



UvA-DARE (Digital Academic Repository)

FAIR Data in Medical Research

Incorporating the FAIR Principles in the Research Data Life Cycle

Kersloot, M.G.

Publication date

2022

[Link to publication](#)

Citation for published version (APA):

Kersloot, M. G. (2022). *FAIR Data in Medical Research: Incorporating the FAIR Principles in the Research Data Life Cycle*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 3

FAIRification efforts of clinical researchers

The current state of affairs

Martijn G. Kersloot
Philip van Damme
Ameen Abu-Hanna
Derk L. Arts
Ronald Cornet

Studies in Health Technology and Informatics 287, 35-39 (2021).

DOI: [10.3233/SHTI210807](https://doi.org/10.3233/SHTI210807)

Abstract

The FAIR Principles are supported by various initiatives in the biomedical community. However, little is known about the knowledge and efforts of individual clinical researchers regarding data FAIRification. We distributed an online questionnaire to researchers from six Dutch University Medical Centers, as well as researchers using an Electronic Data Capture platform, to gain insight into their understanding of and experience with data FAIRification. 164 researchers completed the questionnaire. 64.0% of them had heard of the FAIR Principles. 62.8% of the researchers spent some or a lot of effort to achieve any aspect of FAIR and 11.0% addressed all aspects. Most researchers were unaware of the Principles' emphasis on both human and machine readability, as their FAIRification efforts were primarily focused on achieving human readability (93.9%), rather than machine readability (31.2%). In order to make machine-readable, FAIR data a reality, researchers require proper training, support, and tools to help them understand the importance of data FAIRification and guide them through the FAIRification process.

Supplementary material referenced in this chapter can be accessed through doi.org/10.6084/m9.figshare.c.5617396.

Introduction

In order to improve and support the reuse of scholarly output, a multidisciplinary group of researchers published the fifteen FAIR Guiding Principles for scientific data management and stewardship in 2016, stating that scholarly output should be Findable, Accessible, Interoperable, and Reusable, both for machines and for people [18]. The paper describing the Principles explicitly emphasizes FAIRness for machines, and in another publication, Mons et al. reiterate that “FAIR is not just about humans being able to find, access, reformat and finally reuse data”. Researchers are increasingly required, either by their institutions or funders, to FAIRify their data (i.e., to make their data more FAIR). However, FAIR provides guidance for research data management and is not a standard [16], hence researchers need to interpret the Principles for their use case and make implementation choices accordingly [39].

Within the biomedical community, there are various initiatives that endorse the Principles and aim to develop guidance and tools for researchers [58]. Sinachi et al., for example, developed a FAIRification workflow specific to health research [34]. Practical examples of the implementation of the FAIR Principles include the collection of linked, human- and machine-readable data of COVID-19 patients [24] and rare disease patients [62, 63] and the reuse of such machine-readable data to perform distributed analyses [64]. Despite these initiatives, little is known about the knowledge and efforts of individual researchers regarding data FAIRification. Therefore, we sought to gain insight into clinical researchers' understanding of and experience with data FAIRification.

Methods

An online English questionnaire was developed ([online supplementary file](#)). In the questionnaire, researchers were asked to report if they were familiar with the FAIR Principles, if they knew the meaning of each of the letters of FAIR (Findability, Accessibility, Interoperability, and Reusability), and if they could describe what each aspect of FAIR entailed. In addition, they were asked to rank (no effort, very little effort, some effort, a lot of effort) and describe their current efforts for making data more FAIR. Clinical researchers were invited by email by Research Data Management departments in five out of the seven Dutch University Medical Centers (UMCs). PhD student associations of three UMCs sent out invitation emails to their members. In addition, users of Electronic Data Capture (EDC) platform Castor EDC [50] received an invitation via an popup in the platform. The invitation did not include any mention of FAIR, to ensure that potential participants that did not know about FAIR were also included. Researchers were eligible to participate in our study if they were setting up databases for clinical research. Consent of respondents was required before the questionnaire could be opened. Data were collected in Castor EDC [50] between November 27, 2020, and February 27, 2021. Statistical analyses were performed using R (version 4.0.2) [51]. Questionnaires with complete answers to mandatory questions were included in the analysis. Free-text answers given by researchers describing their interpretation of the FAIR Principles and their FAIRification

efforts were assessed by two medical informaticians (MK, PvdD). We assessed whether the descriptions of the Principles were related to human or machine readability. The FAIRification efforts were divided into categories by way of induction, and the contribution of each category to human and machine readability was evaluated.

Results

A total of 164 researchers completed the questionnaire. Their demographics are summarised in [Table 3.1](#). 51.2% (n = 84) of them opened the questionnaire via an invitation in a UMC and 48.8% (n = 80) did so via the popup in the EDC system. The majority (87.2%, n = 143) worked in a UMC and was a PhD candidate (84.8%, n = 139). 64.0% (n = 105) of all researchers had heard of the FAIR Principles, 45.1% (n = 74) claimed to know what the Principles entailed, and 56.1% (n = 72) claimed to know the meaning of at least one of the letters in FAIR (Findable, Accessible, Interoperable, and Reusable). Not all researchers that claimed to know the meaning of one or more of the letters, gave a correct meaning ([Figure 3.1](#)). Findable was the aspect that was understood best, followed by Accessible, Reusable, and Interoperable. The minority of the descriptions of the Principles and aspects given by researchers were related to machine readability ([Figure 3.2](#)).

After explaining the Principles to the researchers, 81.1% (n = 133) of them stated that they have spent at least very little effort to make their data more Findable, Accessible, Interoperable, or Reusable (*any* aspect of FAIR). For *all* aspects of FAIR, this was 25.6% (n = 42). 62.8% (n = 103) of the researchers spent some or a lot of effort to achieve *any* aspect of FAIR. For *all* aspects this was 11.0% (n = 18).

88.0% (n = 117) of the researchers that spent at least very little effort in making their data FAIR provided a description of their FAIRification efforts. An overview of the most common reported efforts is listed in [Table 3.2](#). Of all mentioned efforts (N = 231), 93.9% (n = 217) focused on human readability (e.g., using a data dictionary or using social media to share research outputs) and 31.2% (n = 72) focused on machine readability (e.g., submitting data and metadata to a data repository or using standardized terminologies).

Discussion

In this study, we conducted an online questionnaire to gain insights into clinical researchers' understanding of the FAIR Principles and their FAIRification experiences. We found that 64.0% (n = 105) of the researchers have heard of the Principles and that 62.8% (n = 103) of all researchers spent at least some effort to achieve to achieve *any* aspect of FAIR, and 11.0% (n = 18) regarding *all* aspects.

A strength of our study is that we invited researchers via an EDC platform and via UMCs, ensuring that we included researchers who were responsible for data management, and thus FAIRification. Moreover, we made certain that our invitation did not mention FAIR, in order to recruit both researchers who were and were not familiar with the Principles. Unfortunately, we were unable to verify the background of the respondents, because potential respondents were not directly invited by the research team to

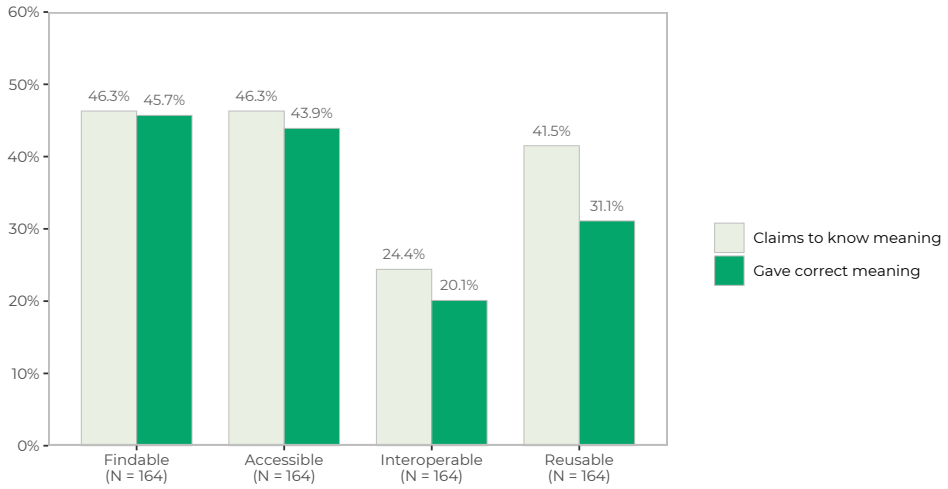


Figure 3.1: Researchers' knowledge of the individual FAIR aspects

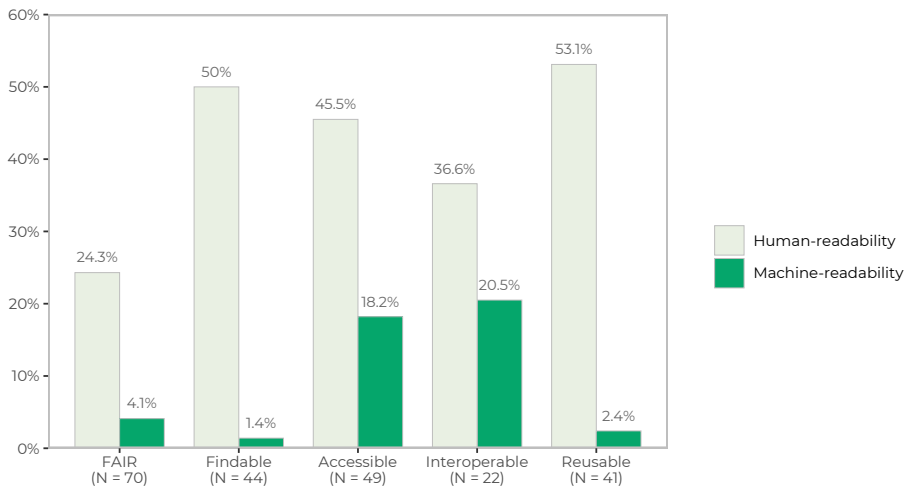


Figure 3.2: Researchers' descriptions of the individual FAIR aspects

Table 3.1: Demographics of the included respondents

	All respondents (N = 164)		FAIR knowledge (N = 74)		No FAIR knowledge (N = 90)	
	n	(%)	n	(%)	n	(%)
Profession						
PhD candidate	139	(84.8)	59	(79.7)	80	(88.9)
Post-doc	11	(6.71)	8	(10.8)	3	(3.33)
Assistant professor	4	(2.44)	2	(2.70)	2	(2.22)
Professor	3	(1.83)	2	(2.70)	1	(1.11)
Associate professor	2	(1.22)	1	(1.35)	1	(1.11)
Other	5	(3.05)	2	(2.70)	3	(3.33)
Age						
< 30	110	(67.1)	45	(60.8)	65	(72.2)
30 - 39	41	(25.0)	21	(28.4)	20	(22.2)
40 - 49	10	(6.10)	6	(8.11)	4	(4.44)
50 - 59	1	(0.61)	1	(1.35)	0	(0.00)
≥ 60	1	(0.61)	0	(0.00)	1	(1.11)
Rather not tell	1	(0.61)	1	(1.35)	0	(0.00)
Primary Institution						
Academic hospital	143	(87.2)	65	(87.8)	78	(86.7)
University	8	(4.88)	2	(2.70)	6	(6.67)
Teaching hospital	6	(3.66)	3	(4.05)	3	(3.33)
Hospital	5	(3.05)	2	(2.70)	3	(3.33)
Other	2	(1.22)	2	(2.70)	0	(0.00)
Research experience						
< 1 year	19	(11.6)	7	(9.46)	12	(13.3)
1 - 2 years	41	(25.0)	21	(28.4)	20	(22.2)
2 - 4 years	53	(32.3)	16	(21.6)	37	(41.1)
4 - 6 years	29	(17.7)	13	(17.6)	16	(17.8)
≥ 6 years	22	(13.4)	17	(23.0)	5	(5.56)

Table 3.2: Efforts to make data more FAIR, as reported by researchers

	Findable (N = 72)		Accessible (N = 84)		Interoperable (N = 64)		Reusable (N = 78)		Total (N = 117)	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Standardize data collection	5	(6.9)	1	(1.2)	23	(35.9)	12	(15.4)	41	(35.0)
Add dataset descriptions	5	(6.9)	6	(7.1)	4	(6.3)	17	(21.8)	32	(27.4)
Using a shared drive	2	(2.8)	19	(22.6)	0	(0.0)	1	(1.3)	22	(18.8)
Using a data capture system / EMR	6	(8.3)	9	(10.7)	4	(6.3)	2	(2.6)	21	(17.9)
Deposit data in a data archive	6	(8.3)	12	(14.3)	1	(1.6)	0	(0.0)	19	(16.2)
Create study protocol and register study	7	(9.7)	1	(1.2)	4	(6.3)	4	(5.1)	16	(13.7)
Provide data availability/access statement	8	(11.1)	6	(7.1)	0	(0.0)	1	(1.3)	15	(12.8)
Standardized file management	11	(15.3)	2	(2.4)	0	(0.0)	0	(0.0)	13	(11.1)
Add data as supplementary material	5	(6.9)	3	(3.6)	0	(0.0)	1	(1.3)	9	(7.7)
Storing data in a database	2	(2.8)	4	(4.8)	2	(3.1)	1	(1.3)	9	(7.7)
Use social media	5	(6.9)	2	(2.4)	0	(0.0)	1	(1.3)	8	(6.8)
Add documentation	1	(1.4)	0	(0.0)	0	(0.0)	6	(7.7)	7	(6.0)
Using a Data Management Plan	3	(4.2)	1	(1.2)	1	(1.6)	1	(1.3)	6	(5.1)
Publish open access	1	(1.4)	3	(3.6)	0	(0.0)	0	(0.0)	4	(3.4)
Add keywords to dataset or publication	4	(5.6)	0	(0.0)	0	(0.0)	0	(0.0)	4	(3.4)
Ask participants for informed consent	0	(0.0)	0	(0.0)	0	(0.0)	2	(2.6)	2	(1.7)
Using standardized data/file formats	1	(1.4)	0	(0.0)	1	(1.6)	0	(0.0)	2	(1.7)
Using a digital research environment	0	(0.0)	1	(1.2)	0	(0.0)	0	(0.0)	1	(0.9)

Chapter 3

comply with the General Data Protection Regulation. Lastly, partially filled-in questionnaires were excluded from the analysis, which could have introduced selection bias, for researchers who were unaware of FAIR could have stopped filling in the questionnaire.

The definition of the Principles and meaning of the individual letters of FAIR were known only by a minority of the researchers participating in our study and only a small amount of researchers spent effort to achieve FAIRness. Moreover, it is still unclear to researchers which efforts contribute to which aspect of FAIR. For example, using a data capture system to collect data does not specifically mean that the data collected in that system are more Findable (F), Accessible (A), or Interoperable (I) with data collected in other systems. The same applies to the use of standardized variable names for collecting data, where free-text variable names do not specifically make data more Interoperable (I) for machines. This indicates the need for simple, easy-to-follow explanations of what the Principles mean in practice and how they affect the (daily) work of clinical researchers. Specifically, researchers should receive training and guidance to help them understand which steps they should take to make their data more FAIR and how these steps are related to the individual aspects of the Principles. A first step toward achieving this is to ensure that there is convergence on FAIR implementations in the healthcare domain. Community-specific FAIR Implementation Profiles (FIPs) [65] can drive this convergence.

Most descriptions of researchers focused on human readability, rather than machine readability, and only a minority of the researchers focused on achieving machine readability. This might be due to the fact that the current workflows mentioned in the literature and software available to make (meta)data machine-readable require a significant amount of background knowledge (e.g., knowledge of data modeling, terminology systems, or metadata schemes). Researchers should move away from the use of such “professorware” and turn to sustainable systems [37]: professional products and services that support them in the creation and use of FAIR data [66]. To ensure that the systems are integrated into the researchers’ existing working processes, research software vendors should develop these systems in collaboration with researchers and research support staff, and integrate them with systems that are currently used.

Conclusion

A large number of clinical researchers is currently unaware of the definition of the FAIR Principles and their emphasis on both human and machine readability. Researchers are undertaking efforts to make their data more FAIR, but their focus is primarily on human readability, rather than machine readability. In order to make machine-readable, FAIR data a reality, researchers need proper training, support, and tools to help them understand the importance of data FAIRification and guide them through the FAIRification process.