

Shelly Hunt

North Carolina State University

Department of Agricultural and Biological

Engineering 3100 Faucette Drive, Raleigh, NC 27695

Ethical Data Management in Biological and Agricultural Research

Proper handling, storage, and use of sensitive data is an ethical obligation of all research professionals. When academics are excited to dive into data exploration, visualization, and analysis, it is easy to overlook important boundaries and limits concerning data access, privacy, and confidentiality, elements that should be established at the start of projects. First and foremost, data should never be abused with malicious intent. Data must also be safeguarded with proper storage and infrastructure. Accidental leaks due to carelessness are as harmful as intentional data misuse. On the other hand, there is a fine balance between keeping data private and ensuring transparency in research. It is unethical to falsify information, so researchers must be transparent about their data sources, data processing steps, and analysis methods. Academics are morally obligated to uphold both aspects and should conduct research honorably.

Background

“The transition to electronic records during the development of the digital age has left many people increasingly uneasy about the use of records relating to themselves and their families” (De Guise, 2020). Most research projects incorporate data, whether it be from conducting a survey or a planned statistical experiment. Private information must be safeguarded to respect the intellectual property of others and keep individuals safe. The Association for Computing Machinery (ACM) is a professional organization aimed at propelling scientific computation education and professions (ACM, 2021). ACM has published a code of ethics to guide individuals working in computing. These guidelines define ethically correct behavior for managing data and information systems. The code should not only be followed by individuals working strictly within the computing industry, but any researcher whose work utilizes data. This includes the agriculture and environmental engineering sectors.

Data Privacy

Most individuals engage in some form of data transaction every day. Whether it be liking posts on a social networking site, providing your phone number to stores when making purchases, or tracking commodity prices. In this digital age, people are almost always contributing to or utilizing the constant

flow of information and data which surround our lives. The ACM recognizes in its code of ethics that all people are stakeholders with a right to autonomy in computing (Gotterbarn et al., 2018). When personal information is recorded, it is the duty of the data holder to be stewards of that information. Code 1.6 “Respect privacy” of the ACM states that collected personal information must be kept to a minimum and only used for its originally intended purposes. All potentially identifiable features must be masked and re-identified. Code 1.7, “Honor confidentiality,” implores researchers and computational professionals to keep secure data private unless it is potential evidence of illegal activity, in which case local authorities should be contacted.

A data use agreement (DUA) is an important legal document often created concerning the collection, storage, usage, access, and longevity of data. This document, established and agreed upon by all parties before a project commences, sets clear limitations and expectations on how data ought to be managed and stored. Researchers working with sensitive data must respect the wishes of research subjects and data sources. All activities must adhere within the defined boundaries agreed upon by the stakeholders. DUAs are helpful legal documents which protect all participants and should be used more among research groups in biological and agricultural engineering. One example of a successful agricultural research project utilizing a DUA is Sweet -APPS, a milt-disciplinary advanced analytics project with a three-way DUA in place to protect the university and their industry partners from whom they are gathering information and other resources from (Jones, 2020).

Data Security

Accidental data leaks may occur even if researchers have the best intentions to honor data privacy. Also, carelessly designed data storage solutions may be subject to cyber-attacks. Code 2.9 of the ACM ethics is “Design and implement systems that are robustly and usably secure” (Gotterbarn et al., 2018). Research professionals should do everything in their power to safeguard data through proper practices and creating competent data infrastructures. The US Department of Homeland Security (2018) warns of potential vulnerabilities within precision agricultural technology systems. These systems are widely used within agricultural engineering research. Many farm information management systems have

insufficient update procedures and privacy controls which may lead to intentional data theft or unintentional data breaches (DHS, 2018).

Although not as common in the agricultural industry, there have been many notable data breaches within the past few years. A prime example was the 2017 Equifax data breach, in which private information of over 147 individuals was compromised. (FTC, 2020). Organizations which have suffered data breaches could have been more proactive by ensuring their data storage centers were useably secure as defined in the ACM code of ethics. Researchers must learn from the mistakes of companies who have experienced data breaches and heed the warnings of the Department of Homeland Security to keep confidential biological and agricultural research data secure.

Agricultural producers who collaborate with technology developers such as university researchers are concerned about their data security. 76% of producers interviewed by the Sweet-APPS research team responded that they had concerns about who has access to their production data as they fear they could lose their competitive advantage (Grieger et al., 2022). Data owners are willing to share their information to research teams if they are confident in the researcher's plans to control who has access to their data. Additionally, 69% of respondents believed it was important to have data privacy and protection for specific farm management and phenotypic data, such as fertilizer inputs and planting techniques. However, only 41% of surveyed stakeholders felt it was important to have the same security measures in place for environmental information, such as weather and pest pressure (Grieger et al., 2018). These data, especially weather information, are perceived as publicly and freely available so most respondents didn't feel the need for data protection. Concerns of industry partners should always be used to facilitate and guide research projects and data security practices.

Open Science

It's important to keep private data secure, but research projects must be open and honest about data sources, methods, analyses, and results. "1.3 Be honest and trustworthy" of the ACM code of ethics details the importance of truthfully disclosing system strengths and weaknesses, not falsifying data, not making false claims, and speaking out about any potential conflicts of interests or risks (Gotterbarn et al.,

2018). Researchers should never lie, falsify data, or maliciously manipulate results. When academics share important findings of their work, they should clearly state the steps they have taken throughout the research pipeline. Traditional research papers seldom include enough detail on data analysis, nor do they include code and data necessary to reproduce the results (NASEM, 2020). Open science initiatives encourage researchers to openly share their data and code to be transparent in their methods of handling outliers, augmenting data, assumptions, and their analyses. However, that does not mean sensitive data must be compromised. It is still the responsibility of the researchers to uphold ethical practices when handling data and private information and do all that they can to protect it.

Conclusion

The integrity of a research project starts with how you handle sensitive information. Researchers must be the safeguards of the data others entrust in them. When an individual is in possession of data, whether they own the data or not, it is their duty to serve as a custodian and protect it to the best of their abilities (De Guise, 2020). Many individuals find themselves stewards to sensitive data, and it is the duty of the steward to protect these data by upholding proper data management practices. Companies such as Equifax have faced the consequences of not properly protecting private information, and researchers must learn from these mistakes.

Agricultural stakeholders are willing to partner with researchers to share information, however they have data privacy concerns which must be considered when designing and conducting responsible research. Important steps can be implemented to ensure proper data practices, such as masking certain variable names or creating dummy variables to hide sensitive information and to de-identify data. Implementing a DUA can help define the expectations of all parties involved. These are considerations which every biological and agricultural engineering research professional should take to preserve the ethical and moral integrity of the discipline.

Works Cited

- ACM. (2021). About ACM. Association for Computing Machinery. Retrieved from www.acm.org/about-acm
- De Guise, P. (2020). *Data Protection: Ensuring Data Availability* (2nd ed., Vol. 1). New York, NY: Auerbach Publications. <https://doi.org/10.1201/9780367463496>
- DHS. (2018). Threats to Precision Agriculture. Department of Homeland Security Office of Intelligence and Analysis. Retrieved from https://www.dhs.gov/sites/default/files/publications/2018%20AEP_Threats_to_Precision_Agriculture.pdf
- FTC. (2020). Equifax Data Breach Settlement. Federal Trade Commission. Retrieved from www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement
- Gotterbarn, D., Brinkman, B., Flick, C., Kirkpatrick, M. S., Miller, K., Varansky, K., Wolf, M. J. (2012). ACM Code of Ethics and Professional Conduct. Retrieved from <https://www.acm.org/code-of-ethics>
- Grieger, K., Zarate, S., Barnhill-Dilling, S. K., Hunt, S., Jones, D. S., Kuzma, J. (2022). Fostering responsible innovation through stakeholder engagement: Case study of North Carolina sweetpotato stakeholders. *Sustainability*, 14(4), 2274. <https://doi.org/10.3390/su14042274>
- Jones, D. S. (2020). Sweet-APPS. Retrieved from <https://sites.google.com/ncsu.edu/sweet-apps/>
- NASEM. (2020). *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25303>