact that the original or another (team of) experimenter(s) use these digital artifacts. Results replicability is facilitated by the publication of re-usable data and software in digital repositories. There is a subtle distinction with **Repeatability**, **Technology ↔ Actor**, as guaranteeing that technology yields repeatable results is the responsibility of the original team of experimenters.

Third, **Method Replicability** is a **Task ↔ Actor** relation, which means that actors (i.e., experimenters) can replicate studies by following the same procedures (or tasks). Methodological reproducibility is only possible in case the experimental system is closed. So, the **closedness** of the system implies that the two sides of the experimental system (social and technical) are consistently communicated and operated (hence closedness is depicted as the bridge between social and technical systems).

Finally, a summary of the threats to reproducibility and their associated risks is shown in Table 5. The associated risks and vulnerabilities are compiled from the interviews and the screening of publications. Also, the terminology retained for classifying reproducibility threats, risks, and vulnerabilities are derived from the ISO standard ISO/IEC 27000:2016, which provides a shared vocabulary for information security. Risk is defined as the "effect of uncertainty on objectives" (2.68). A vulnerability is a "weakness of an asset or control that can be exploited by one or more threats" (2.89). Finally, a threat is the "potential cause of an unwanted incident, which may result in harm to a system or organization" (2.57).

Label	Threat	Risk	RDM Vulnerability	Example
Inferential replicability	External experimenters want to use findings from a published study to reach similar conclusions	No relation between reported findings and underlying artifacts which is a risk for Inferential replicability	No (or weak) reproducible reporting	Researchers in biomedical genetics (see interviews) who do not reuse workflows due to poor documentation. (Source: Interviews.)
Method replicability	External experimenters want to produce or acquire new evidence based on published resources and procedures	Underlying artifacts are not disseminated in their original formats which is a risk for method replicability	No (or weak) dissemination strategy	Researchers are reluctant to share code due to their quality perceived as inadequate. (Source: Interviews.)
Result replicability	An independent team wants to conduct an exact or partial evaluation of initial results as reported by the authors using similar (or identical) artifacts (e.g., software).	Custom code, workflows are not available in the laboratory and outside the laboratory which poses a risk for result replicability	No (or weak) dissemination strategy and preservation strategy	Very few computational artifacts are attached to publications. (Source: Screening.)
Closedness	The team of experimenters wants to isolate experimental results from interferences between experimental events and external (uncontrolled) events.	Software versions and computational workflows not preserved which is a risk for closedness as different software versions might give differing results	No (or weak) preservation strategy. No (or weak) reproducible reporting	New laboratories have no systems in place yet to trace experimental processes. (Source: Interviews.)
Repeatability	The team of experimenters wants to obtain similar results by applying the same routines and procedures (i.e., operationalization)	Poor management of software, data and lab notebooks are a risk for repeatability	No (or weak) preservation strategy	Significant turn-over and no follow up on projects. (Source: Interviews.)
Conceptual replicability	External experimenters produce or acquire new evidence to evaluate existing theories.	Experimental conditions not sufficiently described is a risk for conceptual replicability	No (or weak) reproducible reporting	A majority of artifacts are not deposited on curated repositories. (Source: Screening.)
Availability	Make the evidence underlying a preliminary report (i.e., a scientific article) available to readers for further verification and reuse.	Poor planning, versioning of artifacts, sharing habits, unforeseen privacy issues are a risk against the availability of artifacts	No (or weak) research data planning	Few projects consistently plan data management. (Source: Survey.)

Table 5 Overview of threats to reproducibility

6 Discussion and Limitations

We presented an approach to reproducibility, articulated in dimensions and threats, which help to categorize the different challenges of reproducibility encountered in experimental sciences. To the best of our knowledge, no such conceptualization of reproducibility and data management has been previously suggested in the literature.

We have introduced an initial framework to answer the question: "What are reproducibility threats occurring in experimental systems stem from vulnerabilities in research data management?". We have seen that preservation, dissemination, planning and reporting practices, standards in experimental science vary per domain (see interviews), and publishers (see publication screening). We worked towards a framework to capture these elements and position them according to reproducibility risks, threats, and RDM vulnerabilities. By doing so, we depict reproducibility threats and RDM vulnerabilities by considering (1) researchers and other stakeholders (2) introduce challenges experienced by the researchers (3) seek to grasp how these challenges translate into the scholarly communication infrastructure. To achieve that, we bridged a gap between experimental systems and socio-technical systems as successful reproduction of experimental work rely upon factors beyond technology.

Besides, reproducibility mechanisms, as depicted in the framework (Figure 2) goes beyond scientific experimentation in natural sciences. For instance, in IS research, design science research (DSR) is confronted with similar issues regarding transparent reporting and dissemination of reusable artifacts (Gleasure, Feller and Flaherty, 2012; Iivari, Rotvit Perlt Hansen and Haj-Bolouri, 2018). While some authors cast doubts on the applicability of terms such as reproducibility on a DSR paradigm (Baskerville and Pries-heje, 2016), the challenges experienced during artifact design and experimentation are similar from the perspective of working scientists. A dynamic view on the production of artifacts requires another type of sources than interviews and reports. A suggestion is, therefore, to include laboratory forensics (LF) findings into the current reproducible framework. According to Lefebvre and Spruit (2019), LF adds a perspective from practice by investigating digital files on storage systems and offer insights on RDM vulnerabilities going beyond what can be obtained from interviews and the study of publications alone. Another suggestion is to pursue the evaluation of experimental artifact reusability similarly to evaluation criteria for the reuse of design principles (Iivari et al., 2018), where experimental artifacts and their descriptions in method sections are not only evaluated by their accessibility but with a broader range of criteria such as appropriate guidance and effectiveness of disseminated experimental artifacts.

There are several limitations to our study design and findings, which we elaborate on here. The first limitation is that we conceptualize reproducible for experimental sciences with data covering only a limited sample of experimental scientists in biomedical science. However, we attempted to mediate this narrow view on experimental science in our study by adopting a more general view on scientific experimentation with experimental system theory. Similarly, to the limitations of our interview data, other disciplines might reflect other data sharing and usage patterns that the patterns we found in biomedical disciplines (Gregory et al., 2018).

7 Conclusion

A socio-technical approach on experimental systems highlights the dynamics of scientific experimentation from the point of view of working scientists (i.e., the actors) operationalizing experimental design (i.e., the tasks) using laboratory instruments and computers (i.e., technology) to communicate novel findings on the scholarly communication infrastructure (i.e., structure). We believe that understanding RDM practices and reproducibility challenges depends on the capability to frame experimental work in all its dimensions.

However, from the survey, interviews, and screening of publications, we saw that dissemination and preservation strategies are challenging to implement. RDM deals with the fragmentation of policies, ad-hoc data governance in laboratories, and few constraints put on systematic and structured sharing of computational resources in publications. Our results show that reproducibility risks need to be better understood to redesign the research data management and scholarly communication infrastructures effectively.

References

Ahmad, A., K. Lyytinen and M. Newman. (2011). "The evolution of process models in is research: from a punctuated social process model to a socio-technical process model." In: *AIS Electronic Library (AISeL)*.

Andersen, H. and B. Hepburn. (2016). "Scientific Method." In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 201). Metaphysics Research Lab, Stanford University.

Baesens, B., R. Bapna, J. R. Marsden, J. Vanthienen and J. L. Zhao. (2016). "Transformational Issues of Big Data and Analytics in Networked Business." *MIS Quarterly*, *40*(4), 807–818.

Baskerville, R. and J. Pries-heje. (2016). "Design Theory Projectability," 446219–232978.

Bergman, M. (2008). *Advances in Mixed Methods Research*.

Borgman, C. L. (2008). "Data, disciplines, and scholarly publishing." In: *Learned Publishing* (Vol. 21, pp. 29–38). John Wiley & Sons, Ltd.

Borgman, C. L. (2012). "The conundrum of sharing research data." *Journal of the American Society for Information Science and Technology*.

Brown, S., A. Dennis, B. Samuel, B. Tan, J. Valacich and E. Whitley. (2016). "Replication research: Opportunities, experiences and challenges." *ICIS 2016 Proceedings*.

Casadevall, A. and F. C. Fang. (2010). "Reproducible science." *Infection and Immunity*, *78*(12), 4972–5.

Dennis, A. R. and J. S. Valacich. (2014). "A Replication Manifesto." *AIS Transactions on Replication Research*, *1*(1), 1–5.

Donoho, D. L. (2010). "An invitation to reproducible computational research." *Biostatistics (Oxford, England)*, *11*(3), 385–8.

Editorial. (2014). "Journals unite for reproducibility." *Nature*, *515*, 7.

European Commission. (2015). *Access to and preservation of scientific information in Europe*.

European Commission. (2016). "Guidelines on Data Management in Horizon 2020." Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Freire, J., P. Bonnet and D. Shasha. (2012). "Computational reproducibility: state-of-the-art, challenges, and database research opportunities." *Proceedings of the 2012 ACM SIGMOD ...*, 593–596.

Gleasure, R., J. Feller and B. O. Flaherty. (2012). "PROCEDURALLY TRANSPARENT DESIGN SCIENCE RESEARCH: A DESIGN PROCESS MODEL." In: *ICIS* (pp. 1–19).

Goodman, S. N., D. Fanelli and J. P. A. Ioannidis. (2016). "What does research reproducibility mean?" *Science Translational Medicine*, *8*(341), 341ps12-341ps12.

Gregory, K., H. Cousijn, P. Groth, A. Scharnhorst and S. Wyatt. (2018). "Understanding Data Search as a Socio-technical Practice." *Journal of Information Science*, 016555151983718.

Iivari, J., M. Rotvit Perlt Hansen and A. Haj-Bolouri. (2018). "A Framework for Light Reusability Evaluation of Design Principles in Design Science Research." In: *DESRIST*.

Knuth, D. E. (1984). "Literate Programming." *The Computer Journal*, *27*(2), 97–111.

Laine, C., S. N. Goodman, M. E. Griswold and H. C. Sox. (2007). "Reproducible Research: Moving toward Research the Public Can Really TrustMoving toward Research the Public Can Trust." *Annals of Internal Medicine*, *146*(6), 450–453.

Latour, B. and S. Woolgar. (1986). *Laboratory life the construction of scientific facts*. Princeton, NJ : Princeton University Press.

Leavitt, H. J. (1965). "Applied organisational change in industry: Structural, technological and humanistic approaches." In: *Handbook of organizations.*

Lefebvre, A., E. Schermerhorn and M. Spruit. (2018). "How Research Data Management Can Contribute to Efficient and Reliable Science." In: *The 25th European Conference of Information Systems*. Portsmouth.

Lefebvre, A. and M. Spruit. (2019). "Desiging laboratory forensics." In: I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccher, J. Krogstie, & M. Mäntymäki (Eds.), *18th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2019*. Trondheim, Norway: Springer.

Lyytinen, K. and M. Newman. (2008). "Explaining information systems change: A punctuated socio-technical change model." *European Journal of Information Systems*.

Miksa, T., R. Mayer, S. Strodl, A. Rauber, R. Vieira and G. Antunes. (2014). "Risk Driven Selection of Preservation Activities for Increasing Sustainability of Open Source Systems and Workflows." *IPres 2014*, 91–100.

Niederman, F. and S. March. (2015). "Reflections on Replications." *AIS Transactions on Replication Research*, *1*(1), 1–16.

OECD. (2007). "Principles and Guidelines for Access to Research Data from Public Funding."

OED Online. (2019). ""reproducibility, n."." Oxford University Press.

Olbrich, S., U. Frank, S. Gregor, F. Niederman and F. Rowe. (2017). "On the Merits and Limits of Replication and Negation for IS Research." *AIS Transactions on Replication Research*, *3*(1), 1–19.

Peng, R. D. (2011). "Reproducible research in computational science." *Science*, *334*(6060), 1226–7.

Plesser, H. E. (2018). "Reproducibility vs. Replicability: A Brief History of a Confused Terminology." *Frontiers in Neuroinformatics*, *11*, 76.

Radder, H. (1992). "Experimental Reproducibility and the Experimenters' Regress." In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 1992, pp. 63–73).

Radder, H. (2012). *The Material Realization of Science* (Vol. 294). Dordrecht: Springer Netherlands.

Reilly, S., W. Schallier and S. Schrimpf. (2011). "Report on Integration of Data and Publications." *Report on Integration of Data and Publications*, *5*, 1–87.

Rheinberger, H.-J. (1997). "Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube." Stanford University Press.

Sandve, G. K., A. Nekrutenko, J. Taylor and E. Hovig. (2013). "Ten simple rules for reproducible computational research." *PLoS Computational Biology*, *9*(10), e1003285.

Schloss, P. D. (2018). "Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research." *MBio*, *9*(3), e00525-18.

Silver, M. S. and M. L. Markus. (2013). *Conceptualizing the SocioTechnical (ST) Artifact*. *Signs & Actions An International Journal on Information Technology, Action, Communication and Workpractices* (Vol. 7). Retrieved from http://www.sysiac.org/

Stevens, H. (2013). *Life out of sequence: a data-driven history of bioinformatics*. Chicago, USA: Univeristy of Chicago Press.

Tabb, D. L., L. Vega-Montoto, P. A. Rudnick, A. M. Variyath, A. J. L. Ham, D. M. Bunk, … C. Spiegelman. (2010). "Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry." *Journal of Proteome Research*, *9*(2), 761–776.

Wallis, J. C., E. Rolando and C. L. Borgman. (2013). "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology." *PLoS ONE*, *8*(7), e67332.