



Ethical Use of Data with FRED[®]

[Diego Mendez-Carbajo, PhD](#), Federal Reserve Bank of St. Louis

GLOSSARY

Data mining: Analyzing large amounts of data to discover patterns.

Data mirroring: Replicating data obtained from a different location.

Data robots: Software-based automated processes that complete data-related tasks, such as downloading or checking for updates.

Hacking: Accessing data in a computer or a network without prior authorization from the owner.

Scraping: Copying and/or extracting information from a website, including the data, usually performed without permission via automated software.

Seasonal patterns (in data series): Ups and downs in data values that occur because of events that more or less follow a regular pattern each year.

Compelling Question

How should we use data?

Description

FRED[®] provides access to a wide range of time series data from more than 100 sources. Its terms of use describe the acceptable ways to gather and use data from FRED[®]. This guide describes ethical considerations regarding gathering, analyzing, storing, and distributing data for new data users and serves as a reference for advanced data users.

Introduction

Researchers must act ethically when gathering, analyzing, storing, and distributing data. If a dataset is constructed unethically, its use is also unethical, and no acceptable work can be done with it. Also, if a dataset is analyzed unethically, the conclusions drawn from it become invalid. Finally, if a dataset is not stored according to the conditions under which it was put together, or it is distributed after disregarding the conditions under which it was obtained, the researcher may be banned from conducting that type of work again.

The Ethics of Gathering Data

Data must be ethically obtained. If not, their use is rendered unethical—no matter the goals of the researcher. Stealing data is an example of an unethical way of constructing a dataset. **Hacking** into a data server or walking away with documents from someone's office breaks laws, and the person involved may be subject to criminal and civil charges. If the data describe persons, researchers must secure approval from an institutional review board protecting the people subject to the research before data can be collected. For example, an instructor studying how to improve the design and delivery of an academic course must receive clearance from her school's institutional review board to distribute surveys or collect data from the registrar's office about her students.¹

Businesses and government organizations gather large amounts of data when people go shopping, pay taxes, or request public services. An ethical data-collection process lets people know they are the subject of data gathering and describes the intended use of those data.

Computer-Assisted Data Gathering

Using computers to collect data from websites is an effective strategy to gather large amounts of data. For example, in 2018 Alice Wu used that technique, known as **scraping**, to create a dataset from all the natural language postings in the job search discussion forum econjobrumors.com. By downloading more than 2 million postings covering a four-year period, she was able to describe the unwelcoming or stereotypical culture on display in the forum.² Because the webpage describing the privacy policies of the website did not ban the practice of scraping, the dataset was constructed ethically, and the research was published in scholarly venues.³

Scraping a website for data slows down access to the data for all other users, and sometimes the website administrator explicitly bans the practice. This is the case for the FRED® database, which, as listed in its terms of use, prohibits

- **data mining**;
- **data mirroring**;
- scraping;
- **data robots**; and
- other similar data-gathering or extraction methods.⁴

An alternative way to access large volumes of FRED® data is to use its application programming interface (API). This piece of software code allows software programs to communicate directly with the database.

After a dataset is constructed, its quality is documented with citations describing its source.⁵ A good data citation allows a reader to find the data used in an analysis or referenced in a report. This way, the work itself can be replicated and checked for robustness.

The Ethics of Using Data

Ethical practices when analyzing data are part of the set of best practices that all researchers must follow. The same way a researcher tests for the presence of **seasonal patterns** in a series before attempting to forecast its values, ethical principles must guide the process of working with data. The Federal Data Ethics Tenets identify several guiding principles of data-based work, which include

- respecting the public, individuals, and communities; and

- acting with honesty, integrity, and humility.⁶

An ethical use of data benefits the public good and is mindful of its impact on unique communities and localities. At the same time, honesty, integrity, and humility when working with data are demonstrated when the limitations and known biases of the analytical techniques are openly stated.

In the case of the FRED® database, its public domain data can be used for personal, educational, and non-commercial purposes. This means that creating and embedding data graphs in documents or websites for purposes of teaching or personal use is allowed. The ethical use of FRED® requires citing the source of the data and noting they were accessed via the information service provided by the Federal Reserve Bank of St. Louis. Altering the data visualizations created through the FRED® portal is unethical because it may deceive readers into believing that the St. Louis Fed backs up the authors' work and endorses their conclusions.

Using Large Datasets Ethically

In the professional world of big datasets, where artificial intelligence technologies take a leading role in developing algorithms and automations that make probabilistic recommendations and decisions, the ethical use of data requires a thoughtful review to ensure they benefit the common good.⁷ For example, facial recognition technology makes use of digitized images, and the sources of those images (e.g., criminal records or driver's licenses) can introduce biases both for and against particular populations. Careful human supervision of the process and outcome of machine-assisted learning is needed to minimize unintended harm.

Because all analytical efforts are limited by the choice of methods and the scope of the data, their conclusions and recommendations must take into consideration their limitations. Describing how a dataset was put together can reveal some of the shortcomings of the analysis on which it was based. For example, if data are obtained from self-reported surveys or if the assembled dataset is small, the potential presence of response bias and the difficulty of conducting valid statistical tests undermine the robustness of the conclusions and prevent making generalizations. But the use of technical language in data descriptions, like that in the previous example, can

obscure their implications. Acknowledging data limitations demonstrates thoroughness and fosters additional research on the topic, potentially inviting additional work and new perspectives.

Finally, when data are owned by specific population groups, such as indigenous or aboriginal peoples, their own institutions can reserve the right to authorize the distribution of research findings. This principle, known as Indigenous data sovereignty, is a safeguard against data interpretations capable of casting whole populations in a demeaning or an insensitive light.⁸ For example, repeatedly highlighting relative disadvantages and disparities among those populations implicitly projects a cultural viewpoint to which they might object. In this case, good ethical practices related to studying social or economic issues extend beyond the stages of collecting and analyzing data we have described earlier in this essay.

The Ethics of Storing and Distributing Data

The ethical aspects of working with data extend to the management and distribution of datasets. Data gathered for analysis must be stored in accordance with the conditions stated when the dataset was assembled. For example, deleting people's names from a dataset is a common requirement to ensure a minimum degree of privacy before conducting statistical analysis. Similarly, creating sufficiently broad data groups so that no particular individual can be recognized is an ethical research practice. By contrast, carelessly allowing data breaches and potential misuse of personal data can have serious negative consequences for the people whose information is exploited and for the reputation of the researcher. Failure to maintain high ethical standards in the storage of information can ban researchers from doing that work again.

In the case of FRED®, redistributing copyrighted data series for commercial use is not allowed unless the data copyright owner authorizes it. This means that even if data are used for one of the authorized purposes described in the previous section of this essay, the user needs to obtain permission from the persons or organizations providing the data to make them available commercially. For example, commercial publishers and websites must secure written permissions to use data obtained through FRED® in their publications and digital platforms.

The last step in the ethical use of data involves making that work reproducible by other people. It involves documenting each of the steps taken in assembling and analyzing a dataset and allowing someone else to follow each of those steps. With documentation, the process of working with data is made transparent and any unethical practices are easier to see. Since 2019, the American Economic Association publishes articles in its journals only after authors have submitted copies of the data and computer code used in the analysis.⁹

Summary

There are ethical implications to gathering, analyzing, storing, and sharing data. This article has described best practices, providing examples related to the FRED® database. ■

Additional Resources

The Alan Turing Institute Data Ethics Group

<https://www.turing.ac.uk/news/alan-turing-institute-data-ethics-group>

European Data Protection Supervisor

https://edps.europa.eu/data-protection/our-work/ethics_en

DataEthics

<https://dataethics.eu/>

Notes

¹ See “Frequently Asked Questions About Institutional Review Boards.” American Psychological Association; <https://www.apa.org/advocacy/research/defending-research/review-boards>.

² See Wu, 2018.

³ See Wu, 2020.

⁴ See “Legal Notices, Information and Disclaimers.” FRED®, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/legal/>.

⁵ To learn more about data citations, see Mendez-Carbajo, 2020.

⁶ See “Federal Data Strategy: Data Ethics Framework”; U.S. Government; <https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf>.

⁷ See European Parliamentary Research Service, 2020.

⁸ See Walter et al. (eds.), 2021.

⁹ See “Data and Code Availability Policy.” American Economic Association; <https://www.aeaweb.org/journals/data/data-code-policy>.

References

European Parliamentary Research Service. “The Ethics of Artificial Intelligence: Issues and Initiatives.” European Parliament, March 2020; [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU\(2020\)634452_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf).

Mendez-Carbajo, Diego. “Data Citations with FRED®.” Federal Reserve Bank of St. Louis *Page One Economics*, October 2020; <https://www.stlouisfed.org/education/page-one-economics-classroom-edition/data-citations-with-fred>.

Walter, Maggie; Kukutai, Tahu; Carroll, Stephanie R. and Rodriguez-Lonebear, Desi, eds. *Indigenous Data Sovereignty and Policy*. Open Access Publishing in European Networks: Taylor & Francis, 2021; <https://library.oapen.org/handle/20.500.12657/42782>.

Wu, Alice H. “Gendered Language on the Economics Job Market Rumors Forum.” *AEA Papers and Proceedings*, 2018, 108, pp. 175-79; <https://doi.org/10.1257/pandp.20181101>.

Wu, Alice H. “Gender Bias in Rumors Among Professionals: An Identity-based Interpretation.” *Review of Economics and Statistics*, 2020, 102(5), pp. 867-80; https://doi.org/10.1162/rest_a_00877.

Please visit our website and archives at <http://research.stlouisfed.org/publications/page1-econ/> for more information and resources.

© 2021, Federal Reserve Bank of St. Louis. The views expressed are those of the author(s) and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.