

FAIRification Efforts of Clinical Researchers: The Current State of Affairs

Martijn G. KERSLOOT^{a,b,1}, Philip VAN DAMME^a, Ameen ABU-HANNA^a,
Derk L. ARTS^b and Ronald CORNET^a

^a*Amsterdam UMC, University of Amsterdam, Department of Medical Informatics,
Amsterdam Public Health Research Institute, Amsterdam, The Netherlands*

^b*Castor EDC, Amsterdam, The Netherlands*

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.

Keywords. FAIR data, medical research, Research Data Management

1. 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 [1]. 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 [2], hence researchers need to interpret the Principles for their use case and make implementation choices accordingly [3].

Within the biomedical community, there are various initiatives that endorse the Principles and aim to develop guidance and tools for researchers [4]. Sinachi et al., for ex-

¹Corresponding Author: Martijn G. Kersloot, m.g.kersloot@amsterdamumc.nl.

ample, developed a FAIRification workflow specific to health research [5]. Practical examples of the implementation of the FAIR Principles include the collection of linked, human- and machine-readable data of COVID-19 patients [6] and rare disease patients [7,8] and the reuse of such machine-readable data to perform distributed analyses [9]. 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.

2. Methods

An online English questionnaire was developed [10]. 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 [11] received an invitation via a 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 [11] between November 27, 2020, and February 27, 2021. Statistical analyses were performed using R (version 4.0.2) [12]. 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, PvD). 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.

3. Results

A total of 164 researchers completed the questionnaire. Demographics can be found at [10]. 51.2% (n = 84) 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 1.1). 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 1.2).

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).

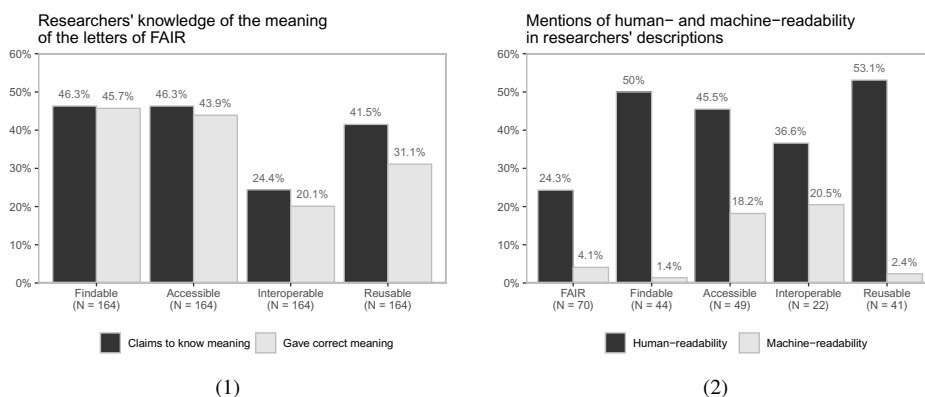


Figure 1. Researchers' knowledge and descriptions of the individual FAIR aspects

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 1. 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).

Table 1. Efforts to make data more FAIR, as reported by researchers

	F N = 72		A N = 84		I N = 64		R 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

This table does not include efforts reported by less than 5% of the researchers.

4. 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 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 FAIR 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 [13] 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 [14]: professional products and services that support them in the creation and use of FAIR data [15]. 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.

5. 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.

References

- [1] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016 Mar;3(1). Available from: <https://doi.org/10.1038/sdata.2016.18>.
- [2] Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*. 2017 Mar;37(1):49–56. Available from: <https://doi.org/10.3233/ISU-170824>.
- [3] Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, et al.. FAIR Principles: Interpretations and Implementation Considerations. MIT Press - Journals; 2020. Available from: https://doi.org/10.1162/dint_r_00024.
- [4] Trifan A, Oliveira JL. Towards a More Reproducible Biomedical Research Environment: Endorsement and Adoption of the FAIR Principles. In: *Biomedical Engineering Systems and Technologies*. Springer International Publishing; 2020. p. 453–470. Available from: https://doi.org/10.1007/978-3-030-46970-2_22.
- [5] Sinaci AA, Núñez-Benjumea FJ, Gencturk M, Jauer ML, Deserno T, Chronaki C, et al. From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. *Methods of Information in Medicine*. 2020 Jun;59(S 01):e21–e32. Available from: <https://doi.org/10.1055/s-0040-1713684>.
- [6] Reisen M, Oladipo F, Stokmans M, Mpezamihgo M, Folorunso S, Schultes E, et al. Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. *Advanced Genetics*. 2021 Jun;2(2). Available from: <https://doi.org/10.1002/ggn2.10050>.
- [7] Jannik S, Dennis K, Jens G, Christian-Alexander B, Marco R, van Enckevort David, et al. OSSE Goes FAIR - Implementation of the FAIR Data Principles for an Open-Source Registry for Rare Diseases. *Studies in Health Technology and Informatics*. 2018;253:209–213. Available from: <https://doi.org/10.3233/978-1-61499-896-9-209>.
- [8] Groenen KHJ, Jacobsen A, Kersloot MG, dos Santos Vieira B, van Enckevort E, Kaliyaperumal R, et al. The de novo FAIRification process of a registry for vascular anomalies. *Orphanet Journal of Rare Diseases*. 2021 Sep;16(1). Available from: <https://doi.org/10.1186/s13023-021-02004-y>.
- [9] Beyan O, Choudhury A, van Soest J, Kohlbacher O, Zimmermann L, Stenzhorn H, et al. Distributed Analytics on Sensitive Medical Data: The Personal Health Train. *Data Intelligence*. 2020 Jan;2(1-2):96–107. Available from: https://doi.org/10.1162/dint_a_00032.
- [10] Kersloot MG, van Damme P, Abu-Hanna A, Arts DL, Cornet R. FAIRification efforts of clinical researchers: the current state of affairs. *figshare*; 2021. Available from: <https://doi.org/10.6084/m9.figshare.c.5617396.v1>.
- [11] Castor EDC. Castor Electronic Data Capture; 2020. Available from: <https://www.castoredc.com>.
- [12] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. Available from: <https://www.R-project.org/>.
- [13] Schultes E, Magagna B, Hettne KM, Pergl R, Suchánek M, Kuhn T. Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In: *Lecture Notes in Computer Science*. Springer International Publishing; 2020. p. 138–147. Available from: https://doi.org/10.1007/978-3-030-65847-2_13.
- [14] Mons B. Data Stewardship for Open Science. Chapman and Hall/CRC; 2018. Available from: <https://doi.org/10.1201/9781315380711>.
- [15] van Vlijmen H, Mons A, Waalkens A, Franke W, Baak A, Ruiters G, et al. The Need of Industry to Go FAIR. *Data Intelligence*. 2020 Jan;2(1-2):276–284. Available from: https://doi.org/10.1162/dint_a_00050.