The A of FAIR - As Open as Possible, as Closed as Necessary (Postprint)

Authors: Landi, Annalisa, Thompson, Mark, Giannuzzi, Viviana, Bonifazi, Fedele, Labastida, Ignasi, Santos, Luiz Olavo Bonino da Silva, Roos, Marco, Landi, Annalisa

Date: 2022-11-16T19:16:34+00:00

Abstract

In order to provide responsible access to health data by reconciling benefits of data sharing with privacy rights and ethical and regulatory requirements, Findable, Accessible, Interoperable and Reusable (FAIR) metadata should be developed. According to the H2020 Program Guidelines on FAIR Data, data should be "as open as possible and as closed as necessary", "open" in order to foster the reusability and to accelerate research, but at the same time they should be "closed" to safeguard the privacy of the subjects. Additional provisions on the protection of natural persons with regard to the processing of personal data have been endorsed by the European General Data Protection Regulation (GDPR), Reg (EU) 2016/679, that came into force in May 2018. This work aims to solve accessibility problems related to the protection of personal data in the digital era and to achieve a responsible access to and responsible use of health data. We strongly suggest associating each data set with FAIR metadata describing both the type of data collected and the accessibility conditions by considering data protection obligations and ethical and regulatory requirements. Finally, an existing FAIR infrastructure component has been used as an example to explain how FAIR metadata could facilitate data sharing while ensuring protection of individuals.

Full Text

Preamble

Related to other papers in this special issue: 3 (p30); 20 (p199); 7 (p66)
Addressing FAIR principles: F1, F2, F3, F4, A1, R1

The "A" of FAIR – As Open as Possible, as Closed as Necessary

Authors: Annalisa Landi[1]†, Mark Thompson[2], Viviana Giannuzzi[1], Fedele Bonifazi[1], Ignasi Labastida[3], Luiz Olavo Bonino da Silva Santos[4] & Marco Roos[2]

[1]Fondazione per la Ricerca Farmacologica Gianni Benzi Onlus, 30 – 70010 Valenzano (BA), Italy

[2]Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands

[3]Learning and Research Resources Centre (CRAI), Universitat de Barcelona, Catalunya 08007, Spain

[4]GO FAIR International Support & Coordination Office (GFISCO), Leiden, The Netherlands

Keywords: Accessibility of health data; FAIR metadata; GDPR; Data sharing; ELSI

Citation: A. Landi, M. Thompson, V. Giannuzzi, F. Bonifazi, I. Labastida, L.O. Bonino da Silva Santos & M. Roos. The "A" of FAIR – As open as possible, as closed as necessary. *Data Intelligence* 2(2020), 47–55. doi: 10.1162/dint_a_{00027}

Abstract

In order to provide responsible access to health data by reconciling benefits of data sharing with privacy rights and ethical and regulatory requirements, Findable, Accessible, Interoperable and Reusable (FAIR) metadata should be developed. According to the H2020 Program Guidelines on FAIR Data, data should be "as open as possible and as closed as necessary"—open to foster reusability and accelerate research, yet closed to safeguard subject privacy. Additional provisions on protecting natural persons regarding personal data processing have been endorsed by the European General Data Protection Regulation (GDPR), Reg (EU) 2016/679, which came into force in May 2018. This work aims to solve accessibility problems related to personal data protection in the digital era and achieve responsible access to and use of health data. We strongly suggest associating each dataset with FAIR metadata describing both the type of data collected and the accessibility conditions, considering data protection obligations and ethical and regulatory requirements. Finally, an existing FAIR infrastructure component is used as an example to explain how FAIR metadata could facilitate data sharing while ensuring individual protection.

† Corresponding author: Annalisa Landi (E-mail: al@benzifoundation.org, ORCID: 0000-0001-9368-6424)

---

1. Introduction

"Recent rapid expansions of health and biological data, combined with the availability of sophisticated computational technologies, offer unprecedented opportunities to benefit public health," as discussed during a European Medicines Agency workshop in November 2016 [1]. Significant scientific insights may derive from matching human experts' knowledge with computers' capability to analyze and share large datasets. Health data hold real value for scientific research, and there is an urgent need to reconcile data sharing benefits with privacy rights, constraints, and ethical and regulatory requirements. FAIR metadata should provide responsible access to health data by safeguarding subjects while enabling data sharing for decision-making and improving computational technologies [2]. The FAIR principles support treating data as a resource for formulating and testing scientific hypotheses, aiming to increase data worth by enabling reuse. Increasing linkage between different data types—electronic health records, genetic data, patient-reported outcomes, clinical trial data—can produce numerous benefits, including enhanced disease knowledge, earlier diagnosis, better treatment selection (possibly personalized), and greater patient involvement in decision-making [3].

In recent years, "Interoperability" was considered the bottleneck of the FAIRification process [4], but today another major concern is "Accessibility," especially when processing sensitive data. Growing concerns about data accessibility relate primarily to its connection with "Reusability," as ensuring research data are reusable is a main FAIR principle focus, maximizing value and accelerating scientific research. Nevertheless, despite dramatically increased health data processing and reuse, widespread FAIR principle adoption, and rapid technological development enabling fast scientific advances, challenges for personal data protection persist and must be addressed. Since GDPR took effect in May 2018 [5], additional obligations on protecting natural persons regarding personal data processing have been introduced. Sometimes, even when data usage licenses allow sharing, access is not guaranteed for various reasons. Most restrictions derive from original consent obtained from data owners, institutional data protection policies, data sharing/use agreements between providers and recipients, and the significant effort required to transform data into machine-readable formats.

Notably, given growing concerns about health data accessibility, numerous initiatives aim to properly address data protection, ethical, and regulatory issues regarding health data sharing and access while promoting greater patient involvement in decision-making. For instance, Luxembourg's ELIXIR node is developing the open-source Data Information System (DAISY) for a research data registry as part of GDPR training activities. DAISY will enable data owners to create Data Information Sheets including essential data protection metadata such as data use restrictions, de-identification methods, and legal processing basis [7]. The "Automatable Discovery and Access Matrix" (ADA-M) initiative, released by the Global Alliance for Genomics and Health (GA4GH) and International Rare Diseases Research Consortium (IRDiRC) in 2016, enables re-

sponsible sharing of biomedical data and biospecimens, allowing data owners to choose visibility levels and availability degrees for different data types. Owners may establish re-contact possibilities for new findings and describe accessibility conditions (e.g., limitations and permissions) in metadata by expressing preferences regarding data users [8][9].

A similar initiative, the Beacon network—an open sharing platform developed by GA4GH and ELIXIR—allows data owners to choose access levels (open, registered, or controlled) for publishing genomics data and define which data types may be provided to each requestor category [10].

---

2. FAIR Metadata Describing Accessibility Conditions

Automating accessibility conditions while ensuring compliance with regulatory, ethical, and data protection obligations requires addressing the right balance between safeguarding individual privacy and generating potential benefits from automated machine processing and data reuse for research. Solving accessibility challenges and achieving responsible data access and use demands a scientific, multidisciplinary approach involving healthcare professionals, researchers, technical experts, and ethical/regulatory/legal specialists in developing FAIR metadata.

The H2020 Program Guidelines on FAIR Data introduced the concept that "data should be as open as possible and as closed as necessary" [11]. The "A" in FAIR stands for "Accessible under well-defined conditions." The Accessibility principle (A1) foresees creating metadata describing all conditions for responsible data access and use. According to FAIR principles, the metadata schema should be standardized to simplify processes across data sources and identify data requestors through authentication and authorization procedures. For instance, ELIXIR launched an Authentication and Authorization Infrastructure (AAI) in 2016 [10][12] to control and manage data user access rights, create different access levels, and meet privacy and data protection legislation obligations. Users may be identified through home organization credentials or community/commercial identities (e.g., ORCID, LinkedIn). Thus, achieving responsible data access requires infrastructures to use FAIR metadata associated with each dataset describing accessibility conditions and AAI infrastructures to authenticate and authorize requestors.

Furthermore, FAIR metadata should describe accessibility conditions in clear, plain language and remain publicly accessible even when data are no longer available. Considering that GDPR aims to strengthen individuals' rights to be better informed about lawful, fair data processing and give them greater control, it is important to verify whether existing metadata account for original consent and all applicable data sharing/use agreements and protection policies. Moreover, investigation is needed into whether metadata have implemented information

required for data subjects under GDPR Articles 13 and 14, and whether all data subject rights have been and will be respected.

When processing data, information must be provided regarding processing type, current and future purposes, and storage duration or criteria for determining persistence. Additional required information includes any personal data transfer to third countries and appropriate safeguards, any automated decision-making (including profiling), and responsible figures for data processing (the controller, Data Protection Officer (DPO), and personal data recipients). Finally, data subjects must be aware of their main rights: right to request access, deletion, or rectification; right to restrict or object to processing; right to data portability; right to lodge complaints with supervisory authorities; and right to withdraw consent [5].

Particular attention is required when processing health data from rare disease patients, especially involving non-EU countries. Such health data should be considered more sensitive (e.g., identifiable), requiring additional safeguards. Regarding non-EU countries, examples in Guidelines 3/2018 on GDPR territorial scope [13] should be consulted, as different situations may occur (e.g., data processed in non-EU countries while the controller/processor is EU-established).

As data use in scientific decision-making increases, compliance with all applicable ethical, regulatory, and data protection requirements is essential. The goal is creating FAIR operational metadata describing accessibility conditions for each dataset, considering all applicable protection, ethical, and regulatory requirements. To achieve this, existing regulatory and ethical documentation (e.g., informed consent forms, data sharing/use agreements, privacy policies) should be converted into computable, machine-readable, and "machine-actionable" formats, sufficiently structured and optimized for computer processing to speed up processes and improve data quality while protecting data subject rights. This appears to be the best approach for addressing complex ethical and regulatory requirements. Therefore, to realize a responsible data access model, actors in the access process (e.g., data subject, controller, processor, DPO) should be identified in advance, along with their responsibilities and applicable accessibility conditions (e.g., limitations and permissions) according to existing consent forms, agreements, and policies. Roles and responsibilities must always be well-defined when processing personal data, and data subjects must be provided with the controller's name and contact details and those of the DPO (if available) [5].

Technologies must implement mechanisms and technical protocols detailing accessibility and use constraints (limitations, obligations, permissions) to be provided to data users when requesting access. Notably, FAIR infrastructures urgently need sections enabling GDPR compliance to guarantee that data access and use are grounded on a legal basis. To achieve this goal and demonstrate practical application of our considerations, an existing FAIR infrastructure component—the FAIR Data Point (FDP) [14]—is used as an example. For detailed FDP design, specification, and implementation, we refer to online doc-

umentation [15][16]. We review the four main basic conceptual components of the FDP—Metadata Provider, Data Accessor, Security Enforcer, and Metrics Gather—in relation to the aforementioned accessibility considerations.

The **Metadata Provider**, responsible for providing data accessibility conditions (e.g., limitations, permissions), should create sufficient FAIR metadata for each dataset considering conditions foreseen by applicable informed consent forms, data sharing/use agreements, and privacy policies, while implementing GDPR requirements. The controlled access model currently foresees a data access committee performing regulatory and ethical checks to allow or deny data access and use. As mentioned, automating this process by converting documentation into machine-readable format would be optimal. Thus, all applicable data protection, ethical, and regulatory requirements should be considered during FAIR metadata creation to outline the entire framework of accessibility conditions for automatic evaluation, structuring, combination, and machine-readability. This would reduce data access committee effort, leading to net efficiency and cost-effectiveness gains, particularly where accessibility conditions were previously ill-defined or data users belong to different institutions.

Furthermore, the Metadata Provider should modify accessibility conditions according to data subject requests to guarantee rights respect. If a data subject wishes to withdraw consent, restrict processing, or erase/modify data, the Metadata Provider should evaluate requests according to GDPR requirements (Articles 15-22 and 34) and decide how to proceed. A copy of processed data should also be made available to the data subject upon request to verify processing lawfulness [5].

The **Security Enforcer**, acting as gatekeeper and protector against unauthorized access, should allow or reject access after evaluating requests. To perform complete assessments, the Security Enforcer should consider both dataset accessibility conditions and requestor identity and rights. The Metadata Provider should make specific accessibility conditions available as FAIR metadata associated with each dataset. Additionally, the Security Enforcer should be tightly coupled to an AAI infrastructure, as authentication and authorization procedures are needed for each dataset or individual data item access. For different access levels, requestors may be identified through ORCID, certified email addresses, online profile links, identification document copies, or (federated) AAI based on validated institutional accounts.

The **Data Accessor**, responsible for providing dataset access when granted by the Security Enforcer, should provide a machine-readable interface enabling human and machine data access.

The **Metrics Gather** component monitors all FDP components, tracking all data access requests (including numbers and types of requests, submission dates, requestor name and contact details, decisions taken, and data shared/accessed). Additionally, mechanisms to collect and display data user feedback on the process should be developed.

Importantly, while we describe possible implementation through the four basic FDP components, any FAIR technology should address these access considerations for personal data. Other FAIR infrastructures may consider these same four aspects to guarantee responsible data sharing and access compliant with all data protection, ethical, and regulatory requirements.

---

3. Conclusion

This paper proposes considering individual rights as integral to implementing FAIR principles for personal data. We described access considerations to guide FAIR principle implementation, emphasizing that these ethical and legal considerations are inevitable for personal data processing. For instance, while making data findable is the primary FAIRification goal, access considerations necessarily apply when personal data are involved. This should not lead to avoidance—such as preferring data anonymization techniques—but rather to describing access policies in human- and machine-readable terms, enabling automated access procedures for efficient large-scale data access.

In conclusion, achieving responsible data access requires balancing data sharing needs with data subject protection. Creating value from health data demands attention to ethical, regulatory, and data protection issues related to data access and use. Therefore, to solve digital-era personal data protection accessibility problems and achieve responsible health data access and use—reconciling data sharing benefits with privacy rights and ethical and regulatory requirements—FAIR metadata describing accessibility conditions must be developed.

Accessibility conditions in FAIR metadata should reflect original consent obtained from data owners and all applicable ethical and regulatory documentation for each dataset. GDPR obligations must be complied with by verifying that data subjects receive all required information and ensuring respect for their data processing rights. Unnecessary restrictions should be avoided, and access conditions—including all limitations, obligations, and permissions—must be described in machine-readable format by converting documentation into FAIR metadata for run-time processing and access requirement verification. These considerations should be implemented in FAIR infrastructures by identifying data processing roles and responsibilities and creating FAIR metadata that ensures data subject respect and responsible health data access. When FAIR metadata are developed to facilitate data free flow while ensuring high individual privacy protection levels, an efficient, informed, and responsible data access process becomes possible.

Author Contributions

A. Landi (al@benzifoundation.org) prepared the first draft. M. Thompson (m.thompson@lumc.nl) and M. Roos (M.Roos@lumc.nl) contributed to methodology setup. V. Giannuzzi (vg@benzifoundation.org) and M. Thompson collaborated in drafting. F. Bonifazi (fb@benzifoundation.org), I. Labastida (ilabastida@ub.edu), and L.O. Bonino da Silva Santos (luiz.bonino@go-fair.org) provided general revision. M. Thompson and M. Roos provided final draft review.

Acknowledgements

Part of this work is funded by the NWA program (project VWData - 400.17.605), Netherlands Organization for Scientific Research (NWO), European Joint Program Rare Diseases (grant agreement #825575), and ELIXIR-EXCELERATE (H2020-INFRADEV-1-2015-12).

References

[1] European Medicines Agency. Workshop on identifying opportunities for "big data" in medicines development and regulatory science. (2016). Available at: https://www.ema.europa.eu/en/events/workshop-identifying-opportunities-big-data-medicines-development-regulatory-science.

[2] S.O.M. Dyke, A.A. Philippakis, J.R. De Argila, D.N. Paltoo, E.R. Luetkemeier, B.M. Knoppers, … & S.T. Sherry. Consent codes: Upholding standard data use conditions. *PLOS Genetics*. 12(1)(2016), e1005772. doi:10.1371/journal.pgen.1005772.

[3] C. Ohmann, R. Banzi, S. Canham, S. Battaglia, M. Matei, C. Ariyo, … & J. Demotes-Mainard. Sharing and reuse of individual participant data from clinical trials: Principles and recommendations. *BMJ Open* 7(2017), e018647. doi: 10.1136/bmjopen-2017-018647.

[4] A. Jacobsen, R. Kaliyaperumal, L.O. Bonino da Silva Santos, B. Mons, E. Schultes, M. Roos & M. Thompson. A generic workflow for the data FAIRification process. *Data Intelligence* 2(2020), 56–65. doi: 10.1162/dint_a_{00028}.

[5] European Parliament and Council of the European Union. Regulation (EU) 2016/679 of The European Parliament and of The Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*. May, 2016.

[6] I. Labastida & T. Margoni. Licensing FAIR data for reuse. *Data Intelligence* 2(2020), 199–207. doi: 10.1162/dint_a_{00042}.

[7] GDPR activities. ELIXIR-LU. Available at: https://elixir-luxembourg.org/gdpr-activities.

[8] GA4GH/ADA-M. Available at: https://github.com/ga4gh/ADA-M.

[9] J.P. Woolley, E. Kirby, J. Leslie, F. Jeanson, M.N. Cabili, G. Rushton, … & A.J. Brookes. Responsible sharing of biomedical data and biospecimens via the "Automatable Discovery and Access Matrix" (ADA-M). *npj Genomic Medicine* 3(1)(2018), Article No. 17. doi: 10.1038/s41525-018-0057-4.

[10] M. Linden, M. Procházka, I. Lappalainen, D. Bucik, P. Vyskocil, M. Kuba, … & T. Nyrönen. Common ELIXIR service for researcher authentication and authorization. *F1000Research* 7(2018), 1199. doi: 10.12688/f1000research.15161.1.

[11] European Commission. Directorate-General for Research & Innovation. H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. 26 July 2016.

[12] ELIXIR AAI: Authentication and authorization infrastructure. Available at: https://www.elixir-europe.org/services/compute/aai (accessed February 17, 2019).

[13] European Data Protection Board. Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) - Version for public consultation. Adopted on 16 November 2018.

[14] L.O. Bonino da Silva Santos, M.D. Wilkinson, A. Kuzniar, R. Kaliyaperumal, M. Thompson, M. Dumontier & K. Burger. FAIR data points supporting big data interoperability. In: M. Zelm, G. Doumeingts & J.P. Mendonça (eds.) *Enterprise Interoperability in the Digitized and Networked Factory of the Future*. ISTE Press, 2016, pp. 270–79.

[15] FAIRDataTeam/FAIR Data Point metadata specification. GitHub. (2019). Available at: https://github.com/FAIRDataTeam/FAIRDataPoint-Spec/blob/master/spec.md.

[16] FAIRDataTeam/FAIR Data Point design specification. GitHub. (2019). Available at: https://github.com/FAIRDataTeam/FAIRDataPoint-Spec/blob/master/README.md.

Note: Figure translations are in progress. See original paper for figures.