




## What if It Didn't Happen: Data Management and Avoiding Research Misconduct

**Heather Coates**, Indiana University - Purdue University Indianapolis, IN, USA 

**Abigail Goblen**, University of Illinois at Chicago, IL, USA, [agoblen@uic.edu](mailto:agoblen@uic.edu) 

**Kristin Briney**, California Institute of Technology, Pasadena, CA, USA 

---

### Abstract

As research misconduct has created reproducibility and researcher reputation concerns, there is an opportunity to recommend data management techniques to assist researchers as they seek to prevent these issues. Also central to the discussion are issues of power in the conduct of research, particularly in upholding the values of honesty and accountability. This commentary discusses how data professionals can engage in practical strategies to protect against allegations of misconduct.

---

**Received:** June 21, 2023 **Accepted:** November 14, 2023 **Published:** December 20, 2023

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), which permits unrestricted use, distribution, and reproduction in any medium non-commercially, provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc/4.0>.

∞ OPEN ACCESS

## Introduction

Among the challenges of data governance, emerging concerns of reproducibility, and ongoing publish or perish culture of the academy, increased concerns have arisen around research misconduct. Recent case examples and the rise of Retraction Watch have demonstrated the frequency where inadvertent and potentially deliberate research misconduct has been caused or facilitated by a lack of research data management. Research misconduct has impacted the trust that we are able to put into past and current research, and as end users of that research, we care about having reproducible scientific findings in healthcare, food science, psychology, transportation, economic policy, etc. While data management has, to date, often been focused on grant related planning and the early stages of data capture, data management professionals have a growing role in prevention and response of research misconduct—providing transparency and assisting with power dynamics amongst research teams. In this commentary, inspired by an RDAP Summit workshop, we explore case studies as a guide for understanding research misconduct investigations, identifying data management concerns where things have already gone wrong. We subsequently use data management activities as a mechanism for preventing inadvertent misconduct and engaging in relationship building and education. These goals allow for data management professionals to develop new mechanisms for outreach and instruction in order to support the myriad data challenges of research teams and communities.

## Research Misconduct

In the United States, the commonly used definition of research misconduct comes from [42 CFR Part 93](#), which describes three types: 1) fabrication, or “making up data or results and recording or reporting them;” 2) falsification, or “manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record;” and, 3) plagiarism, or “the appropriation of another person’s ideas, processes, results, or words without giving appropriate credit.” This code also clarifies that research misconduct “does not include honest error or differences of opinion.” In practical terms, this definition raises the bar relatively high for proving that fabrication or falsification was intentional and can be substantiated. At most universities research misconduct falls under the Office of Research and the person overseeing this may be called a Compliance Officer. This puts an emphasis on proving misconduct and intent rather than the pursuit of research with integrity, or higher ethical conduct.

In contrast, the Singapore Statement (Resnik & Shamoo 2011) offers a more holistic view of research integrity as an ideal state of practice. It describes four principles: 1) honesty in all aspects of research; 2) accountability in the conduct of research; 3) professional courtesy and fairness in working with others; and, 4) good stewardship of research on behalf of others. Additionally, the document articulates 14 responsibilities. This more inclusive approach encourages adoption of proactive strategies to demonstrate the integrity of research conduct and products.

However, the disciplinary culture and acceptable standards of conduct varies widely across disciplines; thus, there is no single set of practices or norms that can be universally applied. Each community must agree upon locally relevant norms, resources, constraints, and a particular vision of research integrity.

### **Value of Case Study Examples**

The majority of research misconduct cases are not publicized, limiting the ability of both the academic community as well as the public to understand the frequency of investigations, the outcomes, and the potential impact on scholarship. While the blog *RetractionWatch* has significantly improved the awareness of misconduct where it led to publication retraction, this is only a part of the investigations conducted. Further, due to the stigma of investigations even where the researcher is not found to have committed misconduct, it is unlikely that we will ever fully know how often research misconduct is asserted. Additionally, institutions are often constrained from making the results of internal investigations public, so offenders are able to obtain employment elsewhere and continue their patterns of misbehavior.

As such, we must look to completed and published investigations as examples to draw from when considering research misconduct and research data management. These extreme stories illustrate how data management is central to supporting or refuting misconduct, though we recognize that the majority of allegations are unlikely ever to rise to this level of severity.

We present three examples of known cases of research misconduct which could be used to better understand how a lack of data management may have allowed research misconduct to happen or continue over time:

- Diederick Stapel was dean of faculty in social and behavioral sciences at Tilburg University and a social psychologist in the Netherlands. The investigation against him began in 2011 and examined publications on which he was an author from 1993-2011, spanning three universities. The final report, released in 2012 found a long-standing pattern of data manipulation and fabrication along with deception of collaborators regarding the origins of the data (Levelt Committee, Noort Committee, and Drenth Committee 2012).
- Ching-Shih Chen was the chair of cancer research at Ohio State University and a professor of medicinal chemistry. The university began its investigation into Chen in 2016, with an initial inquiry determining that a more comprehensive and further examination was necessary. The full investigation found that Dr. Chen alone falsified data through image manipulation and failed to adequately review data provided to him. The Weinberg Group was contracted by OSU to investigate allegations related to Investigational New Drug (IND) AR-42, including four publications related to its safety and efficacy. They determined that the inconsistencies in Western blot data had no impact on the safety or development of AR-42 (The Ohio State University 2017).

- Brian Wansink the director of the Cornell University Food and Brand Lab. In 2018, a Cornell investigation, spurred by post-publication re-analysis of his work, found that Wansink committed scientific misconduct. Unfortunately, the Cornell investigatory report is no longer available. Based on the available evidence, it appears that Dr. Wansink used irresponsible reporting practices, including overstating conclusions based on the available evidence and inconsistently reporting key details of studies. Regardless of intent, he was not able to produce the data to support published results (Robinson 2017; van der Zee et al. 2017).

In all of these instances, the researcher in question held significant positions of power within their institution. These positions likely protected them from any questions that arose from students, staff members, or other reliant members of their research teams.

### **Misconduct and the Relationship to Data Management**

There is not empirical proof, nor is there likely to be, that data management activities in and of themselves will prevent misconduct. Much as with not having a computer crash or a server fail, implementing these activities does not guarantee lack of misbehavior or malintent. However, by identifying activities that have led to findings of misconduct, we can propose education and practices that encourage appropriate research conduct and facilitate record keeping, documentation, and data sharing. Documentation and transparency can prevent temptation towards data manipulation and allow for easy verification from within or beyond the research team. Data management activities and education also enable research teams to engage in self-examination of their current practices and identify weak points of communication or data handling which may lead inadvertently to research misconduct, such as inappropriate reuse of data or data going missing.

Table 1 provides a chart of data management activities which could be implemented within research teams in order to avoid research misconduct. When considering implementing these activities, there are four areas which you may wish to consider: the benefits to the research team of adopting this activities; the power issues inherent in adopting or not adopting this activities; how this might specifically reduce the potential for or mitigate misconduct; and the creation of talking points to assist in starting these conversations with researchers.

**Table 1:** Example Data Management Activities for Avoiding Research Misconduct

Create template language for the data repository	Catalog data sets held by the lab/research team/university	Write a project closeout document
Identify storage back ups	Implement metadata standards	Convert files to a non-proprietary and/or preservation format
Write a data management plan	Assess dataset credibility	Develop a data policy for lab/unit/college/campus
Cite data	Use permanent identifiers such as DOIs	Use a university sponsored Github
Release data under an open license	Develop a research group file naming convention	[And other practices]

An example of one of these activities is to consider the use of file naming conventions.

- **Benefits:** Among the benefits for using file naming conventions on a research team is ensuring that all members of the team can organize, find, and disambiguate files easily and quickly. It assists with documentation across experiments, which can reduce the need for replication.
- **Power Issues:** File naming conventions can be used to address power issues within the lab because it introduces a lower context culture, where it is not inherent on new or more junior members of the lab to guess or learn how files are named or formatted, but instead presents a clear path towards understanding. This enhances the potential for all members of the research team or, later on, other researchers to be able to follow or reproduce the work. It may also reduce the ability of individual research team members to hoard files, which may otherwise reinforce research team hierarchies.
- **Reducing Potential Misconduct:** File naming conventions can assist researchers in preventing research misconduct as it can be easier to document which files were used for analysis, which assists in proving that research was conducted in the manner it was described. File naming can additionally show the history of documents and files as they go through various stages of data cleaning, storage, and use to demonstrate veracity of data collection and standardization as well as preventing data duplication.
- **Creating a Talking Point:**
  - “When you name your files consistently, it helps you by making sure you are analyzing the correct file and documenting your research process as transparently as possible.”

## Questions of Power

Many case studies of research misconduct involve misuse of power. Primary Investigators carry significant power within collaborative research, which many have used to hide their misconduct. Graduate students, on the other hand, often have the least power within research yet can be in the best position to identify where misconduct is occurring. An example of these power imbalances is data fabricator Diederik Stapel, who used his power over his students and collaborators to hide the true source of the data he supposedly collected (Levelt Committee, Noort Committee, and Drenth Committee 2012). Complicating these power differences is the role of the university, which often reinforces a PI's power until accusations are serious enough to merit formal investigation. Once any accusations arise, there is also the possibility of retaliation to further tangle questions of power within research relationships. Case studies involving grave misuse of power are common, though smaller misuses of power may also be rampant but better hidden or more difficult to prove.

Navigating issues of power is central to investigating research misconduct and so must be addressed in linking data management activities to research ethics. There are many questions to ask when thinking about power and research misconduct:

- What happens when institutional authority is not recognized? When researchers feel above/beyond it?
- What happens when institutional leaders fail to recognize their authority or abdicate their responsibility to hold researchers accountable?
- What power do trainees and students have? What can they do?
- How do junior researchers practice integrity in a culture that doesn't reward it?
- How can power be lent or borrowed?

Data management specialists should also consider the power they bring into the research process. There is power in observation. There is power in asking questions. There is power in documentation. While these may be softer powers, data management provides tools to help shift power imbalances by making clear exactly what is happening with the data. Our own power, as well as the power of all participants in research, must be a key consideration in interrogating data management's role in research ethics.

## Translation to our Institutions

While research misconduct is often handled by separate offices from those that support data management at academic institutions, it is beneficial for data management specialists to be aware of how misconduct can occur and how this relates to data management. Data librarianship involves relationship building and thinking about the entire data lifecycle, making research ethics a natural avenue for building further connections.

There are many ways to strengthen work in this area. One mechanism for this is to develop talking points to bring up when discussing data management with colleagues. Opportunities abound to incorporate this topic in teaching about data management, including everything from adding misconduct examples to existing workshops up through giving an entire presentation for the campus' Responsible Conduct of Research (RCR) series—a requirement for all NIH and NSF trainees. Those looking to pursue this topic more fully should consider partnering with their institution's Research Compliance or Research Integrity Office.

## Conclusion

With the continued pressures on researchers to obtain competitive grant funding and publish or perish in order to obtain and retain academic positions, there will remain the risks of deliberate or inadvertent research misconduct. Expansive cases of data fabrication or manipulation may lead to further mistrust of academic researchers and institutions. While data management cannot wholly prevent or mitigate these issues, such activities serve as an additional mechanism to promote well documented and, where appropriate, transparent research behaviors in order to reduce errors, duplication, and potentially lessen the impact of power dynamics. We encourage data professionals to seek opportunities to incorporate understanding of research misconduct, work with their institutions to add data management training in RCR programming, and partner with their research compliance officers to aim for the goal of research misconduct not happening.

---

## Acknowledgements

The content of this article is based on the workshop "What if It [Didn't] Happen: Data Management and Avoiding Research Misconduct" originally presented at [RDAP Summit 2023](#).

## Competing Interests

The authors declare that they have no competing interests.

---

## References

- "42 CFR 93.103 -- Research Misconduct." n.d. Accessed June 6, 2023. <https://www.ecfr.gov/current/title-42/chapter-I/subchapter-H/part-93/subpart-A/section-93.103>.
- Levelt Committee, Noort Committee, and Drenth Committee. 2012. "Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel." Tilburg, Netherlands: Tilburg University.
- The Ohio State University. 2017. "Final Report of the College of Pharmacy Investigation Committee Concerning Allegations of Misconduct in Research under the University Policy and Procedures Concerning Research Misconduct."
- Resnik, David B., and Adil E. Shamoo. 2011. "The Singapore Statement on Research Integrity." *Accountability in Research* 18(2): 71–75. <https://doi.org/10.1080/08989621.2011.557296>.

Robinson, Eric. 2017. "The Science behind Smarter Lunchrooms." e3137v1. *PeerJ Inc.*  
<https://doi.org/10.7287/peerj.preprints.3137v1>.

van der Zee, Tim, Jordan Anaya, and Nicholas J. L. Brown. 2017. "Statistical Heartburn: An Attempt to Digest Four Pizza Publications from the Cornell Food and Brand Lab." *BMC Nutrition* 3(1): 54.  
<https://doi.org/10.1186/s40795-017-0167-x>.