

FAIR Versus Open Data: A Comparison of Objectives and Principles

Putu Hadi Purnama Jati^{1,2†}, Yi Lin¹, Sara Nodehi¹, Dwy Bagus Cahyono¹, Mirjam van Reisen^{1,3,4}

¹Leiden University, 2333 CA Leiden, the Netherlands

²Politeknik Statistika STIS

³Tilburg University, P.O. Box 90153 5000, the Netherlands

⁴Leiden University Medical Centre (LUMC), Leiden University, 1310 Leiden, the Netherlands

Keywords: FAIR Guidelines; FAIR data; Open data; Data management

Citation: Purnama Jati, P.H., Lin, Y., Nodehi, S., Cahyono, D.B., Van Reisen, M.: FAIR versus open data: A comparison of objectives and principles. *Data Intelligence* 4(4), 867–881 (2022). doi: 10.1162/dint_a_00176

Submitted: March 10, 2021; Revised: June 10, 2022; Accepted: July 15, 2022

ABSTRACT

This article assesses the difference between the concepts of ‘open data’ and ‘FAIR data’ in data management. FAIR data is understood as data that complies with the FAIR Guidelines—data that is Findable, Accessible, Interoperable and Reusable—while open data was born out of awareness of the need to democratise data by improving its accessibility, based on the idea that data should not have limitations that prevent people from using it. This study compared FAIR data with open data by analysing relevant documents using a coding analysis with conceptual labels based on Kingdon’s theory of agenda setting. The study found that in relation to FAIR data the problem stream focuses on the complexity of data collected for research, while open data primarily emphasises giving the public access to non-confidential data. In the policy stream, the two concepts share common standpoints in terms of making data available and reusable, although different approaches are adopted in practice to accomplish these goals. In the politics stream, stakeholders with different objectives support FAIR data and from those who support open data.

ACRONYMS

AI artificial intelligence
BOLD big and open link data

[†] Corresponding author: Putu Hadi Purnama Jati, Leiden University (Email: putu.hadi.purnama.jati@umail.leidenuniv.nl; ORCID: 0000-0002-6533-3709).

LOD	linked open data
FAIR	Findable, Accessible, Interoperable, Reusable
VODAN	Virus Outbreak Data Network

1. INTRODUCTION

The FAIR Guidelines—that data are Findable, Accessible, Interoperable, Reusable (FAIR)—are an important tool for data management worldwide. However, the philosophy of FAIR—that digital objects should be human and machine readable—is not a new idea. The scientific community, among others, well understands the benefits of a system that allows data to be findable, accessible, and reusable, as well as being interoperable, but those advocating for such an approach were previously not connected and there was no structured attempt to coordinate [1]. The FAIR Guidelines (then called the FAIR Guiding Principles) were first discussed in 2014 at a workshop in Lorentz, Leiden, which was attended by several stakeholders from the research community [2]. During this workshop, several standards were established to facilitate manual and automated filing, as well as the retrieval, sharing, and reuse of data. Since then, the FAIR Guidelines have been rapidly developed and accepted by various organisations, including the European Union, G7, G20, Big Data to Information (BD2K) from the United States, and African Science Cloud [3]. These Guidelines have been widely implemented in European regions (67%), and to a lesser extent in the Americas (14%), together accounting for 81% of all implementation activities [3]. This broad acceptance of FAIR Guidelines is due to the ability of FAIR to be implemented in any situation or by any organisation. As FAIR is not a standard for data management, it is not necessary to comply with all of the guidelines to implement FAIR.

Before the FAIR Guidelines, several concepts recognised the importance of machine-readable data that changes the way data flows in a web application. The Semantic Web approach originated from the need to solve the problem of current knowledge management systems, which have significant weaknesses in relation to searching information, extracting information, maintaining weakly structured text sources, and automatic document generation. The Semantic Web is conceived as an extension of the current Web, where documentation is annotated with meta-information and is human and machine-readable using World Wide Web browsers. This metadata defines the information (documents) to be processed by the machine [4]. In recent years, the Web has evolved out of a global information area for linked documents into a link between documents and data. A set of best practices underpins this development to publish and connect structured data over the Web, which is called linked data.

Linked data uses a site to link information from different sources. These sources can be as different as databases maintained by two organisations or heterogeneous systems within an organisation that historically have not interoperated easily at the data level [5]. One of the implementations of linked data that is used for data privacy is the Solid project. Solid is a decentralised platform for social web applications that manages user data independently of the applications that create and consume it. The user data is saved in a personal online datastore accessible via the Internet (or pod). The Solid platform grants users one or more pods from various pod providers, while also allowing users to switch between providers [6].

Another principle, called ‘open data’, has also gained worldwide public attention, especially in the public sector. The term open data refers to non-private and non-confidential data made available through public means, without restricting its use or dissemination [7]. An Open Data Barometer study showed that 55% of the countries surveyed now have an open data initiative and a domestic data catalogue that provides access to reusable datasets [8]. New open data projects have been launched or announced in many countries, including Jamaica, Ecuador, Saint Lucia, Nepal, Thailand, Botswana, Ethiopia, Uganda, and Rwanda [8]. The popularity of open data in the public sector is primarily motivated by the increased return it generates on publicly-funded data, the generation of wealth through the downstream use of output, and the provision of necessary data for policymakers to solve complex problems. Open data is gaining momentum globally due to the trend towards transparent data management across sectors and the flexibility in applying its principles in private and public sectors.

Until now, far too little attention has been paid to the comparison between FAIR and open data. For example, in the health arena, the World Health Organization (WHO) found that the lack of guidelines on both technology and content has led to a wealth of digital solutions with unknown data and health content, which undermines confidence, which is essential for government and donor investment; prevents the localisation of countries; and prevents the interoperability of healthcare continuity, optimal data utilisation, and accountability [9]. FAIR and (or) open data could be used to optimise health data management. Both principles have some similarities, but they are not identical. Hence, the research question addressed in this article is: What is the difference between FAIR and open data? The objective of this study was to compare FAIR and open data, using Kingdon’s multiple streams theory of agenda setting as a framework for analysis.

2. THEORETICAL FRAMEWORK

Kingdon’s multiple streams theory of agenda setting was used for this study to distinguish between FAIR and open data. Kingdon’s theory purports that agenda setting involves three separate streams—a problem stream, policy stream, and politics stream—which, when joined together, open a window of opportunity for a problem (solution) to reach the policy agenda [10, 11].

The first stream, the problem stream, refers to the concerns, problems, or difficulties that have attracted society’s attention. The policy stream represents the policy options available to researchers, stakeholders, and governing bodies that propose to address the problem. The problem may be brought to the attention of decision makers by monitoring indicators or current legislation, interest groups (e.g., the medical profession), the media, or real events. The people engaging with the policy agenda are referred to as ‘policy entrepreneurs’, and these can be both within the administration and outside of it [12]. The problem must be seen as a public problem that should prompt policy and decision makers to act or find a solution (policy stream). The politics stream refers to policy changes, unique national circumstances, and social constraints affecting the definition of the problem and identification of the policy/solution.

All three elements work largely independently of each other, although there may be some overlap between the streams. When the three streams come together, a policy window opens [11]. This theory

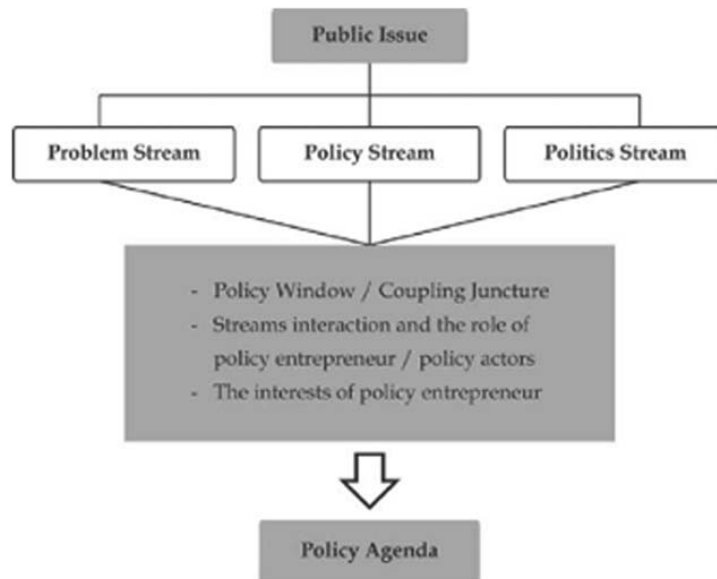


Figure 1. Kingdon’s multiple streams model [11].

clarifies how agenda setting works in the United States, with the three types of independent (and interdependent) variables that interact to create a ‘windows of opportunity’ for agenda setting [12], and has since been applied to problems in many policy fields and in many different circumstances.

3. METHOD

A desk study was undertaken to identify the relevant documents for analysis to compare FAIR and open data according to Kingdon’s multiple streams theory of agenda settings. A criterion for inclusion in the analysis was that the document needed to explain either of the principles. We did not use sources that apply FAIR or open data principles in a specific setting, because a new principle often emerges from the original principles when adapted to a specific setting, such as open government data.

To identify the relevant documents, we used Google Search and Google Scholar to search for documents containing the keyword ‘FAIR’ or ‘open data’. We listed all sources in a spreadsheet and created a three-column structure: ‘FAIR’, ‘open data’ and ‘general topic’. A document was selected if it contained FAIR or open data principles and explained the general idea of the principles (see Supplementary Material 1, Purnama Jati, 2021). We selected 17 out of 28 articles on FAIR and open data for analysis.

A qualitative research and exploratory method was applied by analysing 11 of the documents from the literature review to make a comparison between FAIR and open data. The next step was to identify the aspects of agenda setting mentioned in each document using Kingdon’s theory. The extraction of the content of the documents was done using content analysis, a tool for analysing written, verbal or visual messages.

This approach systematically defines, quantifies and describes the phenomena [13]. To improve data interpretation, content analysis allows researchers to explore theoretical issues [14]. Words can be simplified into smaller content groups through content extraction. Thus, names, phrases, and the like have the same value when grouped in the same class [15].

Open coding, axial coding and selective coding were used as the basis for the analysis of the documents. Open coding was used to create a preliminary marker that summarises the document's content. Then axial coding was undertaken to identify the relationships between the open codes. Finally, for selective coding, we categorised axial coding based on Kingdon's theory of agenda setting. In this process, we identified the problem highlighted by a particular stakeholder in FAIR and open data. The second stream is the policy stream, which consists of proposals or alternative solutions to this problem. In this stream, a number of ideas are discussed, which do not necessarily pan out as solutions. The final stream is the policy stream, which covers factors that influence the policy body, including national mood changes, executive or legislative changes, and stakeholder advocacy initiatives [16]. Identifying the stakeholders in FAIR and open data is important to understand who is driving the agenda that opens the policy window to put these two principles on the policy agenda. We compared all three streams for FAIR and open data and highlighted the differences, with the aim of identifying a significant separation between the two principles, to clearly establish when it is most appropriate to apply FAIR and to apply open data.

4. RESULTS: COMPARISON BETWEEN FAIR AND OPEN DATA

4.1 Problem Stream

The current digital ecosystem is insufficient to reap the full benefits of research. Researchers often require several weeks (or months) of advanced technical effort to collect the data they need. This is not due to lack of appropriate technology, but due to the fact that we do not pay sufficient attention to digital objects when creating and preserving them [17]. To overcome this, improved data processing and management are needed.

The concept of FAIR was developed to maximise investments in global research, by ensuring that datasets and other research artefacts (e.g., workflows) that originate from traditional science and do not serve a specific purpose are considered research objects of 'first-class' value [16]. In general, the problem potentially solved by FAIR is the inability of machines to automatically find and read data, which makes it challenging for the data to be reused by any stakeholder.

In a well-functioning democratic society, the public need knowledge and access to government policies and development information [8]. To address this and fulfil the need for transparency and accountability in government, the concept of open data was developed. Open data is defined by Geiger and von Lucke as making accessible "all stored data of the public sector which could be made accessible by the government in the public interest without any restrictions on usage and distribution" [17]. However, this definition contains unclear statements, such as 'public interest', without specifying how 'public interest' is defined. Open data is defined more generally as non-private, non-confidential data made available for publicly

accessible use or distribution without restriction. Private, confidential, and classified data are excluded, because it is not appropriate to publish such data [7]. Similarly, the *Open Data Handbook* defines open data as “data that can be freely used, reused, and redistributed by any person, subject to attribution requirements at most, and must be redistributed in the same manner in which it appears” [18]. Although there is no unified definition of open data, the primary goal of the concept is to provide data that can be freely used, reused, and redistributed by anyone, while at the same time maximising interoperability [19].

Despite an obvious overlap between these concepts, both which aim to enhance the (re)usability of data, there are several differences. The origin of FAIR comes is the research environment, where the difficulties of collecting data create barriers to maximising the benefits of investments in research. However, the focus of open data is on giving the public access to data and distributing data that is believed to be in the public interest in terms of transparency and democratic control, improved or new private products and services, improved efficiency and effectiveness of government services, or new knowledge from combined data sources and patterns in large data volumes [19]. Open data tends to be more concerned with the accessibility of data to the public, as long as the data is not confidential. In contrast, FAIR is more focused on how to appropriately create and process data to increase the usability of data for research. Making data accessible to the public in open data means that the data provided should be non-confidential. This is different from FAIR, which does not explicitly mention the type of data or users who access the data. FAIR data could be public or confidential data, and the stakeholders/users of this data are not necessarily confined to the public.

In general, therefore, it seems that these two principles do not have the same problem streams. In FAIR, the problem stream focuses on the complexity of data collected for research and making it findable, accessible, interoperable and reusable, which requires better data management. In contrast, open data primarily emphasises giving the public access to non-confidential data, enabling them to participate in data redistribution, and making data reusable in the public interest.

4.2 Policy Stream

FAIR requires data, whether confidential or non-confidential, to be findable, accessible, interoperable, and reusable. ‘Findability’ means that data can be easily found by machines and humans. Metadata is critical for this, as it is important for automatically discovering records and services. ‘Accessibility’ means that authentication and verification should be possible after the user accesses the data. ‘Interoperability’ requires data to be able to be integrated with other data and systems or workflows for analysis, storage and processing, so that it is interoperable. And the last element of FAIR, ‘Reusability’, states that metadata and data should be able to be reused, replicated and mixed in different environments [20].

The FAIR Guidelines are not a standard and do not propose a specific technology to be applied in data management. Rather, the guidelines are a precursor to implementation and serve as a guide for data publishers and managers to determine if their digital artefact is ‘FAIR’ [2]. FAIR is flexible and adaptable, which means that it can be applied in any type of organisation or environment.

In reviewing the literature, no principle was proposed for association with open data. Open data is more of a philosophy than a guiding principle. However, the summary definition of ‘open data’ states that it should have the following characteristic [18]:

- The data must be available in its entirety and at a low reproduction cost, preferably via Internet download. The data should also be available in a functional and modifiable way (availability and access).
- The data should be made available for reuse and redistribution, including mixing with other datasets (reuse and redistribution).
- Anyone should be able to use, reuse and redistribute the data—there should be no discrimination in terms of the areas, individuals or groups that may use/distribute it, including, for example, ‘non-commercial’ restrictions that prevent ‘commercial’ use, or restricting use for certain purposes (e.g., educational use only). In other words, to be open, universal participation should be allowed.

The similarities and differences between FAIR and open data are summarised in Table 1. According to the summary, two intersections are identified in two main areas: (i) the FAIR Guidelines ‘Findability’ and ‘Accessibility’ and the first open data principle; and (ii) the FAIR Guideline ‘Reusability’ and the second and the third open data principles. The FAIR Guideline ‘Interoperability’ has no counterpart in open data, as it is not explicitly facilitated in the open data format.

The first open data principle—data must be available in its entirety at a low reproduction cost—shares a highly similar functional purpose with the FAIR Guidelines of ‘Findability’ and ‘Accessibility’, which emphasise that data should be retrieved in an efficient way as the first step. But we notice that the details of both principles regarding this aspect are somewhat different. For instance, the availability of data in open data refers to data integrity, without specific mention of further conditions, whereas ‘Accessibility’ under FAIR highlights the need for data protection related to the circumstances in which the data is produced. This differs from open data in that data restriction is against the concept of open data, as it excludes people from reusing data [19]. The ‘Accessibility’ principle protects content, by requiring access and authorisation protocols for users [21]. The remaining aspects of open data show a strong bond with the FAIR Guideline of ‘Reusability’, as both concepts have the identical purpose of making data reusable. The open data principles particularly strengthen the need for redistribution neutrality, from the point of view of both data generators and data processors, while the FAIR Guideline ‘Reusability’ highlights the importance of using metadata as a way of making data reusable. The FAIR Guidelines require data and metadata to be expressed in a formal, accessible, shareable, and broadly applicable language, especially ontologies expressed in RDF to drive data integration at all levels. By establishing semantic data in such a manner, a transformation from raw data to highly processed data is realised [21].

The two concepts share common standpoints in terms of making data available and reusable, although different approaches are adopted in practice to accomplish these goals. Open data focuses on no barriers to data accessibility, as well as a high level of decentralisation and redistribution of data, while the FAIR Guidelines highlight the need to protect data ownership and data protection relevant to the place where

Table 1. Comparison between FAIR and open data.

No	FAIR	Open data	Similarity	Difference
1	Findable: Data can be easily found by machines and humans.	Data must be available in its entirety and at a low reproduction cost.	Findability and accessibility in FAIR is quite similar to	Availability in open data refers to data integrity and does not mention other conditions. Open data focuses on no barriers to
2	Accessible: Authentication and verification is possible after the user accesses the data.		availability in open data, because it provides accessibility to the user.	data accessibility, while accessibility in FAIR principles highlights the need for data protection and the conditions for access to be formulated to meet the specific circumstances that relate to the data.
3	Interoperable: Data can be integrated with other data and systems or workflows for analysis, storage and processing are interoperable.		-	Interoperability is promoted by the creation of machine-readable instances of ontologies that the data represent, linked to metadata in languages such as JASON or RDF, widely used for the Semantic Web.
4	Reusable: Metadata and data should be defined for reuse and can be replicated and/or mixed in different environments.	<ul style="list-style-type: none"> The data should be made available under the condition of reuse and redistribution. Anyone should be able to use, reuse and redistribute the data—there is no discrimination based on the purpose for which the data is to be used or the individuals/groups wishing to use it. 	Both principles have the purpose of making data reusable.	Open data does not mention metadata and focuses on redistribution neutrality.

data is produced, by introducing the use of metadata based on proper ontology to allow data visiting based on specifically identified conditions. Open data invites free portability, as the data should, in any case, be available to anyone. The FAIR Guidelines invite data ownership, as the regulations and privacy concerning the data are recognised as relevant to the life of the data, in association with where and for what purpose the data is produced.

The promise of strengthening decision making with digital data is now widely accepted, and the notion of ‘big data’ is growing in popularity. Big data means large amounts of data that need new technologies and architectures to extract value from them through capture and analysis [22]. Both FAIR and open data support big data. Big data and open data, known as ‘big and open link data’ (BOLD), can transform public government and public relations and create new opportunities for governments [23]. The combination of

linked data with open data is also known as linked open data (LOD). An example of LOD is an RDF database like GraphDB ontotext. GraphDB can handle large datasets from different sources and connect them to open data, enhancing knowledge discovery and efficient analytics driven by data [24].

In addition, open data is also crucial to enable artificial intelligence (AI), which can extract deep insights from datasets, because AI systems need big data to function properly. If the data is not in some way open or accessible, it is impossible to reuse the data for other purposes, like AI. Open data access ‘unlocks’ the potential for the AI applications of data-hungry machines [25]. The FAIRification process recommends Semantic Web and linked data technologies to transform non-FAIR data into linkable data. Unlike open data, the accessibility in FAIR means that FAIR does not have to be ‘open’, and not all open data is FAIR, as stated in the motto “as open as possible, as closed as necessary” [21]. FAIR itself does not mandate that all data be openly accessible, but ensures that funded research is made as open as possible and as closed as necessary. FAIR offers powerful new AI analysis access to data for machine learning and prediction. Implementing the FAIR Guidelines as a data management strategy results in several improvements, including the possibility of machine readability (data and metadata) for robotics, digital twin technology and process automation, allowing for reutilisation and scalability. In this way, the complete process of operating data through the acquisition, semantic alignment, integration and analytics chain is simplified and, therefore, more efficient in terms of generating insights [26].

4.3 Politics Stream

Kingdon defines politics as factors that affect the body politic, such as national mood swings, changes in the executive or legislative branches, and lobbying campaigns by interest groups [12]. This section compares FAIR Guidelines with the concept of open data in terms of how both relate to the politics stream. In other words, how receptive the political environment is to these concepts, as one of the factors that play a role in the opening of a policy window. In order to do this, we briefly discuss the context of their origin and evolution, as well as the stakeholders who are pushing the agenda.

The concept of open data emerged over a decade ago in 2007, when a group of 30 Internet thinkers and activists gathered in Sebastopol, north of San Francisco, to discuss the need for transparency and accountability in data. The goal was to recognise and implement the idea of open public records of presidential candidates in the United State [27]. This event established the principles that now allow us to adopt and evaluate open public records. The basic idea was to make public data common property. Hence, the main stakeholders pushing the open data agenda are Internet thinkers and activists. Open data is also supported by the Open Knowledge Foundation, which works with both individuals and organisations with a mission to create a more open world in which all non-personal information is open, free for everyone to use, build on and share [28].

The FAIR Guidelines were first discussed in 2014, during a workshop in Leiden, at which members of the research community discussed the need for data management to enable researchers to more effectively use/reuse data. Following this workshop, the principles were formulated and reformulated many times

before being published in 2016 as a commentary in the journal *Scientific Data* [2]. Since then, this agenda has been embraced by researchers, data publishers, tool makers, funders, and the data science community: researchers want to distribute, get credit for, and reuse data and interpretations; professional data publishers offer their services; tool makers develop data analysis and processing services that offer software and reusable workflows; and funders (private and public) increasingly focus on the long-term stewardship of data [2]. The FAIR Guidelines are supported by GO FAIR, a project seeking to promote the implementation of the FAIR Guidelines, which consists of individuals and institutions working together, through so called Implementation Networks (INs) [29].

This analysis found that FAIR differs from open data from a political perspective. The policy entrepreneurs for FAIR data are from particular groups, such as researchers, data publishers, and the data science community, who seek to improve the process of data stewardship for better data analysis and data processing. Unlike the FAIR Guidelines, there does not seem to be a specific group advocating for open data that has more interest in the concept than any other group. It appears that everyone who advocates for open data is themselves an expert in data for various purposes of analysis and policy action [30]. The literature did not describe which type of group first discussed the idea of open data. A possible explanation for this could be that the term ‘public’ is not mentioned as a specific topic. This means that open data was born out of the need for the democratisation of data, supported by Internet thinkers and activists. The main driver of this group is to push the concept of free access to data, even though it is unclear which group has the most interest in open data. Unlike open data, the motivation for FAIR is to optimise data management to improve the reusability of data. The FAIR concept is supported by various policy entrepreneurs, dominated by researchers, data publishers, and data scientists.

5. DISCUSSION AND CONCLUSION

This research aimed to explore the difference between FAIR Guidelines and open data in order to consider both principles when implementing them in a particular field or organisation. From the perspective of Kingdon’s agenda setting, we conclude that a clear distinction can be made between the concepts of FAIR data and open data in all three streams: the problem, policy, and politics streams. In relation to the problem stream, the difference between FAIR and open data is that FAIR emerged in response to a problem in the research environment related to the need to be able to reuse research data, which has value (and such data is often lost or unable to be easily retrieved or located), while open data focuses on democratising data by ensuring that it can be freely used, reused, and redistributed by anyone—making all non-confidential data open and free for everyone to use, build on and share.

In the policy stream (or the solution), the philosophy of FAIR data has four elements—data should be Findable, Accessible, Interoperable, and Reusable—which act as a guide for data management, but do not suggest a specific technology to be applied. In contrast, in the concept of open data, ‘open’ is defined as “anyone is free to access, use, modify, and share” the data for any purpose [31]. Interestingly, the goal of open data can be achieved if the first three FAIR Guidelines are followed. However, making data accessible under FAIR is not the same as making it ‘open’, as FAIR requires that data users be authorised and verified.

By making data accessible, FAIR does not aim to make it open to the public, but to regulate it according to the nature of the data (some data is by nature sensitive and access needs to be restricted). Open data is commonly used in the government domain to ensure the transparency and accountability of the government to the public. The term ‘public’ in open data should be further elaborated by underlining that the data is categorised as open if it is not confidential. Hence, both concepts are suggesting two different solutions, to two slightly different problems, but with some overlap.

Moving on to the politics stream, open data is about data democratisation for the public good, while FAIR is about getting the most value from scientific data. There are specific groups that focus on FAIR (researchers, data publishers, tool makers, funders, and the data science community), while the stakeholders in open data are those interested in making data available to the public. The spirit to give free access to public data is the primary driver of the open data concept, while optimising data reuse by creating machine-readable data is the central theme of FAIR.

Despite the great benefits of data being both FAIR and open, as the absence of restrictions supports the widest possible reuse and repurposing, we have gained a perspective on what FAIR and open data respectively facilitate. The problem stream between FAIR and open data could serve as a boundary separating objectives related to the use of FAIR and open data. If the focus of the implementation is data democratisation, data management may be more open than FAIR. On the other hand, if the aim is to maximise the value of data that involves confidential and private data, FAIR may be the best fit. The question raised by this result is: How can we make data both FAIR and open to achieve optimum benefits for data management? To answer this question future research is needed to compare FAIR and open data in more detail and investigate the implementation of both principles.

ACKNOWLEDGEMENTS

We would like to thank Misha Stocker for managing and coordinating this Special Issue (Volume 4) and Susan Sellars for copyediting and proofreading. We would also like to acknowledge VODAN-Africa, the Philips Foundation, the Dutch Development Bank FMO, CORDAID, and the GO FAIR Foundation for supporting this research. Lastly, we would like to give a special thanks to Prof. Dr Mirjam van Reisen, for being the editor of this Special Issue.

AUTHORS' CONTRIBUTIONS

Putu Hadi Purnama Jati (putuhadi2808@gmail.com, 0000-0002-6533-3709): Writing—original draft preparation, investigation, conceptualization. Yi Lin (yilin.shcn@gmail.com, 0000-0002-9833-3457): Writing—original draft preparation, resources, investigation. Sara Nodehi (sara.nodehi95@gmail.com, 0000-0002-2919-1336): Writing—review and editing, resources, investigation. Dwy Bagus Cahyono (dwy.bagus.cahyono@gmail.com, 0000-0002-1100-5742): Writing—review and editing, resources, investigation. Mirjam van Reisen (mirjamvanreisen@gmail.com, 0000-0003-0627-8014): Supervision, validation, conceptualization, project administration.

CONFLICT OF INTEREST

All of the authors declare that they have no competing interests.

ETHICS STATEMENT

Tilburg University, Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences REDC#2020/013, June 1, 2020-May 31, 2024 on Social Dynamics of Digital Innovation in remote non-western communities

Uganda National Council for Science and Technology, Reference IS18ES, July 23, 2019-July 23, 2023

REFERENCES

- [1] Purnama Jati, P.H.: Improving Satu Data Indonesia with FAIR elements: A model to extend Satu Data Indonesia Principles in COVID-19 data management. Unpublished Master's Thesis, Leiden University (2020)
- [2] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al.: Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), 1–9 (2016). doi: 10.1038/sdata.2016.18
- [3] Van Reisen, M., Stokmans, M., Basajja, M., Ong'ayo, A.O., Kirkpatrick, C., Mons, B.: Towards the tipping point for FAIR implementation. *Data Intelligence* 2(1–2), 264–275 (2020)
- [4] Davies, J., Fensel, D., Van Harmelen, F.: Introduction. In: J. Davies, *Towards the Semantic Web: Ontology-driven Knowledge Management*, Wiley, Chichester, pp. 1–9 (2003)
- [5] Bizer, C., Heath, T., Berners-Lee, T.: Linked data—The story so far. *International Journal on Semantic Web and Information Systems* 5(3), 1–22 (2009). doi: 10.4018/jswis.2009081901
- [6] Sambra, A.V., Mansour, E., Hawke, S., Zereba, S., Greco, N., Ghanem, A., Zagidulin, D., Aboulnaga, A., Berners-Lee, T.: Solid: A platform for decentralized social applications based on linked data. Paper (2016). Available at: http://emansour.com/research/lusail/solid_protocols.pdf. Accessed 18 December 2021
- [7] Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. *Information Systems Management* 29(4), 258–268 (2012), doi: 10.1080/10580530.2012.716740
- [8] Open Data Barometer: ODB global report (Third edition). World Wide Web Foundation, p. 48 (2016). Available at: <http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf>. Accessed 1 June 2021
- [9] Mehl, G. et al.: Comment: WHO SMART guidelines: Optimising country-level use of guideline recommendations in the digital age. *The Lancet* 7500(21), 8–11 (2021). doi: 10.1016/S2589-7500(21)00038-8
- [10] Babaey, F., Raessi, P., Ravaghi, H.: Agenda setting analysis for maternal mortality reduction: Exploring influential factors using Kingdon's stream model. *Eastern Mediterranean Health Journal* 25(3), 160–171 (2019). doi: 10.26719/emhj.18.025
- [11] Kingdon, J.W.: *Agendas, alternatives, and public policies*. HarperCollins, New York (1984)
- [12] Béland, D., Howlett, M.: The role and impact of the multiple-streams approach in comparative policy analysis. *Journal of Comparative Policy Analysis: Research and Practice* 18(3), 221–227 (2016). doi: 10.1080/13876988.2016.1174410

- [13] Krippendorff, K.: Content analysis: An introduction to its methodology (Second Edition). Sage Publications (2004)
- [14] Elo, S., Kyngäs, H.: The qualitative content analysis. *Journal of Advanced Nursing* 62(1), 107–115 (2008). doi: 10.1111/j.1365-2648.2007.04569.x
- [15] Cavanagh, S.: Content analysis: Concepts, methods and applications. *Nurse Researcher* 4(3), 5–16 (1997). doi: 10.7748/nr.4.3.5.s2
- [16] Hahnel, M., Valen, D.: How to (easily) extend the FAIRness of existing repositories. *Data Intelligence* 2(1–2), 192–198 (2020). doi: 10.1162/dint_a_00041
- [17] Barry, E., Bannister, F.: Barriers to open data release: A view from the top. *Information Polity* 19(1–2), 129–152 (2014). doi: 10.3233/IP-140327
- [18] Open Knowledge Foundation: What is open data? [Online]. Open Data Handbook (n.d.). Available at: <https://opendatahandbook.org/guide/en/what-is-open-data/>. Accessed 1 June 2021
- [19] Open Knowledge Foundation: Why open data? [Online]. Open Data Handbook (n.d.). Available at: <https://opendatahandbook.org/guide/en/why-open-data/>. Accessed 27 February 2021
- [20] Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., et al.: FAIR Principles: Interpretations and implementation considerations. *Data Intelligence* 14, 10–29 (2020). doi: 10.1162/dint_r_00024
- [21] European Commission: H2020 Programme: Guidelines on FAIR data management in Horizon 2020 Version 3.0. European Commission, Directorate-General for Research and Innovation Luxembourg (2016). Available at: <https://repository.oceanbestpractices.org/handle/11329/1259>. Accessed 28 June 2021
- [22] Katal, A., Wazid, M., Goudar, R.H.: Big data: Issues, challenges, tools and good practices. In: 2013 Sixth International Conference on Contemporary Computing, IEEE, pp. 404–409 (2013). doi: 10.1109/IC3.2013.6612229
- [23] Janssen, M., Van den Hoven, J.: Big and open linked data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly* 32(4), 363–368 (2015). doi: 10.1016/j.giq.2015.11.007
- [24] Ontotext: What are linked data and linked open data? [Online]. (n.d.) Available at: <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>. Accessed 1 June 2021
- [25] European Data: AI and open data: A crucial combination [Online]. DataEuropa.com (2018). Available at: <https://data.europa.eu/en/highlights/ai-and-open-data-crucial-combination>. Accessed 1 June 2021
- [26] Wise, J., Grebe de Barron, A., Splendiani, A., Balali-Mood, B., Vasant, D., Little, E., et al.: Implementation and relevance of FAIR Data Principles in biopharmaceutical R&D. *Drug Discovery Today* 24(4), 933–938 (2019). doi: <https://doi.org/10.1016/j.drudis.2019.01.008>
- [27] Chignard, S.: A brief history of open data [Online]. Paris Innovation Review (2013). Available at: <http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>. Accessed 29 December 2020
- [28] Open Knowledge Foundation: A fair, free and open future [Online]. (n.d.). Available at: <https://okfn.org/>. Accessed 1 June 2021
- [29] GO FAIR: GO FAIR Initiative [Online]. (2019). Available at: <https://www.go-fair.org/go-fair-initiative/>. Accessed 1 June 2021
- [30] Gurstein, M.B.: Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16(2) (January 2011). doi: 10.5210/fm.v16i2.3316
- [31] Open Knowledge Foundation: Defining open in open data, open content and open knowledge [Online]. (n.d.). Available at: <https://opendefinition.org/od/2.1/en/>. Accessed 30 December 2020

APPENDIX 1. LITERATURE REVIEWED

Author	Year	Title	FAIR	Open data	General topic	Decision
M. Janssen, Y. Charalabidis, and A. Zuiderwijk	2012	Benefits, adoption barriers and myths of open data and open government	No	Yes	Yes	Use
European Data	2018	AI and open data: A crucial combination	No	Yes	Yes	Use
S. Chignard	2013	A brief history of open data	No	Yes	Yes	Use
Ontotext	-	What are linked data and linked open data?	No	Yes	Yes	Use
J. Wise, et al.	2019	Implementation and relevance of FAIR Data Principles in biopharmaceutical R&D	Yes	No	Yes	Use
M. Janssen and J. Van den Hoven	2015	Big and open linked data (BOLD) in government: A challenge to transparency and privacy?	No	Yes	Yes	Use
A. Zuiderwijk and M. Janssen	2014	Open data policies, their implementation and impact: A framework for comparison	No	Yes	No	Use
P. Murray	2008	Open data in science	No	Yes	No	Not relevant
Open Knowledge Foundation	-	Defining open in open data, open content and open knowledge	No	Yes	Yes	Use
Open Knowledge Foundation	-	What is open data?	No	Yes	Yes	Use
Open Knowledge Foundation	-	Why open data?	No	Yes	Yes	Use
J. Molloy	2011	The open knowledge foundation: Open data means better science	No	Yes	No	Not relevant
W.D. Sunindyo and I.D.P.D.K. Amrita	2019	The development of data publishing tool for Indonesian open government data	No	Yes	No	Not relevant
C.M.L. Chan	2013	From open data to open innovation strategies: Creating e-services using open government data	No	Yes	No	Not relevant
A.S. Correa, P.L.P. Correa, D.L. Silva, and F.S.C. da Silva	2014	Really opened government data: A collaborative transparency at sight	No	Yes	No	Not relevant
H. Tamimi, S.A. Hoshang, and E.J. Al Blooshi	2017	Analysis of UAE open government data usability within mobile application development	No	Yes	No	Not relevant
M.B. Gurstein	2011	Open data: Empowering the empowered or effective data use for everyone?	No	Yes	Yes	Use
OECD	2015	Open government data	No	Yes	No	Not relevant
E. Barry and F. Bannister	2014	Barriers to open data release: A view from the top	No	Yes	Yes	Use

Downloaded from http://direct.mit.edu/infor/article-pdf/44/4/867/2063756/dint_a_00176.pdf by guest on 24 April 2025

Author	Year	Title	FAIR	Open data	General topic	Decision
A. Immonen, M. Palviainen, and E. Ovaska	2014	Requirements of an open data based business ecosystem	No	Yes	No	Not relevant
G. Smith, H.A. Ofe, and J. Sandberg	2016	Digital service innovation from open data: Exploring the value proposition of an open data marketplace	No	Yes	No	Not relevant
J. Lindman, T. Kinnari, and M. Rossi	2014	Industrial open data: Case studies of early open data entrepreneurs	No	Yes	No	Not relevant
P.H. Purnama Jati	2020	Improving Satu Data Indonesia with FAIR elements: A model to extend Satu Data Indonesia principles in COVID-19 data management	Yes	No	Yes	Use
M.D. Wilkinson, et al.	2016	The FAIR Guiding Principles for scientific data management and stewardship	Yes	No	Yes	Use
A. Jacobsen, et al.	2020	FAIR Principles: Interpretations and implementation considerations	Yes	No	Yes	Use
M. Hahnel and D. Valen	2019	How to (easily) extend the FAIRness of existing repositories	Yes	No	Yes	Use
M. Van Reisen, M. Stokmans, M. Basajja, A.O. Ong'ayo, C. Kirkpatrick, and B. Mons	2020	Towards the tipping point for FAIR implementation	Yes	No	Yes	Use
L. Sales, P. Henning, V. Veiga, M.M. Costa, L.F. Sayão, L.O.B. da Silva Santos, and L.F. Pires	2020	GO FAIR Brazil: A challenge for Brazilian Data science	Yes	No	No	Not relevant
I. Margoni and T. Labastida	2020	Licensing FAIR data for reuse	Yes	No	No	Not relevant