

Measuring FAIR Principles to Inform Fitness for Use

Bradley Wade Bishop
School of Information Sciences
University of Tennessee

Carolyn Hank
School of Information Sciences
University of Tennessee

Abstract

For open science to flourish, data and any related digital outputs should be discoverable and re-usable by a variety of potential consumers. The recent FAIR Data Principles produced by the Future of Research Communication and e-Scholarship (FORCE11) collective provide a compilation of considerations for making data findable, accessible, interoperable, and re-usable. The principles serve as guideposts to ‘good’ data management and stewardship for data and/or metadata. On a conceptual level, the principles codify best practices that managers and stewards would find agreement with, exist in other data quality metrics, and already implement. This paper reports on a secondary purpose of the principles: to inform assessment of data’s FAIR-ness or, put another way, data’s fitness for use. Assessment of FAIR-ness likely requires more stratification across data types and among various consumer communities, as how data are found, accessed, interoperated, and re-used differs depending on types and purposes. This paper’s purpose is to present a method for qualitatively measuring the FAIR Data Principles through operationalizing findability, accessibility, interoperability, and re-usability from a re-user’s perspective. The findings may inform assessments that could also be used to develop situationally-relevant fitness for use frameworks.

Received 30 October 2017 ~ *Accepted* 20 February 2018

Correspondence should be addressed to Wade Bishop, 1345 Circle Park Dr. Room 442. Communications Building. Knoxville, TN 37996. Email: wade.bishop@utk.edu

An earlier version of this paper was presented at the 13th International Digital Curation Conference.

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 Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

The FAIR data principles as outlined by the Future of Research Communication and e-Scholarship (FORCE11) provide “a set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable” (Wilkinson et al., 2016). The four foundational principles apply to interlocking parts in the discovery process that precede and follow data creation, such as algorithms, discovery tools, workflows, information-seeking behaviour and other components that appear sequentially and throughout the data lifecycle. FAIR represents a concise, domain-independent, high-level set of data principles that may be applicable in a number of areas and cater to answering the questions both humans and machines will have while discovering and evaluating data prior to use (Wilkinson et al., 2016). FORCE11 points out the limitations of relying on humans alone to process data due to its expanding scope, growing scale, and quickening rate of creation. Rightfully, data must be machine-actionable given the complexity of contemporary scientific data. The utility, versatility, and charm of the FAIR acronym help explain its popularity and application in a variety of fields including biology, life science, plant science, environmental science, and other data-intensive sciences (Wolstencroft et al., 2017; Wilkinson et al., 2017; Rodríguez-Iglesias et al., 2016; Diepenbroek et al., 2017).

In the FAIR framework, the coupling of metadata and data into (meta)data makes clear that the principles apply to both data and metadata. To contextualize the discussion, Table 1 summarizes the FAIR Data Principles.

Table 1. FAIR Data Principles.

Principle	Requirements
To be findable	<p>F1. (meta)data are assigned a globally unique and eternally persistent identifier.</p> <p>F2. data are described with rich metadata.</p> <p>F3. (meta)data are registered or indexed in a searchable resource.</p> <p>F4. metadata specify the data identifier.</p>
To be accessible	<p>F1. (meta)data are assigned a globally unique and eternally persistent identifier.</p> <p>F2. data are described with rich metadata.</p> <p>F3. (meta)data are registered or indexed in a searchable resource.</p> <p>F4. metadata specify the data identifier.</p>
To be interoperable	<p>I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.</p> <p>I2. (meta)data use vocabularies that follow FAIR principles.</p> <p>I3. (meta)data include qualified references to other (meta)data.</p>
To be reusable	<p>R1. (meta)data have a plurality of accurate and relevant attributes.</p> <p>R1.1 (meta)data are released with a clear and accessible data usage licence</p> <p>R1.2 (meta)data are associated with their provenance.</p> <p>R1.3 (meta)data meet domain-relevant community standards.</p>

The purpose of this paper is to report on an active study that explores how to measure the FAIR principles for re-use of data. This requires operationalizing findability, accessibility, interoperability, and re-usability from a consumer's perspective. Although these principles derive from many different questions that require answers to determine fitness for use for any re-user, this study also illuminates issues that may arise when assessing some parts of the FAIR principles for machines as well. In short, some principles are more easily measured than others, particularly from a system (machine-centered), rather than user (human-centered) orientation. Assessment in practice from a re-user/consumer perspective will require particular metrics.

The principles intuitively harken back to other work in the area of assessing data quality, such as the Joint Declaration of Data Citation Principles (JDDCP)¹ and the Data Seal of Approval (DSA)². Certainly, the fitness for use determinations of any user to inform the assessment of FAIR-ness could also be useful for assessing those other data quality metrics.

The intent of the study is to propose and vet a qualitative interview method for operationalizing and assessing FAIR principles from a fitness for use approach. As such, a variety of existing intake questionnaires and submission agreements to facilitate curators' receipt of data were reviewed. Indeed, several of the FAIR principles are quantitatively or qualitatively captured during ingest-aligned activities, such as data formats, provenance, workflows, and software used throughout the data lifecycle, from creation to preservation. These may be machine-generated or derived from submission agreement negotiations between data producers and data curators or the data archive. While some FAIR principles are observable and lend themselves to automation through objective measurements, others require more qualitative, subjective measures. Further, the perspectives of the data producers and data curators differ from data re-users/consumers. Therefore, additional considerations must be made to determine FAIR-ness beyond data management concerns. Several discovery tools could benefit from informed design considerations based upon actual information-seeking behaviour of data re-users. In this study, the focus is on humans; however, machines will likely have similar information needs if not behaviour.

Measuring re-usability and some of the other FAIR principles from solely the data themselves is not plausible. The potential for re-use goes beyond being findable, accessible, and interoperable. This paper presents recommendations to create context dependent questionnaires to assess the FAIR principles in a way that captures the FAIR-ness from data re-user/consumer perspectives, presenting a proposed set of questions representing each letter of the FAIR acronym. Similar work to assess FAIR from the perspectives of data producers, managers, and curators would also be beneficial and likely differ from this approach. Luckily, the original FORCE11 authors have now provided their own exemplar metrics for machines that were the only perspective that FAIR-compliance was intended for as well as simplified into dichotomous requirements (Wilkinson et al., 2018).

1 Joint Declaration of Data Citation Principles: <https://www.force11.org/group/joint-declaration-data-citation-principles-final>

2 Data Seal of Approval: <https://www.datasealofapproval.org/en/>

Literature Review

Data should be discoverable and usable by a variety of potential re-users for open science to function. The future of science relies on the sharing and re-usability of data by humans, as well as computers through machine-actionable data. The elements listed in the FAIR principles codify efficiently the necessary details each data set needs to provide in the data itself or accompanying metadata to meet the needs of potential re-users for discovery and evaluation. Sharing data improves and advances science by permitting others to verify results; enabling the repetition of experiments; and leading to new research through data re-use (Pryor, 2012). The overarching principles will not necessarily directly translate into measuring data's FAIR-ness, and a better understanding of how re-users discover and evaluate data likely differs from other aspects that make data curatable and preservable. The concept of fitness for use summarizes this discovery and evaluation for the user's perspective. This paper assumes that assessing FAIR-ness from a re-user's orientation would be a goal in implementing FAIR principles, and a useful exercise.

The origin of the fitness for use concept comes from a 1951 private sector quality control book, now in its fifth edition (Juran, 2002). At its inception, the idea was simply that customer needs are met by the specifications of a product. Similar to a gap between data producers and data users, suppliers and customers can differ in their expectations of the quality for a product. The specifications for any product may be numerous and only some factors matter to the re-users/consumers of digital science data.

In 1984, Chrisman defined data fitness for use as “the foundation of data quality (is) to communicate information from the producer to a user so that the user can make an informed judgment on the fitness of the data for a particular use” (Chrisman, 1984). To determine suitability for a particular application or purpose, a user may need to know many details about the data, including data quality, scale, interoperability, cost, metadata, syntactic and semantic heterogeneity, and still others (Veregin, 1999). On the face of it, this seems reasonably simple with users asking producers how they collected the data or locating the needed specifications somewhere within or adjacent to where they found the data. In many observable instances, this process cannot be automated and may require cumbersome ancillary searches for data documentation. Additional searches may be necessary given that semantics vary between fields and within disciplines. Also, data from multiple sources may be collected at different time intervals and/or geospatial scales. An anecdotal search behaviour common in science is contacting the data producer and asking them directly, which requires the producer to be responsive to these types of requests. In addition, the technical expertise of users varies significantly, potentially diminishing the ability for many re-users to make informed evaluation for the limitations that also relate to fitness for use (Bishop and Grubestic, 2016). A World Data System (WDS) and Research Data Alliance (RDA) joint working group exists to create an assessment of quality criteria metric to measure FAIR-ness of the data themselves³; however, this paper presents some alternatives to assess the FAIR principles that may not be as conducive to automation and from the user's perspective.

³ WDS/RDA Assessment of Data Fitness for Use Working Group: <https://www.rd-alliance.org/groups/assessment-data-fitness-use>

Methodology

While this paper reports on an active study to design and vet a qualitative interview approach to assess FAIR-ness from a consumer perspective, several intake questionnaires already exist to prepare curators to receive data. For example, data curation profiles (DCP) allow data provenance information to be gathered and described the data in detail, including its intended and anticipated use, storage requirements, potential stakeholders and access restrictions (Carlson, 2010). DCPs provide science data a tool to capture the ‘data story’ at both individual research project and data aggregator levels, but these types of curation intake questionnaires focus on the data with input from the researchers that produced the data (Bishop and Hank, 2016). Indeed, several of the FAIR principles could be qualitatively captured in ingest data-gathering/validation activities, such as data formats, provenance, workflows, and software used throughout the data lifecycle from creation to disposition.

The following sections discuss the writing of potential questions to operationalize the FAIR principles in a way that qualitatively captures the FAIR-ness from the perspectives of re-users/consumers. The perspectives of the data producers captured through tools like the DCP do not necessarily account for findability, accessibility, and interoperability from a re-user’s view; it may be assumed rather than known. A data producer should meet all FAIR principles with their own data. A data curator likely must approach meeting all the FAIR principles to navigate within their own collections. Measuring re-usability with only input from data curators, data producers, or simply the data themselves is problematic.

The requisite initial step required of rewriting each FAIR principle into answerable questions was to transpose each principle into a meaningful element for a re-user/consumer of data. The following sections present a proposed set of questions for each letter of the FAIR acronym, and discuss the process and choice of these questions to potentially be used to measure the FAIR principles from a re-user’s perspective with the intention of one day automating them.

Findability

Some of the first FAIR principle items on findability are such that data managers can quickly assess them as being present or not, including: *F1. (meta)data are assigned a globally unique and eternally persistent identifier*. Given the dichotomy of this principle, a human or computer can determine if a persistent identifier is present or not. The FAIR-ness findability begins with the idea of data being uniquely identified and persistent at some virtual locality. A Digital Object Identifier (DOI) standardized in 2000 by the International Organization for Standardization is commonly used and sufficiently fulfils this FAIR aspect.

The second findability principle lacks the clarity of measurement that F1 presents in a compliance/non-compliance dichotomy: *F2. data are described with rich metadata*. Rich metadata presents a qualifier that may not be easily quantified. Rich could confuse some data producers and data re-users. Rich metadata is complex and may entail compliance with a metadata standard, but any data that has been found had a minimal amount of metadata to be located. If only findability is taken into account, then minimal metadata would be rich enough metadata for finding, even if it is only minimally useful for re-usability. After conferring with some of the original FAIR Data Principles

authors, rich meant ‘good enough’ to locate. Operationalizing the variable of rich requires additional qualitative considerations.

Much like F1, the F3 element, *(meta)data are registered or indexed in a searchable resource*, presents as dichotomous. Either the data and/or metadata are indexed in a searchable database or not. A great deal of the data that purports or aspires to be FAIR, at a minimum, should exist somewhere virtually. Finally, for findability, *F4. metadata specify the data identifier* could be automatically assessed on the single point of whether the data identifier is specified or not.

The following are proposed F1-F4 questions to assess findability from the re-user/consumer perspective, which may require revision depending on the data type and stakeholders.

1. How did you find the data?
2. Did the data have a persistent identifier?
3. Did the data have metadata?
4. Did the metadata help you locate the data?

Although the first question does not directly map to any of the FAIR principles, through explaining their information-seeking behaviour that helped them find the data, participants address aspects of principles F2 and F3. The second question allows for assessment of F1 and F4. The third question allows a data re-user/consumer to explain F2 in their own words. Participants telling their search story related to findability may actually explain the level of richness required to locate the data and what elements within metadata led them to the data found. Throughout the explanation of the information-seeking behaviour, F3 will be revealed. In addition, the final proposed question allows participants to determine the actual impact metadata played in the data’s findability. There is a chance many users locate data without fully acknowledging or understanding the role metadata plays unless they are also data producers or data curators.

Accessibility

The second principle of FAIR, accessibility, lends itself to dichotomous assessment like some of the findability principles. Access has a multitude of meanings, but within FAIR accessibility takes on the most literal meaning that data and/or metadata can either be accessed or not. The assumption built into this definition of accessibility is that the data and/or metadata could be downloaded and manipulated by end-users and not simply viewed. Further, data may be found, but not accessible. Therefore, this principle goes further to measure if found data can be accessed by re-users/consumers of that data.

The first accessibility principle builds upon the findability principle of having an identifier, and *to be able to use that identifier to retrieve the data using a standardized communications protocol (A1)*. The protocol that allows for retrieval has some additional considerations in the next two nested principles (A1.1 and A1.2). Regarding the former: *A1.1 the protocol is open, free, and universally implementable*. Although openness has its own variations globally, per FAIR, data, especially from a re-user’s perspective, are either open or not, and/or free or not. However, it is not clear what universally implementable may mean and documentation describing that part of the principle is lacking. Regarding the latter: *A1.2 the protocol allows for an authentication*

and authorization procedure, where necessary. Again, this could be interpreted as assessment of permissions allowing re-users/consumers to log in to access data. Finally, *whether or not the metadata itself is accessible is the final principle for this section (A2).*

The following are proposed A1-A2 questions to assess accessibility, which may require revision depending on the data type and stakeholders.

1. How did you access the data?
2. Was the data in an open format?
3. Was the data free?
4. Did the data have use constraints (e.g., limitations of use)?
5. Was the metadata accessible?

As with findability, a participant could more easily, and reliably, relay a story of how the data were accessed rather than provide details of unseen protocols that were necessary for the access to appear seamless. Therefore, assessment should allow a re-user to explain how the data were accessed. A1.1 and A1.2 may be addressed within a re-user's story; however, the dichotomous nature of open and free, as well as authentication and authorization, may be supported through quantitative automation to assess if these guidelines appear in the data or not. If data producers adhere to other guidelines that explain use constraints and limitations of use, then that piece of the assessment may also be automated. Still, many datasets do not detail use constraints in a formal or necessarily public way and, even if so, a re-user's understanding of use constraints might vary. Knowing more detail, as can be generated through these proposed questions, may help inform how to highlight particular limitations of use in a noticeable way. Finally, whether metadata were accessible or not also is simply binary. However, through questioning, participants may provide additional insights beyond a simple binary (yes/no) assessment. Possibly, for example, there were issues in understanding that impacted their agreement or disagreement. Issues such as these are also treated in FAIR's interoperability principles.

Interoperability

The third principle of FAIR, interoperability, may present the most challenging questions for participants, as these can be seen to require a more sophisticated understanding of disciplinary-specific languages and standards, including: *I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation; and I2. (meta)data use vocabularies that follow FAIR principles.* Although most data users should know the format, some data users may not be familiar with how the data they use are encoded. Further, this principle gets at the notion of dependencies between data/metadata for enhanced understanding to facilitate appropriate and effective evaluation and re-use, as captured in *I3. (meta)data include qualified references to other (meta)data.* A multitude of domain-specific ontologies exist, as well as controlled vocabularies; however, unless the re-users worked as data producers or data curators, they might not have an awareness (even if it is in the metadata) that these knowledge organization tools exist. In fact, without reading,

metadata re-users might make assumptions about data from different sources and unknowingly generate errors due to semantic heterogeneity.

The following are proposed I1-I3 questions to address interoperability, which may require revision depending on the data type and stakeholders.

1. Was the data in a useable format?
2. How was the data encoded?
3. Was the data encoded using encoding common to other data used in your research (i.e., syntactically heterogeneous; same format)?
4. Was the data using shared controlled vocabularies, data dictionaries, and/or other common ontologies (i.e., semantically heterogeneous)?
5. Was the data machine-actionable (e.g., to be processed without humans)?

As with other aspects of FAIR principles previously addressed, some of the elements within these interoperability principles can be system-generated and captured automatically, such as encoding and application of controlled vocabularies and data dictionaries. However, there are vital aspects to inform and assess re-use that are not conveniently or effectively automated. For example, in regard to question 1, relevance is a system-centered evaluation and users operate with all means of proprietary and open software and could be resourceful enough to make any format work. However, assessments on value and use are user-centric, and benefit from data collected via survey, including interview methods. This is a common technique in information seeking studies, regardless of format or genre. Responses can inform assessment across the three interoperability principles. Additionally, qualitative data captured as regards encoding, controlled vocabularies, data dictionaries and ontologies may provide indicators in perceived gaps between metadata provided within data archives, and re-users expectations or failure to consult metadata. It can also indicate a lack of understanding and awareness of these requisite aspects for value-added data, and inform approaches to better communicate the metadata's inherent value, to current as well as potential users. Finally, re-users often access and use data via tools for processing and analyses that require data to be machine-actionable. Of note, data may be machine-actionable that make it more re-usable, but not adhere to several of the FAIR principles that could make the data more fit for use.

Re-usability

The fourth and final FAIR principle is re-usability. Re-use without any context is not definable and the assumptions built into re-usability inherently contain usability itself. FAIR data principles have already been applied to several sciences, but each domain (and ultimately each data type) will present its own requirements for re-use. The first principle broadly captures that for data to be reusable in any area, that: *RI. meta(data) have a plurality of accurate and relevant attributes*. Context specific accuracy and relevance require more qualitative, subjective measures to assess. The second principle relates strongly with earlier accessibility principles in that the statement points to usage licensure: *RI.1. (meta)data are released with a clear and accessible data usage license*. The third principle requires data to capture provenance and this data lineage allows re-use to occur fully knowing how the data may have been created and transformed: *RI.2.*

(meta)data are associated with their provenance. Finally, R1.3. (meta)data meet domain-relevant community standards, could only be assessed given selection of some community with likely data re-users.

The following are proposed R1-R1.3 questions to address re-usability, which would require revision depending on the data type and stakeholders. Developing even draft questions requires choosing some swath of all data. For this paper, geospatial data was selected as its structure applies to a great deal of earth science data that must be shared to address grand challenges. Each data type would have different re-use considerations to inquire about.

1. Were there any issues with data quality that impacted re-use of the data?
2. Did the data geographic scale used impact reuse of the data?
3. Did the coordinate systems used impact reuse of the data?
4. Did the metadata provide sufficient information for data re-use?

Although the first question does not directly map to accuracy and relevance, the question allows a user to again provide a narrative of any data quality issues that impacted re-use of data. Geographic scale and coordinate systems are specific facets of geospatial data that are known to impact re-use; therefore, those questions were suggested. Other data types would require differing facets to inform re-use and may require subject-matter experts to draft data quality questions. Data curators would need subject-matter expertise, if not experts, to determine these facets to enable re-use of any data. The final question allows a data re-user/consumer to again either acknowledge or disavow the role of metadata, not rich in this principle but sufficient. There is no chance that sufficiency of metadata could be determined without user input.

Discussion and Future Work

The FAIR data principles address a critical gap in synthesizing and achieving consensus for guidance and assessment on the FAIR-ness of data, a critical and important contemporary imperative to advance science and discovery. Applying these proposed interview questions, derived from FAIR principles, may provide future data producers and data managers some considerations for how to operationalize the principles into metrics. One outcome may be to build fitness for use frameworks to inform what metadata elements are actually used to discover and evaluate data. As each community of re-users seeks different elements to determine fitness for use, many frameworks to guide other assessments could benefit through the use and refinement of these questions. A framework with the most vital facets of fitness for use would outline considerations for the functionality and design of data and metadata, as well as the tools used to find, access, and use both.

Ideally, the information needed to determine fitness for use is transmitted from the producer to the user via metadata, often with an information intermediary (i.e., data curator). Potential problems may arise when producers at the creation and ingest stage, and users at the dissemination and consumption stage, fail to realize the full value and potential of comprehensive, value-added metadata. For users, even when metadata are present, the metadata are unlikely to be consulted to determine fitness for use. When metadata might be housed separately from the data or within interfaces inhibiting search

for some important facets, or even when extraneous metadata overwhelms a re-user's patience for evaluation, it means that accessible metadata may not actually inform re-use. Additional studies should explore this intersection and resulting gaps.

A secondary outcome of applying the fitness for use framework is greater awareness of FAIR data principles. Many re-users of data may not appreciate or understand the efforts occurring throughout the data lifecycle that are quite necessary. In fact, great service anticipates need and those doing the best work in data curation may go unnoticed by re-users. Re-users and consumers of all products do not necessarily have the awareness of how vital metadata is to discovering and evaluating their data. Metadata is not magic, suddenly appearing out of thin air. In addition, re-users of data may not actually make appropriate considerations before using data, for many reasons, if they do not know both the possible limitations and intended uses.

Future work on the assessment of data FAIR-ness should create context dependent questionnaires that take into account re-users' experiences and various, situationally relevant re-use scenarios. Re-uses of data are as unpredictable as science and its questions shift throughout time given that new discoveries inform old data and make it new again. The FAIR principles do provide a guide for data curators and direct assessment from the end-user is not intended, but as outlined in this paper, many considerations emerge when considering FAIR-ness from data re-user/consumer perspectives. If the data do not meet the FAIR principles as imagined by end-users, however, then the F, A, I, and R assessments from data producer and curator perspectives seem to miss the mark. Some FAIR principles may be automated, but others necessitate evaluation at granularities not present in FAIR's present high-level, conceptual framework.

The next steps for these questions are to assess FAIR data principles from re-users' perspectives. Actual data consumers' perspectives on how they discover and evaluate data give new insights into how scientists in specific communities determine fitness for use. Select communities are currently being interviewed using these questions. Some revision is needed, but an open-ended approach has allowed re-users to tell their data stories and reveal their information-seeking behaviour. Data analyses and cross-community commonalities shall inform fitness for use frameworks that in turn would assist with design and assess the most vital metadata for discovery and evaluation for end-users. Ultimately, the data is for re-use and re-uses are not to stem solely from the data creator or play out as any data curator dictates.

In the original fitness for use discussion, the authors acknowledged that many users do not fully understand the technological nature of any product but consumers can make sensory judgments to assess if the "bread smells fresh-baked" (Juran, 2002). Aside from the fact that digital data should not smell, all working with the aspirant FAIR principles and assessment of them must recognize many re-users will not exhibit information-seeking behaviour beyond the smell test. Data producers and data curators should at least acknowledge we may not know what sensory judgments are being made, but invite future study into this area of data research. The look and feel (e.g., information structure) of data itself or where the data is found and accessed do present some level of trust and acceptance to users without any stamp of approval from data organizations. At this point, we are far from truly assessing re-use of data in its totality. The unimaginable re-uses discovered by artificial intelligence and the promise of undiscovered knowledge require a better understanding of capturing what we already do understand. This work all starts with data that are findable, accessible, interoperable, and ultimately re-usable by those beyond its creation or curation.

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

- Arbuckle, P., Kahn, E., & Kriesberg, A. (2017). Challenge paper: Challenges to sharing data and models for life cycle assessment. *Journal of Data and Information Quality*, 9(1), 1-4. doi:10.1145/3106236
- Bishop, B.W., & Grubestic, T.H. (2016). Geographic information: Organization, access, and use. New York: Springer.
- Bishop, B.W., & Hank, C. (2016). Data curation profiling of biocollections. *Proceedings of the Association for Information Science and Technology*, 53: 1–9. doi:10.1002/pra2.2016.14505301046
- Carlson, J. (2010). The data curation profiles toolkit: Interviewer’s manual. Data Curation Profiles Toolkit. Paper 2. doi:10.5703/1288284315651
- Chrisman, N.R. (1984). Part 2: Issues and problems relating to cartographic data use, exchange and transfer: The role of quality information in the long-term functioning of a geographic information system. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 21(2–3), 79–88. doi:10.3138/7146-4332-6J78-0671
- Diepenbroek, M., Schindler, U., Huber, R., Pesant, S., Stocker, M., Felden, J., . . . 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
- Juran, J.M. (2002). Juran’s quality handbook. Norwich, NY: Knovel.
- 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
- 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. <http://doi.org/10.1038/sdata.2016.18>

Wilkinson, M.D. et al. (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)

Veregin, H. (1999). Data quality parameters. In P.A. Longley, M.F. Goodchild, D.J. Maguire, & D.W. Rhind. (Eds.), *Geographical information systems – Volume I, Principles and technical issues*. (2nd ed., pp. 177–189). New York: Wiley.