Let's FAIRify Electronic Health Records*

Derek Corrigan[1,*,†], Qurratal Ain Fatimah[2], Syeda Mah-e-Fatima[3] and Ali Hasnain[4,*,†]

[1]FutureNeuro SFI Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland
[2]University Hospital Galway, Ireland
[3]School of Pharmacy and Biomedical Sciences, Royal College of Surgeons in Ireland, Dublin

Abstract

This paper discusses the importance as well as the practical issues that can be expected while working towards the FAIRification of Electronic Health Records (EHRs). Electronic health records are medical records of a patient that document their illness history as well as other relevant personal information. Patients interact with health care professionals in a multitude of capacities and thus, there is a stream of medical data available at any given time. So far, electronic health records have not been utilized to their full potential, which in part is attributable to the many dimensions the health care profession is split into, and the lack of interconnectivity between disciplines. Having patient data either in one source or alternatively interconnected fulfilling the principles of FAIR data, i.e. findable, accessible, interoperable and reusable, will be of immense benefit to both the healthcare system as well as for the patient. Streamlined FAIR compliance electronic health records will mean all the pertinent data should be linked or connected for better findability and accessibility to make it reusable and interoperable. In this position paper we look at the principles of FAIR data, identify priorities in relation to making data reusable, as well as recognizing points of action and complications in delivering FAIRified EHRs.

Keywords

FAIR (Findable, Accessible, Interoperable, Reusable), Electronic Health Records, FAIR digital objects

1. Introduction

An Electronic Health Record (EHR) is a longitudinal collection of electronic health information about a patient. The ubiquity of EHR systems has transformed the health care landscape over several decades. Yet, even as improved patient care and cost savings have begun to emerge, significant usability impediments have been documented. These include the duplication and fragmentation of EHRs that hampers interoperability efforts and impact on patients' safety. Facilitating better access to and sharing of structured and unstructured health data is crucial to ensure greater accessibility, availability, and affordability of healthcare. It will also stimulate innovation in health and care for better treatment and outcomes and foster innovative solutions that make use of digital technologies.

SeWebMeDA-2022: 5th International Workshop on Semantic Web solutions for large-scale biomedical data analytics, May 29, 2022, Hersonissos, Greece
*Corresponding author.
†These authors contributed equally.
derekcorrigan@rcsi.ie (D. Corrigan); alihasnain@rcsi.com (A. Hasnain)
https://www.rcsi.com/people/profile/alihasnain (A. Hasnain)
0000-0001-7632-6013 (D. Corrigan); 0000-0003-4014-4394 (A. Hasnain)
© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3573/paper8.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

The collection, access, use and re-use of health data in healthcare poses specific challenges in finding the right balance between measures that facilitate data sharing while preserving the interests and rights of individuals, including their personal data protection.

It is important to highlight that unstructured data represent the 80% of data contained in EHRs [1]. Unstructured content is the text of anamnestic notes, physical examination sheets, medical and nurse diaries, surgical forms, specialist test reports, discharge letters written during the patient's stay, and any comment extending the standard patient-reported outcome measures collected during the follow-up phase. This heterogeneous textual content contains relevant, detailed and nuanced information about the illness trajectory and care processes undertaken by and upon the patients. This makes the challenge to automatically extract accurate information from narrative notes worthwhile.

Digital health technology and data pose a critical opportunity and challenge for researching, practicing, and experiencing healthcare, promoting public health policies, and changing the way in which medicine is understood. However, digital health data is currently fragmented and dispersed in different and non-homogeneous repositories.

Standardization and common structures for facilitating interoperability and re-use of these data for clinical practice and for research and innovation is a challenge, especially when it comes to exploiting unstructured information in a secure way and making this information interoperable [2].

Personal data has also emerged in the economic literature as the world economy's new asset class. These data offer a cost-effective and technically feasible pathway to personalized medicine; create wealth and value, attract private and public investment and provide a return to the whole society by means of better and more efficient prevention and treatment management protocols. However, data assetization and exploitation (including its analysis and synthesis) needs further evidence, developments and research into capitalisation, techno-scientific applications and socio-cultural and socio-technical identification of resources and possibilities, including legal and structural facilitators and barriers [3].

Exchange of medical information has been associated with benefits for patients, health delivery systems and the economy. In fact, real-time health information systems, integrating all relevant information on a patient(s) and their healthcare process, can substantially improve coordinated care, patient safety, quality, and efficiency.

For this reason, it is critical to improve EHRs interoperability to achieve their full potential, that is, the ability of health information systems to work together within and across organizational boundaries to advance the effective delivery of healthcare. While electronic health records (EHRs) are of immense value in healthcare and medicine their potential has not been fully exploited due to fragmentation of the healthcare sector. This is due to the lack of interconnectivity and interoperability between different EHR systems.

To this purpose, FAIR data principles [4] provide clear guidelines to increase the findability, interoperability, accessibility and re-usability of data. The sensitive nature of data and information available in EHRs brings challenges to FAIRify it. An important point made in the EU FAIR data report [5] (that was published before the COVID pandemic) was the value of FAIR data - making data more widely available and reusable to support response to public health emergencies in a timely manner will be a valuable asset. This allows easier access and sharing of data across individual nations and beyond the boundaries across the European Union. The need to unlock

the data and the value of national EHR systems across the European Union as a resource for the promotion of population health poses more pressure in the current circumstances of a pandemic.

An argument made is that the research community has not always sufficiently acknowledged the inherent value related to the production of analytical datasets to support research initiatives. One of the recommendations from the EU commission report (step 3, recommendations 12 and 13) [5] is that development of FAIR data ecosystems allows for acknowledgement of the explicit value of research data that can be rewarded through the development of incentives and new metrics that measure data reuse. This can then incentivize and reward those who make it actively do it well through "good data stewardship".

In this position paper, we advocate the importance of FAIRification led interconnectivity of EHRs by presenting two motivational scenarios that arise on a daily basis in clinical settings e.g. accidents and emergency units that can be supported using well-connected interoperable EHRs at least at the national level.

2. Electronic Health Records

Electronic Health Records (EHRs) are primarily known as a digital collection of medical information about an individual patient that includes but is not limited to the information about a patient's health history, diagnoses, prescription medicines, diagnostic and monitoring tests, known allergies and immunizations. It can also include healthcare encounters and referrals to specialists and ambulatory/outpatient encounters in hospitals.

Exchange of medical information has been shown to be associated with higher quality and safer care for patients. It also has clear benefits for healthcare organizations- providers of care in community, outpatient and hospital settings as well as health insurers and the health sector more broadly. EHRs help healthcare providers better manage care for patients and provide better health care by:

- Providing accurate, up-to-date, and complete information about patients at the point of care at specific time
- Enabling quick access to patient records for more coordinated and efficient care
- Exploiting the potential of sharing electronic information with patients and other clinicians in a secure manner
- Helping healthcare providers more effectively diagnose patients, reduce medical or human errors while providing better and safer care
- Improving patient and health care provider interaction and communication, as well as health care convenience
- Enabling safer, efficient and more appropriate prescribing of drugs

The sensitive nature of healthcare data creates fragmented, siloed and mostly private repositories of data by design. Although online portals have increased patients' ability to view their records, the structuring of this access is limited by design. A minority of systems allow direct patient access to data and healthcare results. The degree to which data should be shared directly

with patients is still being debated within the healthcare community due to ever pressing requirements of safety and security of such sensitive data in the digital space.

On the other hand, holding and oversight of health data has largely been in the hands of multinational technology organizations. This data collection, acquisition, storage and holding of health data has evolved around the concept and motivation of the personal Electronic Health Record data that supports collection of data relating to individuals' exercise and lifestyle as well as physiological measurements such as pulse rate, rhythm and variability. Initiatives have emerged to develop these person-based EHRs, the most notable examples being Microsoft's HealthVault[1] and Google Health Google Health,[2]. These technology innovations failed to generate traction because of poor user engagement and concerns about security of personal data. Most critically there were significant limitations in terms of integration and interoperability with EHRs. The overlap between health data relating to diet, exercise and lifestyle activity with more traditional medical data including morbidity conditions, medication codes, and healthcare utilization (diagnostic testing, monitoring of disease, and contacts with health professionals in primary and secondary care) has never been fully resolved. This may well be explained by the confidential nature of medical data and the fact that prescription drugs require input and understanding from health professionals (doctors and pharmacists) who are concerned with prescribing, monitoring and dispensing of medicines needing specialist knowledge and regulatory oversight.

3. FAIR Data Overview

In 2018 the European Commission expert working group on FAIR data produced a detailed report about how to promote the concept of FAIR data to support more open science in the European Union [5]. This detailed report sets out some of the motivations for promoting FAIR data principles:

> *"It has long been recognised that it is not sufficient simply to post data and other research-related materials onto the web and hope that the motivation and skill of the potential user would be sufficient to enable reuse."* [5]

FAIR is a set of principles that collectively describe requirements to develop a FAIR ecosystem supporting digital data reuse [4, 6]. The basic goal of FAIR is to promote wider reuse and sustainability of digital datasets produced from research. This promotes better linkage and connections between different research initiatives both within and across research domains and disciplines. This in turn can accelerate discovery and increase the replicability of science. The FAIR acronym stands for:

- **Findable** – datasets are uniquely identified and indexed in publicly available resources that enable searchable discovery e.g., internet search engines

[1] Microsoft Cloud for Healthcare, https://www.microsoft.com/en-ie/industry/health/microsoft-cloud-for-healthcare?rtc=1. (last accessed 14-03-2022)
[2] Google Health, https://health.google/. (last accessed 14-03-2022)

- **Accessible** – datasets are accessible based on defined authentication requirements in publicly available repositories accessed using open protocols e.g., http
- **Interoperable** – datasets are described with metadata using open standards that determine dataset content based on accepted controlled terminologies and vocabularies to promote wider reuse
- **Reusable** – datasets are richly and accurately described by metadata that uses standards most appropriate to the domain of knowledge and includes the provenance, licensing and dataset content and appropriate use.

3.1. FAIR Digital Object

The basic building block to represent a data resource is the FAIR digital object. The object consists of a link to the dataset itself along with a set of metadata that fully describes the data according to the FAIR data principles. The metadata and the dataset are typically stored separately to allow for authentication and access control (a description of the data may be accessible but the actual data access may not if deemed sensitive).

The digital object ensures that the FAIR resource is uniquely identified, semantically described using open technical standards, and appropriately shared according to defined access controls based on licensing information.

An important consideration as highlighted in the EU report is the need for development of distributed FAIR ecosystems by storing collections of FAIR digital objects across distributed FAIR repositories that belong to searchable and uniquely defined organizations defined in FAIR registries [5]. The goal is the publication of datasets in a FAIR ecosystem, with defined licensing access and provenance that allow for long term sustainability and access to datasets that eliminates the need to make multiple copies of that dataset across multiple sites. It is envisioned that the datasets, once FAIRified are accessed from source:

"Data Federations offer a means to establish agreements between repositories or registries to carry out certain tasks collaboratively and therefore will be essential to this distributed system. Data will increasingly remain at various locations for reasons such as the expense of copying data or because of legal or ethical restrictions. Distributed queries, managed by brokering software, will be used to virtually integrate data. The need for such distributed analysis across multiple data sets is one of the major drivers and use cases for FAIR data: it requires metadata to find the data resources, protocols to access them, agreed specifications such that the data can interoperate and rich provenance information so that the data can be reused with confidence" [5].

An important distinction to be made is that FAIR data does not necessarily imply 'Open data'. As part of the metadata associated with FAIR datasets, it is envisioned that licensing will also be described which may allow full access or restricted access to only the metadata descriptions rather than the actual raw datasets themselves where it is considered that the data is sensitive or should be restricted. This also implies that a FAIR infrastructure must allow for authentication mechanisms that define how data can be shared across different FAIR data repositories, groups and disciplines while respecting the EU legislation across borders as well (such as General Data Protection Regulations). Whilst open data is considered desirable the EU envisions that data should be 'as open as possible, as closed as necessary'.

4. Motivating Scenarios

Integration of different patient repositories as well as electronic health records add value to the case that such integrations could make healthcare safer for healthcare professionals as well as patients. Consideration of FAIR data principles can be a foundational step for the digitization and integration of EHRs that are more findable, accessible, interoperable and reusable for better accessing clinical scenarios. The following practical real world scenarios show the motivation behind the FAIRification led integration of electronic health records.

Motivating Scenario 1 Susy is a patient who is involved in a road traffic accident and presents to a hospital to be seen by the doctor. Due to the nature of injuries sustained, the hospital orders, among other tests, a CT scan of the brain. The patient has the CT scan but is waiting many hours to be seen by the doctor. She eventually leaves without receiving the results of the tests and presents to another hospital. Since no records are available and the patient is unable to tell the doctor at the second hospital what tests she got in the first place, she has another CT scan at the second hospital. The patient is eventually discharged but has now had two CT scans of the head. The radiation exposure is significant. One in every 1800 CT scans leads to one excess cancer. This patient has now had exposure at two facilities due to data inaccessibility between the two facilities. If data is findable and accessible as proposed by the FAIRify principles, better clinical decisions such as limiting the amount of harmful radiation a patient is exposed to, can be made, and repetitive testing can be avoided. This will lead to better patient outcomes as well as reducing cost of patient care.

In a nutshell the accessibility of patient data could have been achieved through FAIRification led digitization of EHRs

Motivating Scenario 2 In a second scenario, a patient has been discharged from hospital following a prolonged in-patient stay. Upon discharge, a discharge summary has been completed by the hospital many days following discharge. The patient had multiple investigations in the hospital including blood tests, electrocardiogram (ECG) and radiological investigations, but none of the results are available to the GP who is the primary physician of the patient in the community. Some of the doses of medications have been adjusted, and needs referrals to multiple outpatient specialties for continuity of care, but as the GP is unaware due to a delay in the discharge summary and recommendations getting to them, there is a delay ultimately in the patient getting optimal care. If data is findable and accessible as proposed by the FAIRify principles to the corresponding GP (obviously in a secure and closed environment), better clinical decisions as well as timely diagnosis and recommended medication modifications would have been possible by the GP. This will also lead to better patient outcomes as well as reducing cost of patient care.

Motivating scenario 3 FAIRifying EHR will play a significant role in integrating research (interoperable data) along multiple platforms. If an interesting clinical scenario such as a rare cancer or genetic condition is observed in facility A, for instance, with EHR being accessible across a network of healthcare, it will be easy and feasible for another institution to not only gain access to the information due to accessibility of data, but also compare the findings and consolidate their own research with research from other sites. EHR can also then be used to recruit for clinical studies, following trends across institutions.

Accessibility of patient data available in hospital in-patient and its interoperability/ connec-

tivity (in a protected, secured and closed environment) with the EHRs available at GP practices could be supported through FAIRification led digitization of EHRs.

5. Describing the FAIR Data Ecosystem for EHRs

The implementation of FAIR data principles are described in the EC report "Turning FAIR into reality" [5], which identifies several priorities in relation to the promotion of the re-use and sharing of health data across Europe that will be the points of action for delivering FAIR data implementations. Key points are:

- Standardized and interoperable data selection
- Long-term data stewardship
- Accessibility (by both person and machine)
- Legal interoperability
- Timeliness of sharing

Based on these priorities, any integrated healthcare data, research outputs, or results can be subject to a FAIRification process [7].

This process of FAIRifying the healthcare data, (more precisely Electronic Health Records) and the implementation of technical components needed to support the process of FAIRification are presented as follow but were originally described in the recommendations of the GO FAIR initiative (Figure 1). The seven distinct points to consider as follows:

1. **Creation** of a FAIR provider registry to uniquely identify digital objects made available by the data provider (e.g., healthcare datasets, publications)
2. **Creation** of **FAIR digital objects** (FDOs) from EHR system extracts
3. **Creation** of **local FAIR data repositories** (FDRs) for digital objects hosted by each data provider
4. **Deployment** of FDOs to FDRs.
5. **Definition** of access policies and licenses to control data sharing and wider access to FDRs & FDOs
6. **Assessment** of FDRs based on FAIR maturity models, certification standards and FAIR metrics.
7. **Development** and deployment of supporting tools and services to allow search, browsing, and controlled access to FDOs between providers and third parties.

It is clear from this description of the FAIRification process that the technical implementation of the FAIR concept in its entirety and at scale requires several separate software components that collectively work together to deliver FAIRificiation of data in practice. These important sets of individual but integrated software components are required to implement FAIR data objects and FAIR data repositories. We propose three most important set of tools namely 1) FAIR Data Object Annotation Tool, 2) FAIR Data Object Manager and 3) Local FAIR Data Repository Manager as described below:

- **FDO Annotation Tool:** A software that will allow health data providers to describe their FDOs using metadata tags in a standardized way using accepted ontologies such as use of appropriate clinical terminologies to provide standardized semantic interoperability and coding for FAIR data (most notable examples but not limited to these are ICD [3], LOINC [8], SNOMED CT [9], etc.). Other proposed standard ontologies include Human Phenotype Ontology [10], Online Mammalian Inheritance in Man (OMIM) [11], ORPHANET [12], providing for cross disciplinary phenotypical and genetic interpretation and understanding of how to use, interpret and link to the FAIR data for research purposes. It is also proposed that the Metadata annotation should be based on the structure of the TRIPOD guidelines [13] for predictive model development and validation to fully describe the research datasets according to recognized best practice.
- **FDO Manager:** A software component for creation, packaging, storage, and curation of FDOs, along with their unique identifiers, dataset description, provenance metadata, licensing and Data Management Plans (DMPs) to local FAIR Data Repositories. The HL7 FHIR [14] for FAIR implementation standard is an example of standards to support the implementation of FAIR repositories.
- **Local FDR Manager:** A software tool for managing local FDRs at each health data provider. This will allow for managing security, governance, and availability of local FDRs to participants and third parties. This will allow for searching and creation of FAIR repositories and making FAIR digital objects available and searchable as collections of publically accessible REST API endpoints. The HL7 FHIR for FAIR implementation standards will be used and extended as needed to implement FAIR repositories.

The pictorial depiction of aforementioned tools, software components and the way these individual components are connected can be seen in Figure 1 below.

It is evident that the delivery of a FAIR data ecosystem that supports sharing of EHR data is therefore not simply a case of adopting health interoperability standards. It must also be supported by a broader ecosystem that uniquely identifies the EHR resources available, makes them searchable, discoverable, and most importantly controls access to only those third parties who are authorized to use those resources.

6. Discussion (Challenges for FAIRifying EHRs)

In the multiple clinical scenarios we have mentioned in this paper, it can be deduced that EHRs, if accessible and interoperable across healthcare modalities, can benefit immensely any given healthcare system in terms of cost reduction by avoiding multiple testing and multiple presentations, as well as by enhancing patient safety. Data collected from example studies of rare genetic disorders or clinical presentations can be consolidated from more than one location and be built upon for clinical studies with more ease.

The original focus for developing FAIR data standards and a broader FAIR ecosystem was cantered around the sharing and transparency of research-related datasets[5]. By their nature, research datasets contain finite datasets that are needed to answer specific research questions guided by the research questions that the study is investigating. Research datasets are also

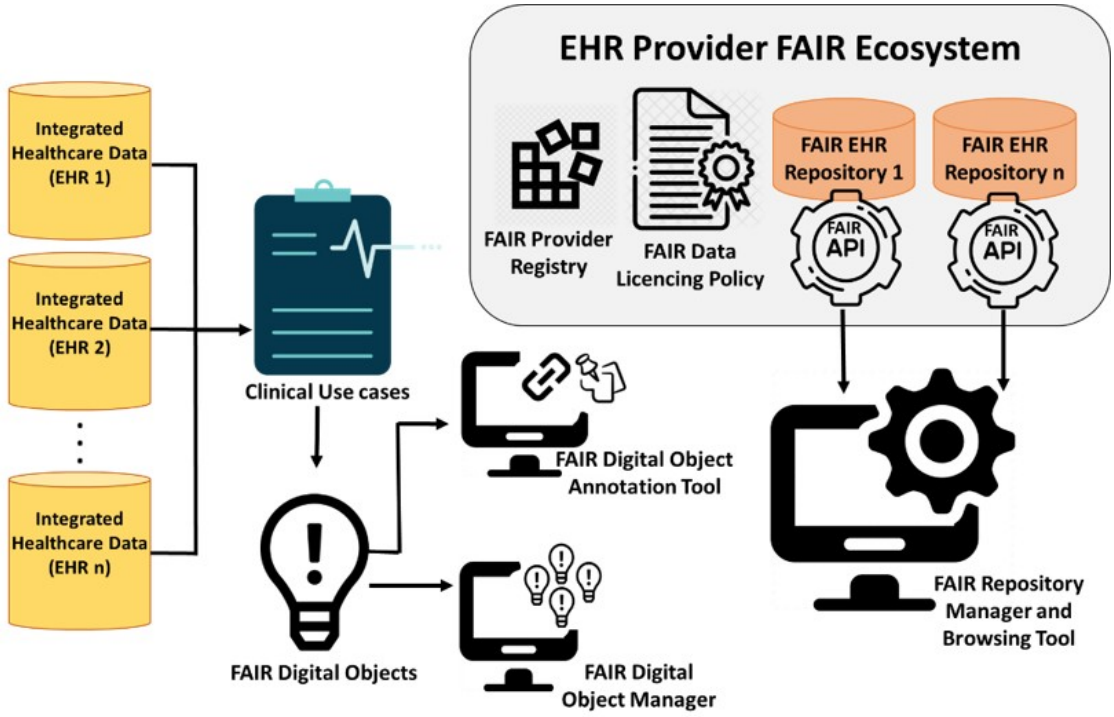

Figure 1: EHR FAIRification proposed architecture

largely expected to be static in nature, unless they are subsequently revised or corrected due to inaccuracies found in the original research. These datasets are captured at a point in time during the conduct of the research and may be reported to support the publication of the research study results. The concept of retaining the integrity of data in research is fundamental and consistent with FAIR data principles where data is published and has a unique resource identifier associated with that static dataset forever, supporting the publication of results which cannot subsequently be changed.

On the contrary, the nature of EHR data is fundamentally different. EHR data typically have very wide coverage of collected data. This data may potentially be captured to support decision making across a much wider spectrum of clinical care including phenotypic data, diagnostics, therapeutics, imaging, and genetics. In any discussion about FAIRifying EHR data we therefore need to be clear about whether it is feasible to FAIRify an entire complex EHR in its entirety, or whether several defined subsets of that EHR data are being constructed and FAIRified separately. The data complexity of a typical EHR system suggests that a more manageable approach would be to extract several data subsets from an entire EHR that can be FAIRified to support the defined needs of research or real-world-evidence creation. In the context of FAIR data, we potentially therefore have multiple FAIR datasets created from a single data source that would appear to have some sort of relationship or linkage between them.

EHR data collected in frontline clinical care is also highly dynamic with changes to the underlying data occurring frequently during a single day. The creation of dataset extracts

as part of a FAIRification process is therefore only reflecting a point-in-time snapshot of a defined subset of the overall EHR data. If multiple FAIR datasets are required to reflect the data complexity of the EHR as mentioned previously then it only makes sense to create them all at the same point in time to maintain integrity between the different related sets of data.

A final distinction needs to be made between research datasets and EHR systems. The data captured as part of the conduct of research studies is typically subject to ethical approval and patient consent to gather such data. The ethical constraints under which EHR data that is captured for the primary purposes of routine clinical care may not be so explicitly defined. Under General Data Protection Regulation (GPDR) [15] guidance, this raises questions about secondary processing of such data and the legal basis for onward data sharing to support the conduct of research or real-world-evidence generation. In the absence of explicit patient consent for onwards data sharing for explicitly informed research purposes, this suggests that anonymization of EHR data is a prerequisite for FAIRification of EHR data which incurs an additional data processing overhead.

These distinctions in complexity, dynamic nature, and ethical approval relating to EHR data suggest that the FAIR data concept was not originally defined to support the concept of sharing of data from frontline clinical care systems in its current form. The concept of 'Findability' of data becomes problematic in this context where each data resource should have a uniquely defined and unchanging identifier that defines it as a uniquely identifiable and shared data resource. It is not clear if multiple EHR extracts should have a single identifier and whether that identifier should dynamically change with each snapshot of data that is extracted from the EHR system [16].

One potential solution, according to Stein et al, [17], that is being used is to associate a unique identifier of an associated research publication based on the underlying FAIR data source and to quote that as the findable resource[3]. This is a workaround rather than a solution to this problem. A proper solution would suggest a different type of resource identifier can be used for these more dynamic and complex datasources that has a core unique identifier that can be adjusted with some sort of time stamp with a common root that indicates where different datasets are related from the same core underlying datasource.

7. Future Directions

The FAIR data concept is still evolving and is now being more universally adopted in the context of sharing data generated from other areas beyond static research datasets. This has been noted as being problematic while indicating a gap in the context of applying FAIR data principles to more dynamic and larger (perhaps linked) datasets that collectively expose EHR data for the purposes of real world evidence more generally [16]. However, this relative limitation of the FAIR data concept and the eagerness of organizations to demonstrate 'open data' credentials has led to differing interpretations of the degree to which there is adherence to FAIR principles, which seems to vary across different data repositories. This suggests that a mutual understanding of minimum requirements is needed, as this has become an ever-pressing need [18].

[3] 4DN Data Portal https://data.4dnucleome.org/help/user-guide/faq. (last accessed 10-03-2022)

Authors also believe that this will emerge and become clearer as expert groups from the EU have been established to measure adherence to and examine alternative metrics some of which may be appropriate for measuring compliance with FAIR data principles [7, 19, 20].

The expert group stated: "A major additional challenge in the data domain is the adoption of a new set of metrics to assess FAIRness, i.e., compliance with the FAIR principles. We propose the following as a basic minimum standard: discovery metadata, persistent identifiers and access to the data or metadata". It will be important to standardize FAIR metrics globally and to coordinate initiatives and some are under way to develop FAIR maturity models or assessment tools [7, 21, 20].

8. Conclusion

FAIR data is a set of principles that provides the guidelines and recommendation for data being findable, accessible, interoperable, and reusable. Applying FAIR data principles to electronic health records is an evolving process. The complex and dynamic nature of medical data, as well as the ethical complexities that do not critically apply to records in any other field, has demonstrated that more dimensions and associated metrics will be needed to be applied to electronic records if they will be interoperable and accessible. Nevertheless, taking the timely decision the EU has already set up committees to address the feasibility, possibilities and complications of the same which is a judicious step in the right direction.

References

[1] H.-J. Kong, Managing unstructured big data in healthcare system, Healthcare informatics research 25 (2019) 1–2.

[2] S. Liu, Y. Wang, N. Hong, F. Shen, S. Wu, W. Hersh, H. Liu, On mapping textual queries to a common data model, in: 2017 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2017, pp. 21–25.

[3] P. Vezyridis, S. Timmons, E-infrastructures and the divergent assetization of public health data: Expectations, uncertainties, and asymmetries, Social Studies of Science 51 (2021) 606–627.

[4] S. Stall, L. Yarmey, J. Cutcher-Gershenfeld, B. Hanson, K. Lehnert, B. Nosek, M. Parsons, E. Robinson, L. Wyborn, Make scientific data fair, Nature 570 (2019) 27–29.

[5] S. Collins, F. Genova, N. Harrower, S. Hodson, S. Jones, L. Laaksonen, D. Mietchen, R. Petrauskaitė, P. Wittenburg, Turning fair into reality: Final report and action plan from the european commission expert group on fair data (2018).

[6] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.

[7] A. Hasnain, D. Rebholz-Schuhmann, Assessing fair data principles against the 5-star open data principles, in: The Semantic Web: ESWC 2018 Satellite Events: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers 15, Springer, 2018, pp. 469–477.

[8] C. J. McDonald, S. M. Huff, J. G. Suico, G. Hill, D. Leavelle, R. Aller, A. Forrey, K. Mercer, G. DeMoor, J. Hook, et al., Loinc, a universal standard for identifying laboratory observations: a 5-year update, Clinical chemistry 49 (2003) 624–633.
[9] K. Donnelly, et al., Snomed-ct: The advanced terminology and coding system for ehealth, Studies in health technology and informatics 121 (2006) 279.
[10] S. Köhler, N. A. Vasilevsky, M. Engelstad, E. Foster, J. McMurry, S. Aymé, G. Baynam, S. M. Bello, C. F. Boerkoel, K. M. Boycott, et al., The human phenotype ontology in 2017, Nucleic acids research 45 (2017) D865–D876.
[11] A. F. Scott, J. Amberger, B. Brylawski, V. A. McKusick, Omim: Online mendelian inheritance in man, Bioinformatics: Databases and systems (1999) 77–84.
[12] S. S. Weinreich, R. Mangon, J. Sikkens, M. E. Teeuw, M. Cornel, Orphanet: a european database for rare diseases, Nederlands tijdschrift voor geneeskunde 152 (2008) 518–519.
[13] G. S. Collins, J. B. Reitsma, D. G. Altman, K. G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod) the tripod statement, Circulation 131 (2015) 211–219.
[14] D. Bender, K. Sartipi, Hl7 fhir: An agile and restful approach to healthcare information exchange, in: Proceedings of the 26th IEEE international symposium on computer-based medical systems, IEEE, 2013, pp. 326–331.
[15] P. Regulation, General data protection regulation, Intouch 25 (2018) 1–5.
[16] M. Löbe, F. Matthies, S. Stäubert, F. A. Meineke, A. Winter, Problems in fairifying medical datasets., in: MIE, 2020, pp. 392–396.
[17] A. Stein, A. Ceol, P. Aloy, 3did: identification and classification of domain-based interactions of known three-dimensional structure, Nucleic acids research 39 (2010) D718–D723.
[18] A. Dunning, M. De Smaele, J. Böhmer, Are the fair data principles fair?, International Journal of digital curation 12 (1970) 177–195.
[19] J. Wilsdon, J. Bar-Ilan, R. Frodeman, E. Lex, I. Peters, P. Wouters, Next-generation metrics: Responsible metrics and evaluation for open science (2017).
[20] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C. T. Evelo, et al., Fair principles: interpretations and implementation considerations, 2020.
[21] N. Krans, A. Ammar, P. Nymark, E. Willighagen, M. Bakker, J. Quik, Fair assessment tools: evaluating use and performance, NanoImpact 27 (2022) 100402.