

# **The Data Life Aquatic: Oceanographers' Experience with Interoperability and Re-usability: Oceanographers' Experience with Interoperability and Re-usability**

Bradley Wade Bishop  
School of Information Sciences  
University of Tennessee

Carolyn Hank  
School of Information Sciences  
University of Tennessee

Joel Webster  
Department of History  
Texas Tech University

## **Abstract**

This paper assesses data consumers' perspectives on the interoperable and re-usable aspects of the FAIR Data Principles. Taking a domain-specific informatics approach, ten oceanographers were asked to think of a recent search for data and describe their process of discovery, evaluation, and use. The interview schedule, derived from the FAIR Data Principles, included questions about the interoperability and re-usability of data. Through this critical incident technique, findings on data interoperability and re-usability give data curators valuable insights into how real-world users access, evaluate, and use data. Results from this study show that oceanographers utilize tools that make re-use simple, with interoperability seamless within the systems used. The processes employed by oceanographers present a good baseline for other domains adopting the FAIR Data Principles.

*Submitted* 21December 2018 ~ *Revision received* 2 September 2021 ~ *Accepted* 2 September 2021

Correspondence should be addressed to Bradley Wade Bishop, 1345 Circle Park Dr. Room 454 Communications Bldg., Knoxville, TN 37996 USA. Email: [wade.bishop@utk.edu](mailto:wade.bishop@utk.edu)

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details, please see <https://creativecommons.org/licenses/by/4.0/>



## Introduction

The FAIR Data Principles infer discovery and re-use, but much of the research so far on the principles focuses on data itself, the systems where data are available, data management policies that enable access, and the perceptions and practices of data producers and curators. Lacking is an understanding of re-use and re-users, necessitating further study to evaluate human behaviour in data discovery and assessment through actual, real-world re-use cases. This article fills a gap by studying the perspectives of data re-users' experiences in determining data interoperability and re-usability. In the years since the Future of Research Communication and e-Scholarship (FORCE11) created the FAIR Data Principles as "a set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable," the acronym has been repurposed and transposed in many ways (Wilkinson et al., 2016). The original Data Principles reverberated a familiar refrain found in other data quality assessment work (e.g., Joint Declaration of Data Citation Principles (2014) and the Data Seal of Approval (2010)). Still, these efforts of data assessment stem from the curator perspective and often focus on assessment of digital repositories and not individual datasets. In particular, defining, measuring, and assessing interoperability and re-usability from either the data producers' or data curators' perspectives does not necessarily reflect or consider any re-users' perspectives. Whether the re-user will be human or machine, the judge for data's re-usability will ultimately be that re-user conducting the analysis, making discoveries, and generating new data. Re-users' needs and behaviours are paramount to any evaluation of the FAIR Data Principles. The principal goal of re-use in the FAIR Data Principles should not be lost in the sea of additional standards and metrics being reimagined to fit this simple acronym, or drowned in the multitude of reinterpretations of old information services now surfacing or sunk in the revamping of existing cyberinfrastructure and ecosystems, without the key driver, which is machine-actionable data (European Commission, 2018).

This paper addresses the literature gap by studying the perspectives of human data re-users' actions in determining data interoperability and re-usability. Using a critical incident technique, ten oceanographers were asked to describe their most recent search for data. A questionnaire derived from the FAIR Data Principles framed participants' behaviours in search and re-use along a sequence of actions, which included finding, accessing, making interoperable, and re-using data (Bishop & Hank, 2018). Through the participants' description of their process of discovery, evaluation, and use, the responses provide insight into the interoperability and re-usability criteria of the FAIR Data Principles. Oceanographers' perspectives on data interoperability and re-usability provide data curators and other key stakeholders valuable feedback into how real-world scientists access and re-use data successfully without any knowledge or acknowledgement of any data principles which, unseen and underappreciated, do undergird the entire research enterprise and make re-use possible.

## Literature Review

Data curatorial work enables findable, accessible, interoperable and re-usable data via the often-unseen labour that *precedes* and *proceeds* efforts of the data creators and data re-users (Bishop, & Grubestic, 2016). "Sharing is at the heart of success, as collecting, storing and making use of data can only come after the means of sharing are in place" (Cragin *et al.*, 2010, p. 4023). The fifteen FAIR Data Principles, as shown in [Table 1](#), represent succinct, discipline-neutral, broadly-applicable guidelines to use in a variety of fields to facilitate discovery and the evaluation of data by machines (Wilkinson *et al.*, 2016). FORCE11 (and others) bring these data principles forward with urgency due to *human* limitations in data processing. Many fields of science heavily rely on machine-actionable data. Machines not only process data, but also, with machine learning,

make discoveries humans cannot find. The variety and volume of contemporary scientific data and the increased use of machine-learning and artificial intelligence informed the development of the FAIR Data Principles. These principles have already been applied in the life sciences, including biology, environmental sciences, and other data-intensive sciences (Wolstencroft *et al.*, 2017; Rodríguez-Iglesias *et al.*, 2016; Diepenbroek *et al.*, 2017).

To be findable:	F1. (meta)data are assigned a <u>globally unique and eternally persistent identifier</u> .
	F2. Data are described with <u>rich metadata</u> .
	F3. (meta)data are <u>registered or indexed in a searchable resource</u> .
	F4. Metadata <u>specify</u> the data identifier.
To be accessible:	A1. (meta)data are retrievable by their identifier using a standardized communications protocol.
	A1.1 the protocol is open, free, and universally implementable.
	A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
	A2 metadata are accessible, even when the data are no longer available.
To be interoperable:	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
	I2. (meta)data use vocabularies that follow FAIR principles.
	I3. (meta)data include qualified references to other (meta)data.
To be reusable:	R1. Meta(data) have a plurality of accurate and relevant attributes.
	R1.1. (meta)data are released with a clear and accessible data usage license.
	R1.2. (meta)data are associated with their provenance.
	R1.3. (meta)data meet domain-relevant community standards.

Table 1: FAIR Data Principles

## Background and context

Several of the original FAIR Data Principle paper authors quickly formed a FAIR Metric group to evaluate claims in response to many repositories and resources that they were already “FAIR.” This FAIR Metric group conducted focus groups to assess whether their metric guidelines address their principles. Finding that not every metric, or even FAIR Data Principle, were always understood how intended, the group published a response to the original Data Principle paper with additional metrics to clarify the principles (Wilkinson *et al.*, 2018). Additionally, initiatives like GOfAIR<sup>1</sup>, the Enabling FAIR Data Project<sup>2</sup> and others took the overall FAIR framework and translated the original principles to serve their own purposes to improve data sharing. The FAIR acronym took off as it was easily understood, translatable, and the inherent meaning of the word, “fair,” is aspirational. The concept of data sharing to advance science by permitting others to verify results, replicate experiments, and lead to new application and discovery through data re-use, are long-standing goals of all open data movements, and are not entirely harmonized with the FAIR Data Principles (Pryor, 2012).

Scientific endeavours are typically expensive, and in the case of oceanography, data collection is often in real-time, collected once in a snapshot or streaming continuously through sensors, and across broad geographic areas. When one drops anchor, so to speak, to gather seafloor sediment data, the fathoming out of what is down below only occurs when the researchers are inland and connected to high and dry performance computing. The data analysed by the oceanographers that participated in this study are doubly valuable. This is due to both the implicit value of the scientific enterprise itself (e.g., contribute to a better understanding of our oceans), but also the value of knowing the contents of the exclusive economic zone (EEZ). The EEZ is each country’s jurisdiction of the seafloor and ownership of the natural resources beneath the oceans in which to manage, conserve, explore, and exploit (National Oceanographic and Atmospheric Administration, 2019). To be clear, the energy resources below the oceans are quite valuable. The investment of resources for any initial data collection and analysis for primary application is

<sup>1</sup> <https://www.go-fair.org/>

<sup>2</sup> <http://www.copdess.org/enabling-fair-data-project/>

intensive; hence, making sure such data is interoperable and reusable is vital, as secondary analysis and other re-use scenarios are a pragmatic extension of any initial data collection investment.

Regardless of the value of data, without human-driven curation, data does not live beyond its original purpose and its creator. Curators' increased understanding of how re-users determine interoperability and re-usability may contribute to the creation of key performance indicators for research data management services, as well as inform data products, tools, and information services. The aspects of the FAIR Data Principles explored in this paper, interoperability and re-use, likely have similar considerations across science data, beyond the subject matter (oceanography) of this study.

## Critical Incident Technique

Critical incident technique (CIT) provides an alternative to direct observations in defined situations to understand human behavior (Flanagan, 1954). This method is ideal because it addresses the limitations related to the costs of conducting direct observation in real-time, while allowing users to describe their thoughts during a search that are not present during direct observation. The method has been applied to a diverse range of disciplines, including education, medicine, social work, marketing, communication, and psychology (Butterfield *et al.*, 2005). Using this method in information-seeking behaviour research allows participants to recall a recent behaviour and describe it with detail. In information science, Tenopir and colleagues have conducted several studies on the information-seeking behaviours of academics' scholarly communication (Tenopir *et al.*, 2003; Tenopir *et al.*, 2009; Tenopir *et al.*, 2015). The most recent search should be the easiest for a participant to describe with some accuracy. Although a participant's most recent search or other behaviour may seem non-generalizable, since participants selectively choose their most recent search, it does introduce a degree of randomness across all participants.

## Methodology

This qualitative study adopted a semi-structured interview approach utilizing CIT. Along with questions concerning professional demographics, such as education and job duties, information-seeking behaviour questions were posed, derived from the FAIR Data Principles. The critical incident prompt simply asked participants to "think of a recent search for data." Again, although a recent search for data may not be representative of most searches, the participant should recall the details of the search more clearly than other, older searches, and these searches likely indicate a general search pattern for data across a community with enough participants. Since the FAIR Data Principles are written from the perspective of data curators to make data machine-actionable, to capture re-users' views of data required transposing the principles into questions for the interview schedule, presented in Table 2 (Bishop & Hank, 2018). Table 1, as provided in the preceding section, shows that some principles lend themselves to dichotomous questions (e.g., either the data had an element or not), but many others are more complex and require qualifiers and context to spur, capture, and assess re-users' explanations of their information-seeking behaviours. In the section on re-use, some questions were created outlining some known fitness for use facets specific to geospatial data re-use. The interview begins with some educational and occupational background questions to ensure the participants were trained and experienced in their domain.

Occupation and Education:	<ol style="list-style-type: none"> <li>1. What is your current job title?</li> <li>2. How many years in total have you been working in your current job?</li> <li>3. How many years in total have you been working with earth science data?</li> <li>4. Describe your work setting.</li> <li>5. Please indicate your credentials and degrees.</li> <li>6. Please provide any other education or training you have received that is applicable to performing your job.</li> </ol>
Interoperability:	<ol style="list-style-type: none"> <li>7. Was the data in a useable format?</li> <li>8. How was the data encoded and was it using encoding common to other data used in your research (i.e., same format)?</li> <li>9. Was the data using shared controlled vocabularies, data dictionaries, and/or other common ontologies?</li> <li>10. Was the data machine-actionable (e.g., to be processed without humans)?</li> </ol>
Re-usability:	<ol style="list-style-type: none"> <li>11. Were there any issues with the data that impacted re-use of the data (e.g., resolution)?</li> <li>12. Did the data geographic scale or resolutions impact re-use of the data?</li> <li>13. Did the coordinate systems used impact re-use of the data?</li> <li>14. Did the metadata provide sufficient information for data re-use?</li> </ol>

Table 2: Interview Schedule

Participants were recruited through purposive sampling, contacting known re-users of data in the Coastal and Marine Geology Program at two U.S. Geological Survey (USGS) Coastal and Marine Science Centers: Pacific Coastal and Marine Science Center,<sup>3</sup> and Woods Hole.<sup>4</sup> Ultimately, ten phone interviews with oceanographers were conducted over a three-month period. The interviews were recorded, transcribed, and analysed using NVivo. Grounded theory application of open, axial, and selective coding generated categories and broad themes concerning interoperability and re-usability. Occupational and educational data were also analysed, providing some demographic insights into this small sample of data re-users. The data are open and available through the Tennessee Research and Creative Exchange (TRACE) (Bishop, 2020).

## Findings

The findings describe the job analyses, interoperability, and re-usability responses from the interviews. The ten participants were consistent and similarities in responses emerged that signalled a saturation in data collection, which means additional interviews would not likely provide more variation in behaviours and perceptions.

### Job Analyses

For most participants, their job title was reflective of their field of study. Half (n=5) identified their job title as Oceanographer or Research Oceanographer. Two were Geologists—with a specialization of the seafloor. For the remaining participants, one title reflects managerial responsibilities (Deputy Regional Manager), while two appear to be more reflective of data management specific roles: Scientific Programmer (n=1) and a Metadata Management Architect (n=1).

<sup>3</sup> <https://walrus.wr.usgs.gov/>

<sup>4</sup> <http://www.whoi.edu/>

The average years spent working with science data, including all time in higher education, was almost 22 years. Participants' time in their current positions varied from 2.5 years to 30 years, with about 13 years being the average. These participants' expertise in locating science data through changes in data formats and information systems was apparent in their responses and detailed descriptions of how they locate and evaluate data.

Participants mentioned their most advanced degree when asked about their education. Six of the ten hold PhDs, with the remainder having master's degrees. All but one were in the sciences; the outlier held a Master's in Art. The degrees demonstrate a high level of education among participants, and an assumed understanding of data in their fields. Although participants were all asked about additional training they received to do their jobs, only two gave specific examples: MATLAB and ESRI workshops.<sup>5</sup> Most participants indicated they were self-taught, and nearly all mentioned gaining new skills and knowledge from self-directed searches and learning online (e.g., YouTube videos). The lack of formal training in data science and data curation for these participants whose primary responsibilities are related to those tasks is a challenge found in many domains.

When asked about their respective "work settings," a few participants referred to field work, such as boats, cameras, and scuba gear. The majority referred to the hardware and software used to analyse science data. Participants referred to their hardware as "heavy duty processing machines" and all types of computers, from laptops to clusters that access high performance computing to the cloud (e.g., Amazon Web Services), to conduct simulations and run models. The most mentioned "tools" were MATLAB, Python, and ArcGIS.

## Interoperability

The interoperability questions focus on the data itself. Eight participants indicated that the data they located was in a useable format. Still, two were uncertain of the useable format of their data and described some issues transposing what they could find and access into a useable format. "We've obtained is a NetCDF 3, let's say. So, we have to use an internal software, which is free, and it mods it to NetCDF 4 and that's about it. And sometimes you convert that into a .mat, which is the MATLAB format." These additional steps to make data interoperable are logical models that could be built into machines to make data in similar transposable formats machine-actionable. The other participant with useable format issues was working with data presented in PDF, requiring considerable transformations to use, but admitted that this was an atypical case for data re-use in their work (even though it was the most recent data search incident). The data for this atypical re-use case was a bathymetry (ocean floor) sheet map, so that data was not actually encoded in something that could be easily made interoperable. Thus, much of the legacy data in oceanography will still require humans to transpose this invaluable dated data.

Nine of the participants indicated that the data they found and used was in a common encoding standard. Five indicated the encoding was NetCDF, two others cited text files, one used .mat, one GRIB 2 (i.e., gridded bathymetry), and one indicated a Shapefile. Again, in reference to a data portal, one participant said the data portal provided the data in any format they might need, serving up automatic translations, and was not specific to any particular format. This customizable system that serves up data in multiple common encodings solves many of the potential interoperability issues faced in re-use of data. "You have similar options for when you download these, you can bring it as text, as a list, CSV, or Excel, or whatever you want." The participant that did not know if the data was in a common encoding was working with a sheet map in .pdf. Since this is a very common encoding for digital objects, there is a chance the participant did not understand the question about common encoding.

Seven participants indicated controlled vocabularies, data dictionaries, and/or common ontologies were used, but three others did not know how values in their data were categorized. The Global Change Master Directory (GCMD) keywords and Southern California Coastal Water Research Project (SCCWRP) were named specifically, but marine sciences with fewer political boundaries and fewer variables does have well-established metadata standards to the

<sup>5</sup> Environmental Systems Research Institute <https://www.esri.com>

point of invisibility to end users. For the participants that did not know if they were using a controlled vocabulary, it seemed as if their interpretation of using keywords from a thesaurus was in fact a controlled vocabulary and the terminology of the question should be revised for each discipline.

The responses also varied for machine-actionable, as some data are not ready for processing without some human intervention. Seven were confident the data they re-used was machine-actionable, but three had issues that would be a barrier for machine-to-machine re-use.

## Re-usability

Each search and data re-use scenario discussed by participants presents unique re-usability issues depending on participants' specific re-use purposes. All data are imperfect, especially when using data for purposes beyond its original collection and intended primary use. As such, the imprecision, imperfect granularity, and versioning proliferation forced these participants to accept and work with these challenges to re-use because the data are unique, valuable, and alternatives do not abound.

For the question on issues with the data that influenced re-use, a few known issues emerged. Although three participants indicated no issues with re-use, seven had challenges. The most common data issue is the lack of version controls. If errors occur in data collection (e.g., buoy drift), the data is only issued in a corrected version overwriting the old data. "Sometimes the data does change from month to month; in other words, the data you extract one month might be different the next month and we don't always know when that remote dataset has been changed or why." This re-use issue of lack of documentation in large datasets is known, but no easy solution exists. A similar complaint raised by one participant is missing data without explanation. Certainly, instruments fail in the field and missing data is overcome in aggregation, but it is a known issue.

Other data re-use issues mentioned include the precision of data and licensing. Prior to GPS much legacy data has limitations because a reduced accuracy in navigation results in a lack of precision of that data. Humans may understand this better than machines without the context to introduce doubt into oceanographic data prior to the mid-1990s. Along the same line, humans face an issue with license agreements in a different approach than is possible for machines. A human may contact a vendor and clarify licensing issues that may be unclear, where a machine would not process data with licensing issues at all. In the U.S., the various agreements between federal agencies, diffusion of proprietary formats, and grey areas in data sharing present challenges for any re-use but are particularly rough for machines to sail through. "Like I look at shorelines, so I can share those shorelines that I generated from it, but actually sharing the imagery, I can't do that." Finally, a participant explained that many more datasets that are needed for analyses must become accessible through web services. There are similar web services that exist for other data, so building out additional services is possible.

Seven of the participants indicated that geographic scale did not impact re-use of the data. This is in part due to a need for data points to match precise locations for accurate analysis, but when research involves a larger area with more data points there is a higher degree of acceptable imprecision for re-use. "You know gauge data is very site specific, so it doesn't really, you can't really extrapolate over a larger spatial area, so it is limited to a point in space." The three participants that did indicate scale mattered were working with larger extents of the ocean. Similarly, eight participants also did not think the coordinate system influenced re-use. Like scale, coordinate systems can be easily converted to other coordinate systems, meaning data collected in various ways can all be re-projected into the same system for analysis. At least one participant said, "we kind of just ignored that. Because many times we couldn't tell what coordinate systems was used." Scale and coordinate systems do influence findings as human and machine re-users may overlook the importance of these details and unknowingly alter findings. For illustration, a few hundred meters in the wrong direction along shorelines could be a big mistake. Two participants understood these data issues related to coordinate systems and

mentioned that the coordinate system used does matter most for instruments that move across the ocean, like buoys and boats, but matter less for instruments that are more static.

Finally, nine of the ten participants said that the metadata provided sufficient information for data re-use. “I mean metadata is harder than data itself sometimes.” One participant was more elaborative, stating “that for the last 20-25% of the assessment ... the metadata might not be enough, you might have to actually get it and try it out yourself or send a query and ask somebody.” This point is key because a machine cannot ask someone for help and question data beyond what it is programmed to do. The education and experience of human experts allows them to assess data quality while using it, but what aspects of human experts’ tactics could be built into machine-actionable quality control? The participant who said that the metadata did not provide sufficient information for data re-use cited this aspect: “in general, there’s always some problems with the metadata, when you think about it from a machine-to-machine perspective.” The perspective of sufficient metadata is likely different for a machine than a human expert user who already has been programmed with all kinds of training experience and knowledge that a machine might not have.

## Conclusion

Investigations of the practices, perceptions, and preferences of digital data producers and data curators have been ongoing for well over a decade now. The results of such work are evident in the FAIR Data Principles, and many other data-curation tools, services, and guidance documentation inspired by the FAIR tsunami. However, attention to the practices, preferences, and perceptions of re-users is lacking and this study of human re-users fills that literature gap and highlights human strengths and machine weaknesses in the push toward machine-actionable data.

Typically, reporting on re-users and re-use scenarios is hypothetical and abstract, rather than resulting from actual, direct data collected from real-world re-users or re-use cases. There are practical reasons for this discrepancy. Paramount, data creation and curation practices and implications need to be well understood in order for discovery and access to get underway. Still, interoperability and re-use issues await downstream from early stages in any data lifecycle. Regardless of this and other factors, the research literature provides treatment that is more extensive from data creator and data curator perspectives than the data re-user. This study is intended to contribute to this vital though less frequently investigated group of data curation stakeholders, and in the domain of oceanography specifically. In conclusion, despite a big push to enable machine-actionable data and many successes with machine learning and artificial intelligence in data analyses, certain aspects of preparing, processing, and legally sharing data require human input to see beyond the current arc of visibility.

## Acknowledgements

The authors would like to acknowledge support from the Gloria and Dave Sharrar Faculty Research Fund 2017, administered by the School of Information Sciences, University of Tennessee.

## References

Bishop, B. W., & Grubestic, T. H. (2016). *Geographic information: Organization, access, and use*. New York: Springer.



- Bishop, B. W. & Hank, C. F. (2018). Measuring FAIR principles to inform fitness for use. *International Journal of Digital Curation*, 13(1). doi: 10.2218/ijdc.v13i1.630
- Bishop, B. W. (2020). Data Fitness for Use Interviews. Knoxville, University of Tennessee, Tennessee Research and Creative Exchange (TRACE). doi: 10.7290/ai1y12d
- Butterfield, L., Borgen, W., Amundson, N., & Maglio, A. (2005). Fifty years of the critical incident technique: 1954-2004 and beyond. *Qualitative Research*, 5(4), 475-497.
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023-4038. doi:10.3138/7146-4332-6J78-0671
- Data Seal of Approval. (2010). Quality guidelines for digital research data. Retrieved from [https://www.datasealofapproval.org/media/filer\\_public/2013/09/27/dsa-booklet\\_1\\_june2010.pdf](https://www.datasealofapproval.org/media/filer_public/2013/09/27/dsa-booklet_1_june2010.pdf)
- Diepenbroek, M., Schindler, U., Huber, R., Pesant, S., Stocker, M., Felden, J., Buss, M., & Weinrebe, M. (2017). Terminology supported archiving and publication of environmental science data in PANGAEA. *Journal of Biotechnology*, 261(Supplement C), 177-186. doi: 10.1016/j.jbiotec.2017.07.016
- European Commission. (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. European Commission Expert Group on FAIR Data. <https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1>
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358. doi: 10.1037/h0061470
- COPDESS. (n.d.). Enabling FAIR data project, commitment statement in the earth, space, and environmental sciences. Retrieved from <http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/>
- GO FAIR. (n.d.). FAIRification process. Retrieved from <https://www.go-fair.org/technology/fairification-process/>
- Joint Declaration of Data Citation Principles. (2014). *Information Standards Quarterly*, 26(1), 28.
- National Oceanographic and Atmospheric Administration. (2019). What is the EZZ? National Ocean Service Website. Retrieved from <https://oceanservice.noaa.gov/facts/eez.html>
- Pryor, G. (2012). Why manage research data? In G. Pryor (Ed.), *Managing research data* (pp. 1-16). London, UK: Facet Publishing.
- Rodríguez-Iglesias, A., Rodríguez-González, A., Irvine, A. G., Sesma, A., Urban, M., Hammond-Kosack, K. E., & Wilkinson, M. D. (2016). Publishing FAIR data: An exemplar methodology utilizing PHI-Base. *Frontiers in Plant Science*, 7(641). doi: 10.3389/fpls.2016.00641
- Tenopir, C. & W. King, D., Boyce, P., Grayson, M., Zhang, Y., & Ebuén, M. (2003). Patterns of journal use by scientists through three evolutionary phases. *D-Lib Magazine*, 9(5). doi: 10.1045/may2003-king.

- Tenopir, C., King, D. W., Spencer, J., & Wu, L. (2009). Variations in article seeking and reading patterns of academics: What makes a difference? *Library and Information Science Research*, 31(3), 139-148.
- Tenopir, C., King, D., Christian, L., & Volentine, R. (2015). Scholarly article seeking, reading, and use: A continuing evolution from print to electronic in the sciences and social sciences. *Learned Publishing*, 28(2), 93-105.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., and Baak, A. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(160018), doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)
- Wilkinson, M. D., Sansone, S., Schultes, E., Doorn, P., da Silva Santos, L. O. B., & Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. *Scientific Data*, 5(180118). doi: [10.1038/sdata.2018.118](https://doi.org/10.1038/sdata.2018.118)
- Wolstencroft, K., Krebs, O., Snoep, J. L., Stanford, N. J., Bacall, F., Golebiewski, M., ... Goble, C. (2017). FAIRDOMHub: A repository and collaboration environment for sharing systems biology research. *Nucleic Acids Research*, 45(D1), D404-D407. doi: [10.1093/nar/gkw1032](https://doi.org/10.1093/nar/gkw1032)