

Association for Information Systems

## AIS Electronic Library (AISeL)

---

AMCIS 2022 Proceedings

SIG DSA - Data Science and Analytics for  
Decision Support

---

Aug 10th, 12:00 AM

### U.S. Department of the Interior: Sharing FAIR Data Fairly

Jason Duke

University of Arkansas - Little Rock, jeduke@ualr.edu

Follow this and additional works at: <https://aisel.aisnet.org/amcis2022>

---

#### Recommended Citation

Duke, Jason, "U.S. Department of the Interior: Sharing FAIR Data Fairly" (2022). *AMCIS 2022 Proceedings*. 6.

[https://aisel.aisnet.org/amcis2022/sig\\_dsa/sig\\_dsa/6](https://aisel.aisnet.org/amcis2022/sig_dsa/sig_dsa/6)

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2022 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# United States Department of the Interior: Sharing FAIR Data Fairly

*Emergent Research Forum (ERF)*

**Jason E. Duke**

University of Arkansas – Little Rock  
jeduke@ualr.edu

## **Abstract**

Government-produced data are consumed by thousands of scientists, researchers, industries, and students around the world daily, but are often difficult to locate because they are collected and stored in a duplicative state at varying levels of quality, inhibiting their usefulness for data science investigations and analysis. To address these challenges, the United States Department of the Interior bureaus have been implementing FAIR Data Principles into their data sharing strategies since 2016. Differing interpretations of the FAIR Data Principles are leading to data that are not documented uniformly and are not properly integrated for reuse. In order to establish a FAIR baseline, analysis of select datasets is being performed with peer-reviewed FAIR assessment tools. Delphi panels are being conducted with DOI Chief Data Officers and DOI Federal data consumers to gain insights as to how to affordably deliver this data according to the FAIR Data Principles.

## **Keywords**

FAIR Data, FAIR Guiding Principles, USDOI Data, DOI Data, DOI FAIR Data, USDOI FAIR Data

## **Introduction**

United States (US) Department of the Interior (DOI) bureaus have been attempting to implement FAIR (Findable, Accessible, Interoperable, Reusable) Principles into their data sharing strategies since the FAIR Guiding Principles for scientific data management and stewardship Data Principles (FAIR Principles) were published in 2016. The Federal Data Strategy is an attempt to leverage data as a strategic asset in order to create a data strategy and infrastructure for the future (Federal Data Strategy. 2022). Government-produced data are consumed by thousands of scientists, researchers, industries, and students around the world daily, but it is often difficult to locate those data, or one learns that the same data are often collected in a duplicative state at varying levels of quality. When data are shared, they often lack the proper artifacts to help the users comprehend what they are receiving from the bureau. Differing interpretations of the FAIR Principles and varying levels of implementation are leading to uneven applications and leading to data that are documented to different standards and not properly integrated for reuse. In order to properly understand how the FAIR Principles can best be applied, it is important to research across DOI bureaus how FAIR Principles are being utilized with data dissemination to the public, and in the metadata that are produced and uploaded into metadata catalogs such as data.gov (Ding et al. 2011). The needs of the data users should be used to craft what data providers are serving to the data users and consumers. DOI bureaus should study and learn what users need and how the FAIR Principles can properly be used to share higher quality data with suitable artifacts for public consumption. The lack of application of FAIR Principles is a societal data problem, as data that are not FAIR are often difficult or impossible to find, are often not available for download in their entirety, lack clear data dictionaries and/or ontology to be integrated with other data, and either have restrictions for reuse or are not at all fit for reuse (European Commission Expert Group on FAIR Data. 2018).

## **Literature Review**

Good data management is not a goal, but a conduit that leads to the discovery of knowledge and innovation (Wilkinson et al. 2016). Poor data management renders much of our data unfindable, due to missing or poor metadata, and results in data of such a low quality that renders them nearly useless. To address these

issues, the FAIR Principles were published in ‘Scientific Data’ in 2016 (Wilkinson et al. 2016). FAIR Principles are guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets. The ultimate goal of the FAIR Principles is to optimize the reuse of data (University of Edinburgh. 2020). It is essential for data science and analytics that data follow FAIR Principles to optimize the use of data by improving data accessibility, quality, and promoting the reuse of data. It is impossible to anticipate how your datasets will be used, or reused, in the future so it is imperative that your data be of the highest possible quality to ensure they are usable for future applications. Better application of the FAIR Principles will likely reduce data littering, that is the result of data that are created and distributed with missing or inadequate metadata that renders the data unusable for future applications and is often referred to as “stingy metadata” (Talbur. 2022). Findable data have metadata and are machine-readable, so they can easily be found by both humans and computers. Accessible data ensure metadata are available even when data cannot be found, and a protocol exists for an authentication and authorization procedure. Interoperable data allow for integration with other data, interoperating with applications and workflows for storage, analysis, and processing. Reusable data are the ultimate goal of the FAIR Principles and to optimize the reuse of the data, metadata should be well-described to be replicated in different settings (Wilkinson et al. 2019). One of the underlying precepts of FAIR relates to needing to include semantic understanding by grounding the data to well-known concepts so that others can readily interpret the meaning of the data.

FAIR Principles are designed to improve all types of data across all segments of our worldwide society, but moving from closed and silo-based approaches into an area of open and networked data sharing requires important changes in science reward and methodological practice (Boeckhout et al. 2018). Increased support is needed for FAIR data-publishing, analytics, computational capacity, virtual machines and workflow systems (Goble et al. 2020). The need for data reuse, the primary objective of FAIR, is so great that it is estimated that time spent by researchers performing “data munging” is 80%. If FAIR Principles were properly implemented, the wasted time could be reduced to a fraction of time spent with data reuse (Mons et al. 2017). The culture of the workforce needs to evolve for FAIR to fully succeed. The current mantra of ‘it’s my lab and my data’; should be ‘it’s the company’s (or the funder’s) data and it is FAIR’. This transformation, possibly incentivized by peer recognition or financial reward, needs total engagement throughout the organization, from the lab to the C-level staff (Wise et al. 2019). The US Federal Government is one of the largest data producers in the world, if not the largest (Brantley. 2018), and the proper application of FAIR Principles by the US Federal Government would vastly advance the availability of data for both humans and machines. A common misconception of data following the FAIR Principles is that FAIR data are all categorized as open access. However, data that are “closed off”, classified, uncontrolled classified information, or data that cannot be shared as open access can still be FAIR (Mons. 2020).

## Research Questions and Research Model

The objective of this study is to determine how DOI bureaus can produce and serve data that are better aligned with the FAIR Principles, and how to better align the data that are served with the needs and requirements of the data consumers. A baseline will be created of selected datasets that are currently shared with the public, using peer-reviewed tools for measuring FAIR Principle adherence. This study will help data leaders better understand and implement FAIR Principles and assess the current state of FAIR in the DOI. The initial focus area was to be across the entire Federal Government, but the sheer number of providers and datasets produced were determined to be too numerous to analyze for this study.

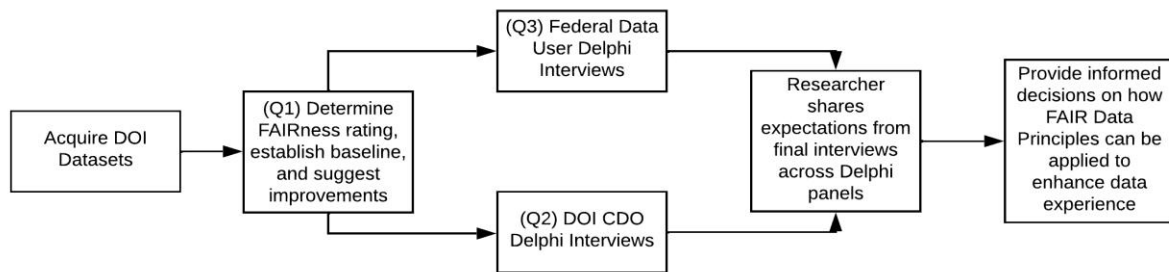
The problem arises in that DOI Chief Data Officers (CDO) are frequently reminded to serve data that are ‘FAIR’ and instructed to adhere to FAIR Principles when collecting and sharing Government data, but they have not been provided sufficient training or guidance in how to do so. Data users and consumers acquire Government data on a daily basis that covers a wide scope of compliance with FAIR Principles that should be optimized for reuse and is machine readable.

My research questions (Q) are:

- Q1.** To what extent are currently shared US DOI datasets following the FAIR Data Principles?
  - Q1.1.** What variability exists, and can it be standardized, with types of artifacts DOI bureaus are sharing with data regarding data dictionaries, field mappings, data structures and other related information?

**Q2.** What additional information is desired by consumers for shared DOI Federal data assets?

**Q3.** To what extent can specificity and consistency, regarding FAIRness, for sharing DOI Federal data assets can be reached by DOI Chief Data Officers without undue burden?



**Figure 1: Research Model**

Q1 will be addressed using 3 FAIR data analysis tools, assessing adherence to the FAIR Principles, for 2-3 datasets, if available, for each bureau in the DOI. Q1 will be answered prior to the completion of Q2 and Q3 to allow the results to be shared with the Delphi panelists. The three resources this study will use to evaluate datasets currently shared by DOI bureaus are: F-UJI, FAIR-Checker, and FAIR Evaluator. These resources have been developed to quantitatively assess compliance to the FAIR Principles and measure data FAIRness. These resources both qualitatively and quantitatively are used to calculate adherence to, and assess, 14 FAIR Metrics (with few exceptions). Those Metrics are (with references to FAIR Principles): 1. Identifier uniqueness (F1), 2. Identifier persistence (F1), 3. Structured metadata (F2), 4. Grounded metadata (F2), 5. Use of GUIDs in metadata (F3), 6. Metadata being indexed in a searchable resource (F4), 7. An open protocol for (meta) data retrieval (A1.1), 8. The protocol's support for authentication/authorization (A1.2), 9. Metadata persistence (A2), 10. The use of a knowledge representation language (loose) (I1), 11. The use of a knowledge representation language (strict) (I1), 12. The use of FAIR vocabularies (loose) (I2), 13. The use of FAIR vocabularies (strict) (I2), 14. Qualified outward links (I3), and 15. Metadata containing a link to a license (R1.1) (Wilkinson et al. 2019).

Q2 will be examined with the first Delphi panel that will be comprised of 10 or more DOI Federal data users who are suggested as regular data consumers by DOI CDOs. In the 4 questionnaire rounds for this group, topics to gather input on will include: Determining what compliance with FAIR Principles looks like, How are DOI Agencies complying with FAIR Principles, What are data needs that can better be addressed by FAIR data, How can DOI bureaus improve data quality and availability through better application of FAIR Principles, and are there any additional items that are not currently part of the FAIR Principles that the users feel would improve the FAIRness of data even more. Other suitable questions revolve around the 4 FAIR Principles resulting in inconsistent interpretations that risk incompatible implementations (Jacobsen. 2020). Q2 will be answered before Q3 is finished to better inform the DOI CDO Delphi panel.

Q3 will be examined by the second Delphi panel that will be comprised of 10 or more DOI CDOs and senior data leaders in the DOI. In the 4 questionnaire rounds for this group, topics to gather input from the CDOs are: How important are the FAIR Principles to the data you serve, Where do you see your compliance is now regarding FAIR Principles with data you serve or are within your bureau in 1, 3, or 5 years from now, What would assist you in implementing FAIR Principles in your bureau, What artifacts do you serve with data and how can those artifacts assist in making your data more compliant with what is expected from the FAIR Principles, and Can you meet the requirements of the FAIR Principles in data you serve without undue burden? For Q3, DOI CDOs will receive questions developed with input received from the Q2 answers of the data users and consumers.

## Methodology

The FAIR Principles are a working framework that can be used as a research framework for this study. These guiding principles make data findable, accessible, interoperable and reusable. They provide specific

guidance for scientific data management and stewardship and are relevant for all stakeholders in the current digital ecosystem. Data producers and publishers are directly addressed in this framework to promote the maximum use of research data (LIBER. 2020). The applicability of FAIR Principles will be assessed against existing shared datasets for DOI bureaus at data.gov using 3 FAIR analysis tools. Action research methods will be used to perform this analysis. This will provide a baseline for current adherence to FAIR Principles for publicly accessible data. Datasets that are shared by bureaus in the future can be compared to this baseline to assess overall improvement by bureaus for FAIR Principles using the Principles as a working framework.

In addition to the FAIR Principles framework, the Deming Voice of the Customer Framework is well-suited for capturing the input from DOI Government data producers and DOI Government data consumers (Willington. 2017). This Deming Framework supports the notion that a business needs to convince customers to ‘pay’ by doing a good job of focusing on the customers and their needs. Organizations that are poorly suited to providing value to customers tend to fail if they do not continually improve the value perceived by their customers. DOI Government bureaus serve a dichotomy of constituents, and those consumers of Government data are like the relationship of clients of a commercial business and their needs from a private entity. This Deming Voice of the Customer perspective goes beyond the notion that an organization is concerned only with customers. Instead, Deming proposes a complete management philosophy and management system to direct resources to continually improve the value provided to customers. The Deming Voice of the Customer Framework comes from the first of Deming’s 14 Points for Management, to ‘Create a constancy of purpose’. It is imperative to use evidence-based management practices to understand the value that customers receive and how that value varies in delivery (Hunter. 2013). It is a somewhat complicated process to determine the Voice of the Customer. This will best be obtained through Delphi panels via multiple rounds of questionnaires to reach consensus. Recent studies have found that Delphi panels have worked well to obtain consensus regarding data management and quality issues (Vogel. 2019). The two Delphi panels for this study will address the Voice of the Customer from the DOI CDOs and the DOI Federal data consumer groups.

## Conclusion

This study will measure how well DOI bureaus are, or are not, complying with the FAIR Principles with shared datasets and determine if better adherence to those Principles will enhance user experiences and enrich the FAIRness of future datasets. It will analyze datasets that are shared by DOI bureaus and artifacts shared with those data to determine how “usable” the data are by the data consumers. Input will be taken from data consumers (via their Delphi panel) that will be provided back to the data providers and producers (the DOI Chief Data Officer Delphi panel) to modify behavior in cases where the data producers can improve their data sharing following FAIR Principles. Special measures will be given to data producers/providers after that group has an opportunity to weigh in on if they can effectively meet changing users’ needs with their available and limited resources and missions. Since the FAIR Principles are encouraged but are not required, consideration will be given to that throughout the study in hopes to encourage their widespread adoption. It is anticipated that this study might also serve the purpose of educating some CDOs more about the FAIR Principles and their usefulness.

## REFERENCES

- Boeckhout, M., Zielhuis, G. A., and Bredenoord, A. L. 2018. “The Fair Guiding Principles for Data Stewardship: Fair Enough?,” *European Journal of Human Genetics* (26:7), pp. 931–936 (doi: 10.1038/s41431-018-0160-0).
- Brantley, B. 2018. The Value of Federal Government Data, Digital.gov, March 14 (available at <https://digital.gov/2018/03/14/data-briefing-value-federal-government-data/>; retrieved April 19, 2022).
- Ding, L., Lebo, T., Erickson, J. S., Difranzo, D., Williams, G. T., Li, X., Michaelis, J., Graves, A., Zheng, J. G., Shangquan, Z., Flores, J., McGuinness, D. L., and Hendler, J. A. 2011. “Twe Logd: A Portal for Linked Open Government Data Ecosystems,” *Web Semant.* (9:3), pp. 325–333.
- European Commission Expert Group on FAIR Data. (2018). “Turning FAIR Into Reality - European Commission,” *Turning FAIR Into Reality*, Directorate General for Research and Innovation (available

- at [https://ec.europa.eu/info/sites/default/files/turning\\_fair\\_into\\_reality\\_1.pdf](https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf); retrieved April 21, 2022).
- Federal Data Strategy, 2022. General Services Administration (available at <https://strategy.data.gov/>; retrieved April 16, 2022).
- Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M. R., Peters, K., and Schober, D. 2020. "Fair computational workflows," MIT Press, MIT Press, January 1 (available at <https://direct.mit.edu/dint/article/2/1-2/108/10003/FAIR-Computational-Workflows>; retrieved April 23, 2022).
- Hunter, J. 2013. "Customer Focus with a Deming Perspective," The W. Edwards Deming Institute, November 13 (available at <https://deming.org/customer-focus-with-a-deming-perspective/>; retrieved February 21, 2022).
- Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., Kaliyaperumal, R., Kersloot, M. G., Kirkpatrick, C. R., Kuhn, T., Labastida, I., Magagna, B., McQuilton, P., Meyers, N., Montesanti, A., van Reisen, M., Rocca-Serra, P., Pergl, R., Sansone, S.-A., da Silva Santos, L. O. B., Schneider, J., Strawn, G., Thompson, M., Waagmeester, A., Weigel, T., Wilkinson, M. D., Willighagen, E. L., Wittenburg, P., Roos, M., Mons, B., and Schultes, E. 2020. "Fair Principles: Interpretations and Implementation Considerations," *Data Intelligence* (2:1-2), pp. 10-29.
- LIBER. 2020. Implementing FAIR Data Principles: The Role of Libraries, Association of European Research Libraries, September (available at <https://libereurope.eu/wp-content/uploads/2020/09/LIBER-FAIR-Data.pdf>; retrieved February 2, 2022).
- Mons, B., Schultes, E., Liu, F., and Jacobsen, A. 2020. "The FAIR Principles: First generation implementation choices and challenges," *Data Intelligence* (2:1-2), pp. 1-9 (doi: 10.1162/dint\_e\_00023).
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., and Wilkinson, M. D. 2017. "Cloudy, Increasingly Fair; Revisiting the Fair Data Guiding Principles for the European Open Science Cloud," *Information Services & Use* (37:1), pp. 49-56.
- Talbur, J. 2022. "Data speaks for itself: Data littering," *The Data Administration Newsletter*, TDAN, February 16 (available at <https://tdan.com/data-speaks-for-itself-data-littering/29122>; retrieved April 18, 2022).
- University of Edinburgh. 2020. "Research Data Support: Quick Guide 6." Quick Guide - FAIR DATA (v1.1) FINAL, University of Edinburgh, June (available at [https://www.ed.ac.uk/files/atoms/files/quick\\_guide\\_6\\_-\\_making\\_your\\_research\\_data\\_fair\\_v1.1\\_0.pdf](https://www.ed.ac.uk/files/atoms/files/quick_guide_6_-_making_your_research_data_fair_v1.1_0.pdf); retrieved April 20, 2022).
- Vogel, C., Zwolinsky, S., Griffiths, C., Hobbs, M., Henderson, E., & Wilkins, E. 2019. "A Delphi study to build consensus on the definition and use of big data in obesity research," *International journal of obesity* (43:12), p. 2573-2586.
- Wilkinson, M. D., Dumontier, M., Sansone, S. A., Bonino da Silva Santos, L. O., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M., and Schultes, E. 2019. "Evaluating Fair Maturity through a Scalable, Automated, Community-Governed Framework," *Sci Data* (6:1), p. 174.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., t Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. 2016. "The Fair Guiding Principles for Scientific Data Management and Stewardship," *Sci Data* (3), p. 160018.
- Willington, L. (2017). "The Voice of the Customer and the Voice of the Process," Deming Alliance - The Voice of the Customer and the Voice of the Process (available at <https://demingalliance.org/resources/articles/the-voice-of-the-customer-and-the-voice-of-the-process>; retrieved February 8, 2022).
- Wise, J., de Barron, A. G., Splendiani, A., Balali-Mood, B., Vasant, D., Little, E., Mellino, G., Harrow, I., Smith, I., Taubert, J., van Bochove, K., Romacker, M., Walgemoed, P., Jimenez, R. C., Winnenburg, R., Plasterer, T., Gupta, V., and Hedley, V. 2019. "Implementation and Relevance of Fair Data Principles in Biopharmaceutical R&D," *Drug Discov Today* (24:4), pp. 933-938.