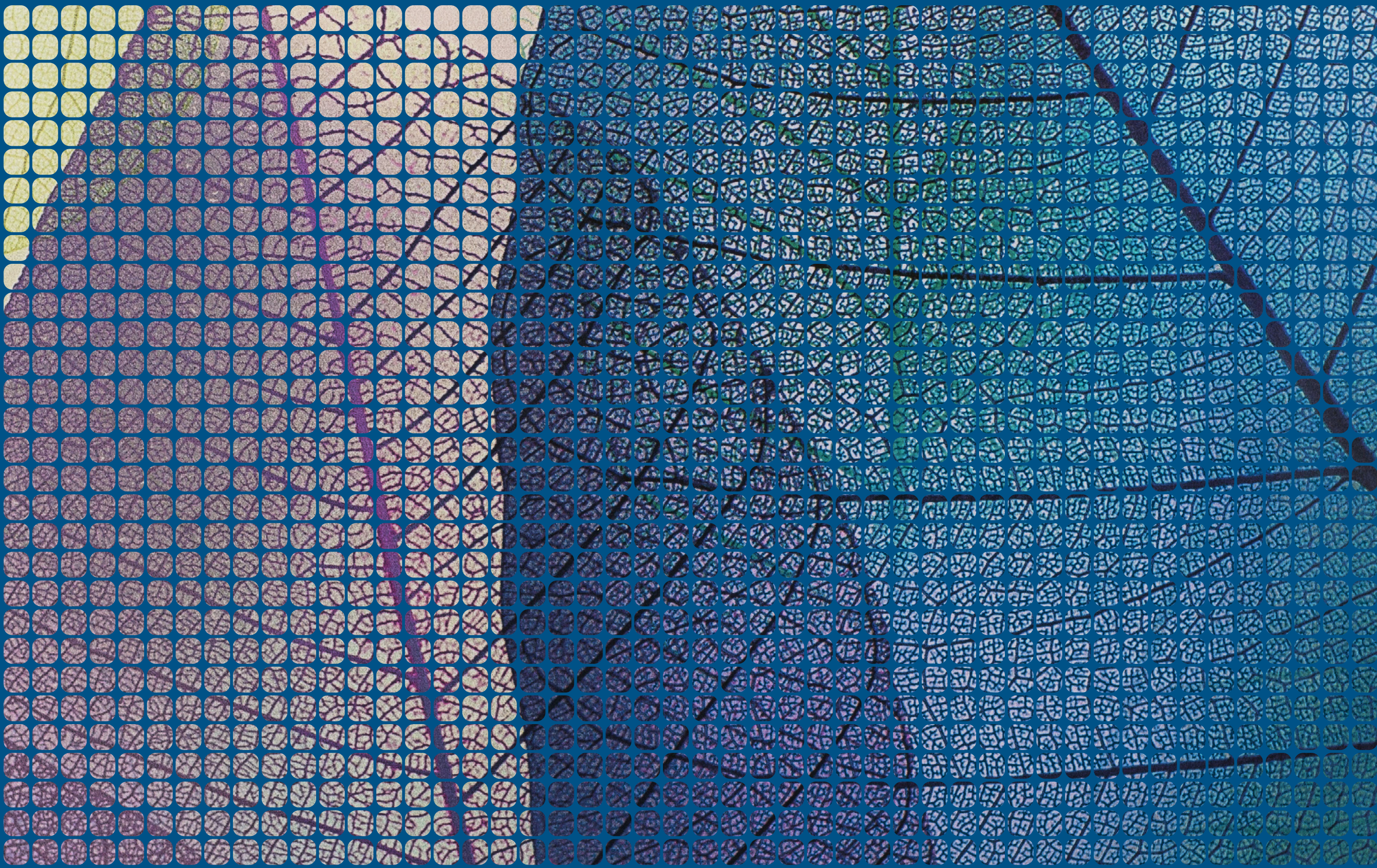


Digital Science Report

Making Science Better: Reproducibility, Falsifiability and the Scientific Method

The increasing importance of failure in supporting modern research

Leslie D. McIntosh, Cynthia Hudson Vitale, Anthony Juehne,
Leah Haynes, Sasha Mothershead and Josh Sumner

SEPTEMBER 2019

DIGITAL science

ripeta

About Digital Science

Digital Science is a technology company working to make research more efficient. We invest in, nurture and support innovative businesses and technologies that make all parts of the research process more open and effective. Our portfolio includes the admired brands Altmetric, Anywhere Access, CC Technology, Dimensions, Figshare, Gigantum, GRID, IFI Claims, Overleaf, ReadCube, Ripeta, Symplectic and Writefull. We believe that together, we can help researchers make a difference. Visit www.digital-science.com

About Ripeta

Ripeta's mission is to assess, design, and disseminate practices and measures to improve the reproducibility of science with minimal burden on scientists, starting with the biomedical sciences. We focus on assessing the quality of the reporting and robustness of the scientific method rather than the quality of the science. Our long-term goal includes developing a suite of tools across the broader spectrum of sciences to understand and measure the key standards and limitations for scientific reproducibility across the research lifecycle and enable an automated approach to their assessment and dissemination. Visit www.ripeta.com

About the authors

Dr. Leslie McIntosh, PhD, MPH is the founder and CEO of Ripeta, and an internationally known consultant, speaker, and researcher passionate about mentoring the next generation of data scientists. She is active in the Research Data Alliance, grew the St. Louis Machine Learning and Data Science Meetup to over 1500 participants, and was a fellow with a San Francisco based VC firm. She recently concluded as the Director of Center for Biomedical Informatics (CBMI) at Washington University in St. Louis. Dr. McIntosh has a focus of assessing and improving the full research cycle (data collection, management, sharing, discoverability) and making the research process reproducible. She holds a Masters and PhD in Public Health with concentrations in Biostatistics and Epidemiology from Saint Louis University.
https://orcid.org/0000-0002-3507-7468

Anthony Juehne, MPH is the Chief Science Officer at Ripeta with specialty skills in epidemiology and biostatistics. He has over five years of experience within the field of healthcare informatics supporting Principal Investigators to define population criteria and research design to capture prospective and retrospective clinical data accurately and efficiently. Within the Center for Biomedical Informatics at Washington University, Anthony managed multiple research and quality-control projects simultaneously. His current work focuses on developing best-practices for conducting and reporting clinical research to enhance reproducibility, transparency, and accessibility.
https://orcid.org/0000-0002-4311-4097

Cynthia R. Hudson Vitale, MA is the Chief Information Scientist with Ripeta. She has over six years of experience providing research data services support to scientific principal investigators in a variety of domains, specifically in the areas of data management and data curation. She has worked with faculty on projects to facilitate data sharing and interoperability while meeting faculty research data needs throughout the research life-cycle. Her current research seeks to improve research reproducibility, addressing both technical and cultural barriers.
https://orcid.org/0000-0001-5581-5678

Sasha Mothershead is a Data Science Intern at Ripeta working on natural language processing. She attends Carleton College, where she specializes in American Studies and Spanish.
https://orcid.org/0000-0001-5095-4257

Josh Sumner is a Data Science Intern at Ripeta using R to assist in analysing. He graduated from Appalachian State University in May 2019 with a BS in Strength and Conditioning and a minor in Statistics. Josh is planning to pursue an MS in biostatistics in 2020 and is particularly interested in statistics, its applications in sports science, and its role in sports science research.
https://orcid.org/0000-0002-3399-9063

Leah Haynes is a Data Science Intern at ripeta working on natural language processing. She is a rising junior at the University of Illinois Urbana-Champaign studying Music Performance and Mathematics. In the future, Leah plans to pursue a graduate degree in Mathematics.
https://orcid.org/0000-0003-4379-2397

This report has been published by Digital Science which is part of Holtzbrinck, a global media company dedicated to science and education.

Digital Science, 6 Briset Street, London EC1M 5NR
info@digital-science.com

DOI: 10.6084/m9.figshare.9633158

Contents

"All research stakeholders have a responsibility to make their work both reproducible and falsifiable"

Introduction

Leslie D. McIntosh, Cynthia Hudson Vitale, Anthony Juehne

Research is, at its heart, a method of answering a question. How well we ask that question and how we choose to explore the routes to a solution plays a critical role in the quality of the research output. Research quality is difficult to define but we venture to suggest here that there are several key factors that objectively underlie good quality research and which do not speak to the novelty or the importance of a result, which we consider to be different from "quality". These factors include the rigor with which the method was employed, the appropriateness of the method undertaken, and the statistical significance of the results.

All research stakeholders have a responsibility to make their work both reproducible and falsifiable. Reproducible: so that anyone can follow the stated method and reach the same conclusions; and falsifiable: so that the method used can appropriately test the hypothesis. Whether the reader is another researcher, a federal agency, commercial company, academic institution, funder or member of the public, the paper should be a route to test and recreate the research that has been carried out. This is the basis of the scientific method.

In this report, we examine falsifiability and reproducibility in a range of contexts related to scientific research, focusing on three areas that our tool, *Ripeta*, supports: Appropriate documentation and sharing of research data, clear analysis and processes, and the sharing of code.

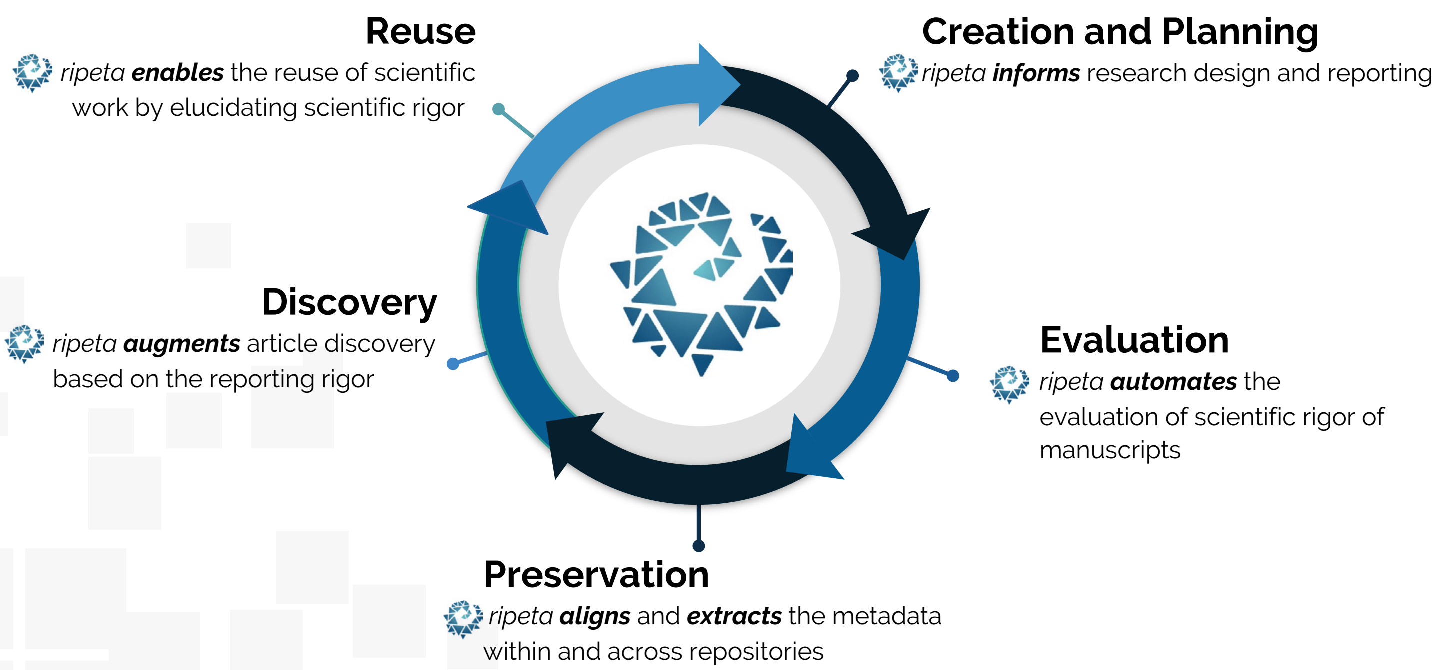


Figure 1: Ripeta improves research reproducibility across the entire research lifecycle.

Reproducibility and Falsifiability in Research

"We believe that the scientific community needs faster and more scalable means to assess and improve reproducibility"

We believe that the scientific community needs faster and more scalable means to assess and improve reproducibility. An important part of that is fundamentally changing how we think about reproducibility. The difficulty is that while we all have a sense of what reproducibility is in our own fields, reproducibility as a concept does not easily translate between fields. If we don't have a generally-defined concept of what constitutes reproducibility then we cannot quantify it objectively and it becomes difficult to advocate for it. We argue that establishing a well-structured framework against which assessments of reproducibility can be made, alongside appropriate reporting, allows the barriers in reusing scientific work, supporting scientific outcomes, and assessing scientific quality to be reduced.

Ripeta provides an intuitive and evidence-supported method for multiple stakeholders to evaluate the potential for reproducibility within published literature. The *Ripeta* process facilitates the understanding of research transparency and accessibility across domains and institutions. *Ripeta* improves the research data lifecycle for many customers and stakeholders by informing research design, automating the evaluation of scientific rigor of manuscripts, aligning and extracting metadata across repositories, augmenting article discovery based on reporting rigor, and enabling the reuse of scientific work (Fig 1).

To Falsify is to Verify, and Begins with a Hypothesis

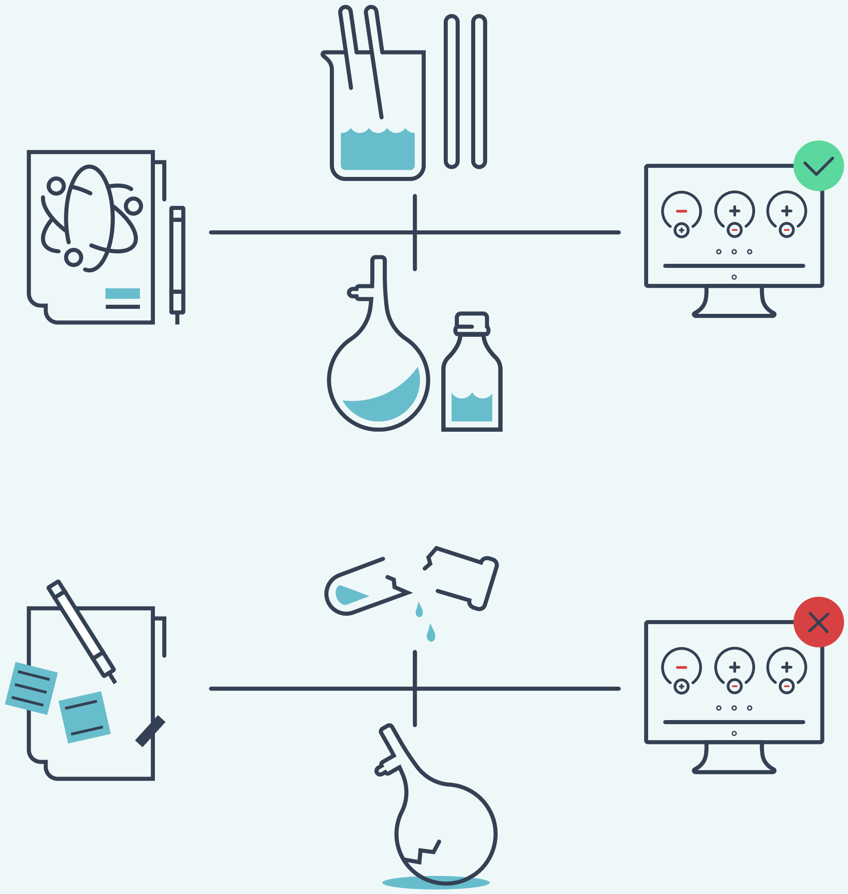
What exactly is falsifiability? Where have we encountered it before? How does it relate to the scientific method? Good-quality research should start with a testable hypothesis that involves well-defined variables and should result in a clear set of observations. These ingredients need to be fully contextualized to be understandable and reproducible by a generally knowledgeable bystander who was not involved in the original research. Critically, the method must be able to stand up to external scrutiny and questioning. Without this structure, we are building vague knowledge to be used in later predictions, and a collection of facts that cannot clearly stand up to critique.

Ripeta is specifically designed with this type of reproducibility and falsifiability in mind - we aim to help scientists double check the structure of their communications to ensure that they are making their research as useful as possible to others. We start by identifying and evaluating all the elements of a scientific workflow that have been reported in an article. We then highlight potential areas of improvement, so that changes can be made to each process to strengthen the credibility of the research being reported.

"*Ripeta* evaluates the potential for reproducibility within published literature"

"It is said that you learn more from failure than from success"

It is said that you learn more from failure than from success and this is definitely true with *Ripeta*. To develop algorithms that can identify the hallmarks of reproducibility and falsifiability, it is not enough to look only at successful experiments. We have extensively studied 'failed' experiments and negative results so that we can understand the true value of a set of results. Due to our current research culture, this is a particularly difficult part of what we do since it is generally difficult to publish negative results in good quality journals: There are sociological barriers to developing tools in this space.



Research can produce both positive and negative results. Negative results are not bad, and are an inherent part of the research progress.

Ripeta helps researchers identify which outputs are as a result of well-conducted and well-reported research, giving validity to the work conducted, and allowing researchers to selectively build on research that comes from good reproducibility and falsifiability.

For example, Ripeta assesses the quality of research methods by finding a clearly-defined, machine-readable statement of the research purpose in a scientific paper.

Figure 2: Good science needs good methods and may have both positive and negative results.

Falsifiability and its Importance Today

"We live in a time when the relationship between research and the data upon which it is based is changing"

Historic breakthroughs have often come from so called 'blue-skies' thinking, often motivated simply by a researcher being curious about the world in which they live. But, little in research remains purely abstract, and over time even the most seemingly abstruse research has led to technologies that we take for granted today. Sadly, today's funding environment often means that this exploratory or curiosity-driven research is not funded, whereas more application-driven research, or research with nearer-term goals or payoffs is supported. Even fundamental research needs to stand up to the same level of scrutiny as more applied areas.

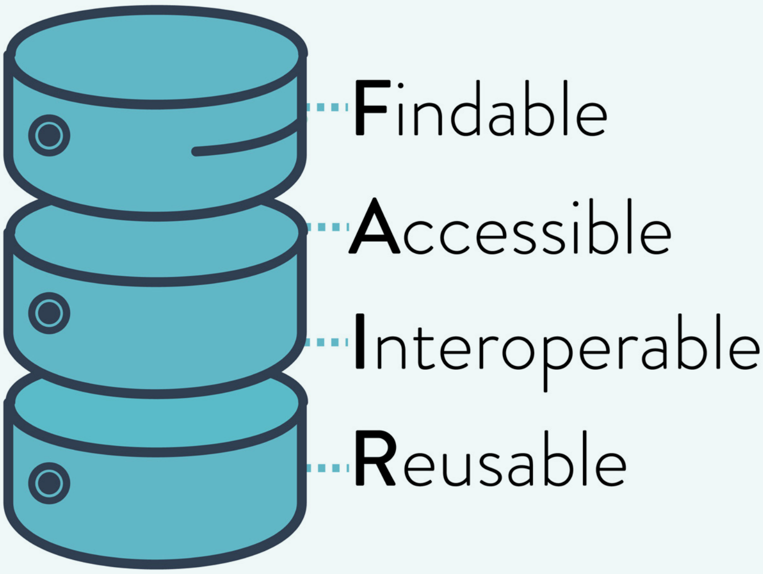
We live in a time when the relationship between research and the data upon which it is based is changing. Datasets are larger. The computational capacity required to parse them and gain insight is becoming more significant, but at the same time more available. The software that performs these complex analyses is becoming not just a research artifact but a research output in its own right. But, all of this complexity makes it significantly more difficult to understand whether each piece of research is valid. It is paradoxical - more data should lead to clearer and more reproducible research, but the volumes of data that are now being produced require new machinery to allow us to understand and interpret them. There are simply too many moving parts. In response, we need to build structure into our research processes that automate the checking of the process itself and alert us to problems when they arise. This new machinery of checks and counterbalances needs to take both falsifiability and reproducibility into account. These new mechanisms need to be informed by carefully crafted questions if they are to be effective in maximising the value of today's research output.

Reproducibility Needs Transparency and Accessibility

Transparency is where reproducibility begins. It can be thought of in simple terms as the sufficiently clear and complete documentation of the research materials, processes, and findings that allows for the re-application of each step of a piece of research. Accessibility is a more multifaceted and more challenging component, requiring that information such as a dataset or analysis code must be made available in a location that is discoverable, and in a form that people can use. Accessibility by definition requires all materials to be available, and also arranged in a manner that makes them fit for the purpose of reproducing the research. For example, data should be formatted in a manner that allows it to be reused, reparsed and reinterpreted; computer code should be sufficiently well annotated and come with sufficient information about the environment in which it was run so that it can be re-run at some future date with consistent and predictable results. Given that many data formats are proprietary or at least project specific, and that much code is custom written, the bar to making this reproducible is quite high and necessarily labour intensive. Finally, while not all research materials need to be accessible due to confidentiality and/or anonymity, achieving adequate transparency is essential to reproducibility.

FAIRification and Falsifiability

'FAIR' data are findable, accessible, interoperable and reusable. FAIR principles describe how data should be contextualized, published and shared in a way that supports its reuse in responsible science. Though FAIR principles have, at their core, the concepts of verifiability and falsifiability, both of which encourage more responsible science, we require more than this. In this aspect, we believe that tools like *Ripeta* can make science better by analysing the research processes employed and allowing improvements to be made to the scientific method, to maximize the credibility of research outputs.



FAIR data and falsifiability share many principles. For quality research, outputs must be easy to find, accessible without barriers, and capable of being reused in consequential work.

Examples of how this can be achieved include authors sharing their data, algorithms, or other information integral to the science in their publications in enough detail to allow others to verify or replicate the claims.

Responsible science requires that researchers publish in a FAIR way.

Figure 3: How FAIR data relates to falsifiability.

Data Sharing and Data Documentation

Leah Haynes

The Importance of Data Documentation

Clear documentation of the entire research data lifecycle is integral to reproducibility. Good data documentation, which includes design, data collection, data cleansing, and analyses leads to “good” science. Well-documented science and research enables further advancement through transparency and adequate data documentation.

How Data Should be Documented

To **validate** and **replicate** findings researchers need access to the same data and certain components of previous studies, as well as documentation and validation of the methods used, for their research to be reproducible. These data need to be documented with a high-level of **clarity**, demonstrating how they represent observable facts, how they were collected, and related limitations in data quality and generalizability. Shared data should also include documentation of relevant ontologies, metadata descriptors, and abbreviations.

Good Data Sharing in Practice

A lot of data are still not made openly available. Although the majority of research articles contain data availability statements, it is rare to find truly accessible data. This may be because of privacy restrictions, consent issues, or simply a poor data-sharing culture, so the research community as a whole needs to find a way to improve data availability and sharing to make research as efficient and productive as possible.

For example, some studies claim to have accessible data, but in reality these data are often difficult to obtain. The authors of a paper with limited useful access to their data may make a statement along the lines of:

> **Data Availability**. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

While one may argue that these data can be obtained if one contacts the corresponding author, the data however are not easily accessible, and this potentially complicates the research replication process. Additionally, “reasonable request” is a vague and subjective term and no

"Clear data analysis reporting is not only related but critical to the practice of good science"

information documenting the verifiable quality and generalizability of this data is available.

Authors of papers with accessible and available data may choose to share their data via a data repository such as Figshare, Github or Zenodo, and include links to the data, such as:

Data Availability. The raw data and R scripts are available from the Figshare repository database doi:10.6084/m9.figshare.2065284

Publicly available data and thorough data documentation facilitates computational reproducibility and improves science as a whole.

Reporting Clear Analysis and Processes for Better Science

Sasha Mothershead

Why is a Clear Analysis Process Important?

Buried within the criteria page for submission to *Nature* is a line on statistical transparency[1]. Other *Nature* requirements, such as an accessible dataset or the admission of conflicts of interest, may seem like obvious components to reputable science, but why specifically demand minute details of an analysis process? The answer is simple:

Clear data analysis reporting is not only related but critical to the practice of good science. This is primarily due to its promotion of two fundamental scientific principles: **reproducibility** and **transparency**.

How is a Clear Analysis Process Related to Reproducibility?

The ability to reproduce scientific findings is an essential tenet of the scientific method; without reproduction, falsifiability, and external verification, scientific claims could be staked without additional investigation or community scrutiny. Valid discovery must therefore be replicable, testable, and reliable across multiple investigations—and clear analytics are a fundamental component of this process. An analysis process, in this context, refers to the information related to how data are treated *after* it has been obtained through the experimental process.

[1] *https://www.nature.com/nature/for-authors/initial-submission*

This includes aspects such as statistical procedures, data preparation, and model-creation. Transparent analysis documentation provides other investigators with easily replicable procedures, and allows researchers to make sense of their results in a comparable manner. Determination of whether or not the same conclusion has been reached across different experimental renditions, for instance, is reliant on the ability to analyze and compare the data in the same way.

"By transparently reporting the chosen analyses, researchers can make more informed decisions about the research methods and generalizability of the outcomes"

How is a Clear Analysis Process Related to Transparency?

Explicit analytics provide a level of transparency to scientific work. Through knowledge of the analysis process, other researchers can evaluate the utility and application of chosen analysis methods, combating misrepresentation of data and providing a level of collaborative scrutiny that is critical for proper, rigorous science. This issue of transparency becomes increasingly important as technological advances create larger datasets and more complex analytics. Increased size and complexity foster greater potential for error, making it imperative that there are easier ways to replicate and authenticate results, even without full experimental reproduction requiring recollection of experimental data and recreation of analysis operations. Thus, transparent analytics are integral to the creation of a high community standard for good science.

What a Clear Analysis Process Looks Like

Clearly stating the chosen analysis method used in their studies, researchers are improving the falsifiability of their research by being open about their data analysis. The following example, though jargon filled, gives an example of how researchers can achieve clear analysis processes:

> The original studies included used the mean and SD to assess the MDI and PDI of the infants. We pooled the MDI and PDI scores of each study separately using the DerSimonian-Laird formula (random-effects model) [31]. Statistical heterogeneity [32] between the studies was assessed using the Q and I2 statistics. Values of p<0.1 and I2>50% indicated high heterogeneity [2] (from Xiao et al., 2018, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208302).[2]

The above extract provides an example of a clearly articulated analysis process, stating the types of statistical tests run, providing links to helpful external information, and specifying how the final values were assessed. This description aids the overall experimental reproducibility and, combined with the article's provision of a full final dataset, provides particularly easy authentication of analyses. By transparently reporting the chosen analyses, researchers can make more informed decisions about the research methods and generalizability of the outcomes. This does not mean that one agrees with the chosen analyses, and it could

[2] *Xiao, Dongqiong, Tingting Zhu, Yi Qu, Xiaoyun Gou, Qun Huang, Xihong Li, and Dezhi Mu. "Maternal Chorioamnionitis and Neurodevelopmental Outcomes in Preterm and Very Preterm Neonates: A Meta-analysis."* Plos One*13.12 (2018): n. pag. Web.*

"Proper documentation of data analysis fosters "good science" by ensuring that findings are reproducible, transparent, and truthful"

be challenging to reproduce the findings, rather it means the analytical methods are transparently reported.

Specification of the Analysis Process is Integral to Good Science

Providing explicit analytical procedure aids replication of scientific analyses, and encourages transparency in the scientific community. Asking researchers to understand and justify their data-management choices allows other scientists to scrutinize the representation of these results. Proper documentation of data analysis fosters "good science" by ensuring that findings are reproducible, transparent, and truthful.

Despite the apparent importance of a clear analysis procedure to the scientific process, the inclusion of an analysis statement is not universal across published manuscripts. A recent review of 200 PLoS One manuscripts, for instance, revealed that 5% were missing a clearly outlined analysis procedure. As the greater problem of reproducibility looms over the scientific community, the provision of clear analytics is an important step on the path to cleaner, better science.

Sharing Code

Josh Sumner

A critical component of the scientific process is making sure that any code used to test hypotheses and draw meaningful conclusions from research is clearly documented to allow for reproducibility.

Sharing Code and the Goals of Science

Modern research studies rely on collecting information, any interventions made, and the conditions under which the information was collected. Most of that information is then analyzed using computer code. Scientific writing tends to have extensive explanations of a researcher's intentions, methodology, and results. Equally important, but often overlooked, is how the data were analyzed, as discussed by Sasha Mothershead earlier in this report, and the computational environment in which the analysis processes were executed. Results can differ when the same data are put through different code, different software, or even different versions of the same software. These various dependencies within the computational environment, if not documented thoroughly, all represent potential errors when attempting to reproduce code-driven data analytics. By supplying code, documenting which version of software was used, and storing code for future reference, science can be made more accurate, more reproducible, and more useful to scientists within and across domains and geographies.

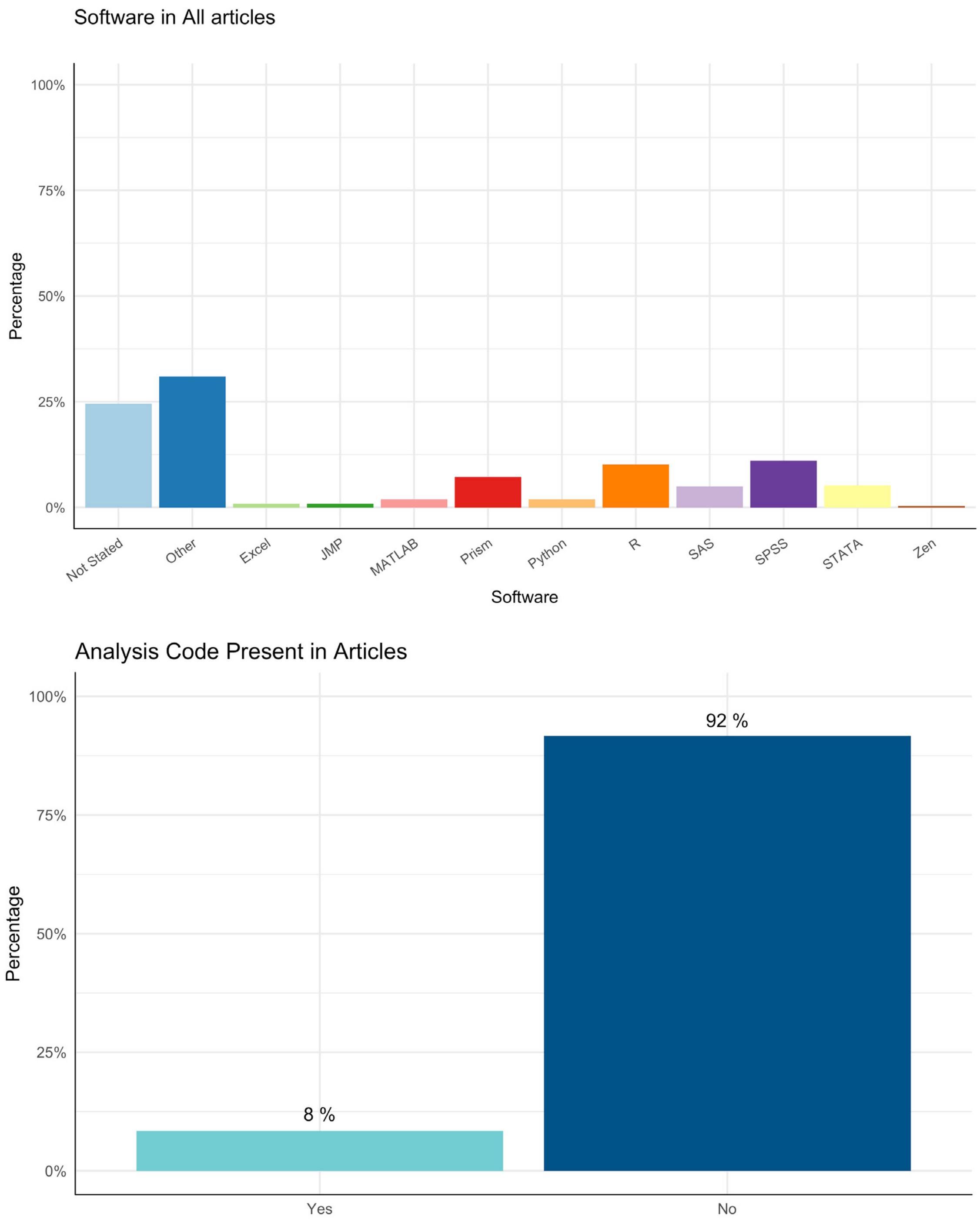
"Results can differ when the same data are put through different code, different software, or even different versions of the same software"

The Importance of Sharing Code

If the analyses are not well documented and cannot be explained or recreated, then the conclusion loses value and cannot be verified for scientific quality, accuracy, or precision. Yet, while authors are describing their analyses, few are sharing their code (Fig. 4). If the analyses are well documented and the analysis code is provided, then the conclusion can be verified and compared to experiments which used similar methods to allow the scientific community to learn more about the generalizability and strength of the result or experimental methodology.

Figure 4: An analysis of publications from the Ripeta database showing software citation, and the small percentage of publications that have shared the code that was used.

Citing Software and Sharing Code



"The good news is there are resources like *Ripeta* that are available to analyze existing procedures and publications, and potentially improve the reproducibility and integrity of research procedures"

Making Science Better: Some Final Thoughts

As reflected in this report, falsifiability is an integral part of the research process. It adds credibility to research and allows further work to build on solid foundations.

Irrespective of whether the findings of a study are disproven at some later date, falsifiability is important. With increasingly specialized and complicated research, this level of scrutiny is now capable of being supported and enforced through both policy and practice by all research stakeholders until more studies can be clearly articulated and conducted, and their findings compared.

The pursuit of knowledge is important, and should be undertaken thoroughly and accurately. Accessible, reproducible research is an important and often overlooked aspect of that pursuit. The good news is there are resources like *Ripeta* that are available to analyze existing procedures and publications, and potentially improve the reproducibility and integrity of research procedures, ultimately making scientific research the best it can be.

Part of **DIGITAL** science

Altmetric

ANYWHERE ACCESS

CC Technology

Dimensions

figshare

gigantum

GRID

ifi CLAIMS

Overleaf

readcube

ripeta

SYMPLECTIC

writefull

DIGITALscience Consultancy

digital-science.com