

# The composition of data economy: a bibliometric approach and TCCM framework of conceptual, intellectual and social structure

*Sunday Adewale Olaleye*

School of Business, JAMK University of Applied Sciences, Jyväskylä, Finland

*Emmanuel Mogaji*

Department of Marketing, Events and Tourism, University of Greenwich, Greenwich, UK

*Friday Joseph Agbo*

School of Computing, University of Eastern Finland, Joensuu, Finland and School of Computing and Data Science, Willamette University, Salem, Oregon, USA

*Dandison Ukpabi*

Jyväskylä School of Business and Economics, University of Jyväskylä, Jyväskylä, Finland, and

*Akwasi Gyamerah Adusei*

Department of Industrial Engineering and Management, Oulun Yliopisto, Oulu, Finland

## Abstract

**Purpose** – The data economy mainly relies on the surveillance capitalism business model, enabling companies to monetize their data. The surveillance allows for transforming private human experiences into behavioral data that can be harnessed in the marketing sphere. This study aims to focus on investigating the domain of data economy with the methodological lens of quantitative bibliometric analysis of published literature.

**Design/methodology/approach** – The bibliometric analysis seeks to unravel trends and timelines for the emergence of the data economy, its conceptualization, scientific progression and thematic synergy that could predict the future of the field. A total of 591 data between 2008 and June 2021 were used in the analysis with the Biblioshiny app on the web interfaced and VOSviewer version 1.6.16 to analyze data from Web of Science and Scopus.

**Findings** – This study combined findable, accessible, interoperable and reusable (FAIR) data and data economy and contributed to the literature on big data, information discovery and delivery by shedding light on the conceptual, intellectual and social structure of data economy and demonstrating data relevance as a key strategic asset for companies and academia now and in the future.

**Research limitations/implications** – Findings from this study provide a steppingstone for researchers who may engage in further empirical and longitudinal studies by employing, for example, a quantitative and systematic review approach. In addition, future research could expand the scope of this study beyond FAIR data and data economy to examine aspects such as theories and show a plausible explanation of several phenomena in the emerging field.

**Practical implications** – The researchers can use the results of this study as a steppingstone for further empirical and longitudinal studies.

**Originality/value** – This study confirmed the relevance of data to society and revealed some gaps to be undertaken for the future.

**Keywords** Big data, Open data, Data privacy, Datafication, Data economy, FAIR data

**Paper type** Literature review

## 1. Introduction

Data refers to either textual or numeric units of information presented using specific machine language systems that enable interpretation by suitable technologies (Monino, 2016). The volume of data is continuously increasing following the proliferation of digital technologies, including smartphones, web services and social networks. The advancement of diver's technologies and diffusion of innovation has increased data

---

© Sunday Adewale Olaleye, Emmanuel Mogaji, Friday Joseph Agbo, Dandison Ukpabi and Akwasi Gyamerah. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

*Erratum:* It has come to the attention of the publisher that the article, "The composition of data economy: a bibliometric approach and TCCM framework of conceptual, intellectual, and social structure" by Sunday Adewale Olaleye, Emmanuel Mogaji, Friday Joseph Agbo, (continued on next page)

---

The current issue and full text archive of this journal is available on Emerald Insight at: <https://www.emerald.com/insight/2398-6247.htm>



Information Discovery and Delivery  
51/2 (2023) 223–240  
Emerald Publishing Limited [ISSN 2398-6247]  
[DOI 10.1108/IDD-02-2022-0014]

Received 14 February 2022  
Revised 24 May 2022  
10 July 2022  
19 August 2022  
16 September 2022  
Accepted 1 October 2022

generation of big data in academia, industry and society. Data economy as an offshoot of big data has created opportunities for data partnership and better environment, cost and user-friendly services. This development is clear visibility of hidden innovation, as suggested by [Edwards-Schachter and Wallace \(2017\)](#). The data economy is at the epicenter of society, from data generation, data cleaning and data engineering to innovative products and services. Through the data economy, ecosystems of the data-driven small, medium and large companies stand to allay consumers' fear of legal, privacy and security issues. Small, medium and large companies can also create a sustainable data moat for competitive advantage and to centralize their data assets with the intervention of artificial intelligence and other emerging technology. One of the earlier studies suggested prioritizing data product or service; it is essential to identify the available opportunity, build the product or design the service, evaluate the first two stages and iterate based on data and the user feedback ([Glassberg, 2018](#)). To maximize the data economy, the timely intervention of governments and societies on the political, economic and social impacts of data-driven artificial intelligence is crucial. This bibliometric study probed into data economy, value and gaps for academia and the practicing managers in a changing landscape of big data opportunity. Therefore, the following research questions guide this study:

- RQ1.* How has data economy been conceptualized and presented by scholars?
- RQ2.* What are the intellectual outputs and contributions of scholars in data economy?
- RQ3.* What social synergy and collaborations exist in the domain of the data economy?

This study contributes to the literature on data economy in multiple ways and presents implications for educators, academic researchers, managers and policymakers. For educators, our analysis provides current teaching materials on essential areas of data economy, providing pedagogical insight relevant to enhancing students' teaching and learning experiences. Our study provides a comprehensive overview of the current state of research on the data economy for academic researchers. We reviewed 591 articles exploring how data economy has been conceptualized and presented by scholars. Our analysis suggested that scientific output

---

Dandison Ukpabi and Akwasi Gyamerah Adusei published in *Information Discovery and Delivery* omitted the first affiliation of Friday Joseph Agbo, and the last name of the author Akwasi Gyamerah Adusei. Friday Joseph Agbo's affiliations are: School of Computing, University of Eastern Finland, Joensuu, Finland and School of Computing and Data Science, Willamette University, Salem, Oregon, USA. The author previously listed as Akwasi Gyamerah should be referred to as Akwasi Gyamerah Adusei. This error was introduced in the editorial process and has now been corrected in the online version. The publisher sincerely apologises for this error and for any inconvenience caused. This article should now be cited as Olaleye, S.A., Mogaji, E., Agbo, F.J., Ukpabi, D. and Adusei, A.G. (2022), "The composition of data economy: a bibliometric approach and TCCM framework of conceptual, intellectual and social structure", *Information Discovery and Delivery*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/IDD-02-2022-0014>.

This work was supported by the Foundation for Economic Education (Liikesivistysrahasto) [grant numbers: 16–9388, 18–10407].

about data economy has remained on the rising curve, suggesting that the emerging field has the potential to grow significantly on an annual basis.

Based on the theories, contexts, characteristics and methodology (TCCM) framework, our review finds that dynamic capability theory has been a dominant theoretical underpinning in the data economy research stream. Data managers gain insights from our comprehensive business model that harmonizes the ethical, legal, technology and societal issues. This proposed business model will proffer solutions to the existing teething problems of the data economy. For policymakers, we posited that the timely intervention of governments and societies on the political, economic and social impacts of data-driven artificial intelligence is crucial; we, therefore, provided insight into developing data-driven strategies that make the stakeholders proactive. Based on the insights gained from our literature review, we develop an agenda for future research, outlining topics and potential research questions based on the TCCM framework. We suggest future research to expand on the intersection of sharing, platform and data economy. In addition, researchers can test existing theories and show a plausible explanation for their investigation of the data economy.

## 2. Review of the literature

There are numerous sources of data, and companies, financial institutions and health service providers generate large amounts of data through their interactions with suppliers, customers and employees. Data is a crucial factor in production that complements physical capital and labor ([Opher et al., 2016](#)). It is nondepletable, and its increased use increases its value. As an asset, its value can deplete over time, as the data becomes less relevant, and its value depends on its unique characteristics. Data is also regarded as a nonrival asset, as multiple users can use it simultaneously ([Agata, 2020](#)). However, it is not automatically labeled as a public good because the data owners reserve the right to exclude individuals from using it, further increasing its value. According to [Nobre and Tavares \(2017\)](#), data can be produced and stored at low costs and households, businesses and individuals constitute the major producers and consumers of data.

### 2.1 FAIR data

FAIR data refers to findable, accessible, interoperable and reusable data ([Dunning et al., 2017](#)). The characteristics of findable, accessible, interoperable and reusable data (FAIR) data must adhere to the FAIR principles, which are used to determine the levels of compliance. This data is assigned a globally unique and persistent identifier and is simple to execute. Suitable examples of the persistent identifier include the digital object identifier, HANDLE (a unique and persistent identifier for Internet resources) and uniform resource name systems ([Dunning et al., 2017](#)). FAIR data is also characterized by several other facets, including being indexed or registered in a searchable resource, and it must be accompanied by a description comprising different attributes.

According to [Tanhua et al. \(2019\)](#), FAIR data enables effective data management through the collaboration of various activities, including quality assurance and control, observations, metadata and data assembly and data publication. Effective data management aims to enhance local and interoperable data discovery access and secures archiving, resulting in long-term preservation. FAIR data is becoming a crucial tool for enabling

digital transformation by supporting research and development (Wise *et al.*, 2019). It capitalizes on analytics tools such as machine learning and artificial intelligence to enable automatic and scalable access to data and support continuous learning.

Wise *et al.* (2019) established that the successful implementation of FAIR data principles would amplify the value of data sets within the companies and external public data by enhancing its discovery and accessibility for humans and machines. This advantage capitalizes on the ranking and rating capabilities of machine learning through algorithmic decision-making. According to Lahoti *et al.* (2019), algorithmic decision-making is continuously becoming pervasive in all aspects of life. Implementing FAIR data principles will help address its societal and ethical concerns.

## 2.2 Open data

Open Data (OD) is scientific data that can be published and reused without any permission or price barriers (Murray-Rust, 2008). It involves publishing data in reusable formats and enhances engagement and innovativeness. Advocacy for OD mainly focuses on the need to increase cyber scholarship. Molloy (2011) supports this assertion by observing that OD enhances science, increasing transparency and societal benefits. OD can be processed and analyzed using data mining tools and automated text analysis to derive valuable findings on business innovation drivers (Molloy, 2011).

According to Huijboom and Van den Broek (2011), OD strategies increase transparency and efficiency in data management. It also fosters services and products innovation. Companies can use available public data to create new businesses, especially digital services, by converting their creativity and ideas into practical solutions to daily challenges (Huijboom and Van den Broek, 2011). Reichman *et al.* (2011) indicate that OD is advantageous because its deployment is bound to enhance and accelerate scientific advancements. Linked data methods avail suitable ways to connect data from distributed sites via standard Web technologies (Reichman *et al.*, 2011). It scales above human limitations by enabling new and improved types of synthetic data studies conducted on larger scales.

Implementing OD policies is meant to stimulate and control data publication to enhance advantageous use. They are mainly implemented within the government systems to increase participation, self-empowerment, social inclusion and interaction (Zuiderwijk and Janssen, 2014). These positive attributes will stimulate economic growth in the countries by supporting business innovation. The availability of OD has been continuously increasing because of the increased pressure on public organizations to publish their data (Janssen *et al.*, 2012). The major motivation for this increase is that increasing access to publicly funded data will increase returns on public investments and enhance wealth generation by using this data to address complex problems. Kassen (2013) also indicates that OD provides a helpful platform for promoting civic engagement and enhancing research and hypothesis testing.

## 2.3 Data ecosystems

Data ecosystems entail the sociotechnical complex networks that allow actors to interact and collaborate to discover, publish, archive, consume or reuse data (Oliveira *et al.*, 2019). These networks also enable them to create value, foster innovation and support new businesses. Oliveira *et al.* (2019) further established

that the emergence of data ecosystems had been influenced by digital technologies that enhance OD production and consumption. The digital technologies supporting data systems include the Internet of Things (IoT), Web technologies and data analytics technologies. Data ecosystems also address the need for a feedback loop between the data providers and data users.

According to Rantanen *et al.* (2019), the significance of data ecosystems has been increasing based on the capability to enrich, use and reuse big datasets by third parties. Various data ecosystems have been formed by groups such as governments, industries and public-private partnerships. Rantanen *et al.* (2019) further established that the data ecosystem has immense potential to provide sustainability in business and enhance competitive advantages. Data ecosystems are formed in different ways and contribute to creating value that individual participants could not realize (Ding *et al.*, 2011). Its key benefits revolve around the sharing of vital resources. These attributes created new business opportunities and increased access to knowledge and data.

Various data ecosystems exist, including directed data ecosystems, collaborative data ecosystems, acknowledged data ecosystems and virtual data ecosystems. Directed data ecosystems are characterized by centralized control structures and are expected within organizational settings (Curry and Sheth, 2018). Acknowledged data ecosystems comprise distributed participants, while virtual data ecosystems focus on pooling decentralized resources to meet specific goals. According to Zuiderwijk and Janssen (2014), OD ecosystems contribute to realizing the benefits of OD and value creation. They capitalize on the original basis of an ecosystem that enhances interdependencies among partners in exchange networks.

## 2.4 Datafication

Datafication entails quantifying human life through digital data for economic value (Mejias and Couldry, 2019). It is applicable within the business and social sciences domains, whereby the data is put in a quantified form for tabulation and analysis. Datafication extends beyond data digitization to make digital data indexable and easily searchable. It enables large-scale processing of various aspects of human life through specific forms of automatic analysis (Ruckenstein and Schüll, 2017). Datafication concept was initially applied in business, and up to date, the amount of commercial data generated exceeds that obtained through the datafication of social life.

Crucial areas in the commercial sector, such as logistics, have continuously advanced to become complex business practices because of datafication. According to Mai (2016), datafication is advantageous because it allows for sophisticated data analysis across large data sets. Its use is predicted to escalate because many digital devices are increasingly becoming connected to the internet (Mai, 2016). This transformation will allow for the digitization of all activities and further extend the scope of datafication. It will result in numerous advantages to the business world because the increased possibilities for analysis enable businesses to create new forms of value. Mai (2016) further established that datafication and predictive analysis would also escalate because more organizations will appreciate the potential to collect and compute user-generated information.

Datafication is widely deployed in social media marketing, whereby metrics from the individuals' use of social networks are quantified to determine market trends (Dourish and Gómez Cruz,

2018). This aspect relates to transforming social actions into online quantified data, which allows for predictive analysis and real-time tracking (Mayer-Schoenberger and Cukier, 2013). Datafication has gradually transformed into a new paradigm for comprehending social behavior (Van Dijck, 2014). This aspect has quantified various social data such as interests, friendships, information searches, emotional searches and casual conversations.

### 2.5 Data economy

The data economy comprises an ecosystem of organizations that use data as their business's main object or source (Opher et al., 2016). It is centered on the production, consumption and distribution of digital data. The data economy thrives on the rapid advancements in digital technologies, especially machine learning, automation and artificial intelligence (Opher et al., 2016). There are no clear distinctions between the producers and consumers in the data economy because supply and demand do not automatically determine the price (Zech, 2016). Data is regarded as a valuable economic resource and players within the data economy can share nonpersonal data to boost economic growth and enhance innovation and interoperability (Bonti et al., 2021). The data economy enables players to derive value from data by transforming it into applications, insights and services. Bonti et al. (2021) further indicate that participation in the data economy also enables organizations to attain the full potential of their data. This potential mainly arises from sharing data within the intercompany systems, developing new capabilities and deploying emerging technologies. Consumers and companies constitute the leading players in the data economy, as they contribute to the value chain associated with data production, collection and analysis (Allen, 2016).

The data economy mainly relies on the surveillance capitalism business model, enabling companies to monetize their data. In this business model, commodities being sold constitute personal data, and this data is collected and produced through massive internet surveillance (Zuboff, 2015). The surveillance allows for transforming private human experiences into behavioral data that can be harnessed in the marketing sphere. Monetization of personal data has led to the emergence of the personal data economy, allowing individuals to share their data with businesses (Elvy, 2017). According to Lammi Pantzar (2019), the personal data economy is supported by the rapid increase in mobile and handheld devices. The devices collect personal data such as geographical information, consumer purchase behaviors and other online metrics (Lammi and Pantzar, 2019). This personal data comprising the consumer's digital tracks and actions is a viable source of economic value creation through the digital economy.

## 3. Methodology

This study focuses on investigating the domain of data economy with the methodological lens of quantitative bibliometric analysis of published literature. The bibliometric analysis seeks to unravel trends and timelines for the emergence of the data economy. Its conceptualization, scientific progression and thematic synergy could predict the field's future. A three-step approach demonstrated in recent studies to conduct a bibliometric analysis was followed by Aria and Cuccurullo (2017) and Agbo et al. (2021a), which consists of:

- 1 article selection and data gathering process;

- 2 data extraction, loading and conversion process; and
- 3 data synthesis process.

In this study, the three main software used were RStudio, Biblioshiny developed by Aria & Cuccurullo and VOSviewer version 1.6.16 by Van Eck & Waltman. The RStudio software is an open-source solution for data science analysis and can be downloaded free from their official website. Biblioshiny is a Web tool that can be launched from the RStudio to provide a Web interface for data visualization and VOSviewer is an open-source software downloadable and used on desktop computers.

The impact of TCCM has been emphasize in review literature (Paul et al., 2021). This study integrates TCCM into a bibliometric study to deepen the understanding of dominant theories, contexts, characteristics and methods employed in data economy research over the past decade (2008 – 2021). The study draws on relevance of theory in domain research as demonstrated by Paul and Feliciano-Cestero (2021) and Paul et al. (2021) that posits that TCCM is efficient for theme-based reviews and emphasizes the impact of TCCM and reviews that develop theories. Based on the earlier proposition on TCCM, this study used TCCM to identify the global used theories, contexts, insightful variables and methods to strike a balance in data economy research and to propose a new direction for future research through gaps identified through TCCM.

### 3.1 Research design

Bibliometric research designs help ascertain the alignment of the data collected and the choice of the data analysis technique (Agbo et al., 2021c). The study starts with a clear idea of research questions that require further investigation. The study chose descriptive and correlation as a subset of quantitative research design. This design intends to give a clearer picture of trends, characteristics and the relationship between authors, coauthors, institutions and stakeholders of FAIR data and data economy through existing literature. Further, the study defined the focus of the study (FAIR data and data economy) and the literature inclusion and exclusion. The study adopts secondary data to expand the scope of the existing literature. In addition, the study used search engines of two extensive databases to collect consistent, accurate and unbiased data and ensure that the results of this study can be easily reproduced (Lai et al., 2020), measured all the necessary concepts and correlate with different measures of the same concept. The data used is well organized and backed up in the cloud for easy data analysis and other researchers' validation and inputs. Olaleye (2020) summarized this process as bibliometric systematic workflow and divided the workflow into six distinct parts:

- 1 research design;
- 2 bibliometric data source;
- 3 bibliometric data analysis;
- 4 data visualisation;
- 5 result and interpretation; and
- 6 conclusion.

Biblioshiny and VOSviewer were used for the data analysis (see details below).

#### 3.1.1 Article selection and data gathering process

The data used in this study were retrieved on June 12, 2021, from Elsevier's Scopus and Clarivate Analytics Web of Science

(WoS) databases, respectively. According to [Saqr et al. \(2021\)](#), the Scopus database warehouses contains over 70 million peer-reviewed articles, whereas, according to the Clarivate website, as of June 2021, WoS contains over 81 million records consisting of published materials from life sciences, biomedical sciences, engineering, social sciences, arts and humanities. Rather than collecting data from a single database to conduct a quantitative bibliometric analysis as showcased in [Agbo et al., 2021b](#); [Aria and Cuccurullo, 2017](#)). This current study undertook a different approach by collecting data from the two earlier mentioned databases to contain the most relevant articles ([Agbo et al., 2021a](#)), which can sufficiently represent the scope of the field under investigation. The main rationale behind using two databases for the data sources was to minimize the tendency to leave out relevant data and conduct an in-depth analysis. On the other hand, the risk of collecting duplicate data in this kind of approach is mitigated by the data conversion processes, as shown in a subsequent section.

The search terms “fair data” OR “data economy” were used to search the two databases. Mainly, the search strings were applied to the title, keywords and abstract metadata of the documents. The structure of the query as used in the respective database search engine is shown in [Table 1](#).

Furthermore, the search was limited to only documents classified as articles and conference proceedings in both databases. The authors decided to limit the data to articles and conference proceedings to allow for analysis that could provide deeper scientific insight because documents from these data points are peer reviewed. [Figure 1](#) presents the preferred

**Table 1** Data gathering procedure demonstrating the search strings and output from databases

Database	Search strings	Output
WoS	("fair data" OR "data economy" OR "fair data" OR "data economy") Refined by: DOCUMENT TYPES: (ARTICLE OR PROCEEDINGS PAPER) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	393
Scopus	(TITLE-ABS-KEY ("fair data" OR "data economy") OR TITLE-ABS-KEY ("fair data" OR "data economy")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp"))	558

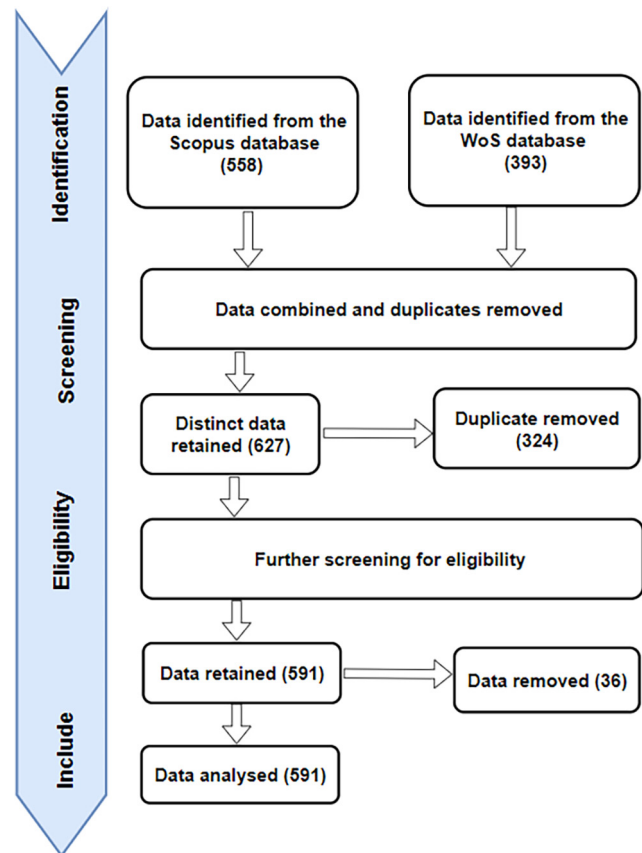
reporting items for systematic reviews and meta-analyses workflow of the data collection and screening process.

3.1.2 Data extraction, loading and conversion process

The data extraction, loading and conversion process are explained following the steps provided by [Agbo et al. \(2021a\)](#). The Bibliometrix R Library ([Aria and Cuccurullo, 2017](#)) was used to combine the two data points. The script for the conversion and combination of the data is shown here in lines 1–10 in [Table 2](#).

Line 1 in [Table 2](#) creates an instance of the directory where the downloaded data from the Scopus and WoS are stored and the combined result. For a detailed explanation of the scripts in each

**Figure 1** PRISMA workflow showing the data collection and screening for this study



**Table 2** Lines of instructions for converting and combining two data sources using RStudio software

Command line	Command
Line 1:	setwd("C:/Users/Intel/Desktop/.. /de2")
Line 2:	getwd()
Line 3:	DigitScopus2 = convert2df("scopus.bib", dbsource="scopus", format="bibtex")
Line 4:	View(DigitScopus2)
Line 5:	digitwos2 = convert2df("wos.bib", dbsource="isi", format="bibtex")
Line 6:	View(digitwos2)
Line 7:	CombinedData = mergeDbSources (DigitScopus2, digitwos2, remove.duplicated = TRUE)
Line 8:	View (CombinedData)
Line 9:	dim(CombinedDatabase)
Line 10:	library (openxlsx)
Line 11:	write.xlsx (CombinedData, file = "XlsCombinedData.xlsx")

line, we refer readers to this previous study by Agbo *et al.* (2021a). Line 7 is the command to combine these data and remove any duplicates. After executing this command in the case of this study, 324 duplicated documents were removed, leaving 627 data used for the data analysis. In order words, only 69 documents were distinct that are indexed in Scopus but not in WoS. This difference implies that when a bibliometric analysis is conducted with a single database such as Scopus or WoS alone, relevant data are left out, which may significantly impact the result of the study. Therefore, our choice of using two databases accords with the previous study by Agbo *et al.* (2021a) justifies this finding.

Moreover, a closer review of the resulting data from our conversion and combination shows some irrelevant documents that can be dropped. As shown in Table 3, the data timespan is from 1970 to 2021. Because the data economy, in our opinion, is an emerging domain in the 21st century, it will make more sense to analyze somewhat recent data. Therefore, we delimited the data between 2008 and 2021 using the Biblioshiny filter function, and the results are presented in Table 4.

### 3.1.3 Data synthesis process

A total of 591 data between 2008 and 2021 were used in the analysis, as shown in Table 4. These documents emerged from

Table 3 Main information of the merged data

Data description	Results
Period	1970–2021
Documents	627
Article	376
Conference paper/proceedings	251

Table 4 Main information about the data set used in the bibliometric analysis

Data statistics	
Documents	591
Period	2008–2021
Data sources (journals, books, etc.)	402
Keywords plus (ID)	3,160
Author's keywords (DE)	1,656
Average years from publication	2.90
Average citations per documents	9.40
Average citations per year per doc	1.90
Document types	
Journal article	355
article; proceedings paper	1
conference paper	201
proceedings paper	34
Authors	
Authors	2,353
Author appearances	2,842
Authors of single-authored documents	109
Authors of multiauthored documents	2,244
Single-authored documents	112
Main documents per author	0.25
Main authors per document	3.98
Main coauthors per documents	4.81
Authors collaboration index	4.70

journals, books, conference proceedings and book chapters. In addition, the data set consists of 2,353 authors, among which 112 are single authors. On the other hand, 1,656 authors have distinct keywords.

### 3.1.4 Data analysis

The study used the Biblioshiny app on the Web interfaced and VOSviewer version 1.6.16 to analyze Scopus and WoS data. First, the study checks for the descriptive values of the literature (details in Table 4). Second, the study carried out analytics and plots based on three different metrics: sources, authors and documents. Third, the study analyzed three knowledge structures of concepts and intellectual and social relationships. Fourth, the study used VOSviewer to filter the theories and methodologies used in FAIR data and data economy literature and later used the values to plot charts in Microsoft Excel for clarity.

Further, the study used VOSviewer for country mapping. We noticed that mapping countries using Biblioshiny could be problematic where it is not easy to differentiate countries, for example, China from Taiwan, whereas, in VOSviewer, they are treated separately. The data analysis generates literature mapping insights discussed later in the study in section four.

## 4. Results and discussion

In this section, we present the results of this study and discuss them based on the research questions to aid the flow of information and understanding.

### 4.1 Conceptual structure of data economy

The conceptual structure of a knowledge domain deals with the representation of concepts to describe specific classification, interrelationships and even taxonomy that can enhance interpretation and understanding of the domain. Because data economy is a vast but emerging area of diverse interest, as highlighted in the background section, this section tries to present its emergence from the conceptual point of view where the underpinning theories and scientific production of articles in the field are analyzed. In addition, the thematic clusters and field evolution based on authors' keywords are examined.

*RQ1.* How has data economy been conceptualized and presented by scholars?

This section begins by examining how data economy has been conceptualized from the perspective of theories through the lens of TCCM framework. The TCCM has recently gained momentum in systematic reviews and bibliometric analysis studies (Sharma *et al.*, 2020; Olaleye *et al.*, 2021). In this framework, scholars posit that the popularity of the TCCM is hinged on the ease with which knowledge gaps can be spotted and a logical process of recommending future research direction (Singh *et al.*, 2020). Thus, we followed the TCCM framework in this study, as detailed in the subsequent sections.

#### 4.1.1 Theory development

The stream of theoretical developments in data economy research embraces theories from different disciplines (Table 5). The integrated theories are trust, security, privacy (Meijer *et al.*, 2014; Kobayashi *et al.*, 2018) and dynamic capability theory (Lee and Yoo, 2019; Shan *et al.*, 2019), knowledge management (Sumbal *et al.*, 2017) and decision theory (Elgendy *et al.*, 2021), graph

Table 5 Theory keywords, frequency and total link strength

No.	Keyword	Occurrences	Total link strength
1	attitudes	1	0
2	awareness and perception	1	0
3	data agency	1	0
4	data trusts	1	0
5	data use	2	0
6	decision-making	2	0
7	dynamic capability theory	1	0
8	firm performance	2	0
9	knowledge management	2	0
10	learning	1	0
11	motivation	2	0
12	network graph	1	1
13	personality type	1	0
14	security and privacy	1	0
15	semantics	1	0
16	social and cultural anthropology	1	0
17	text mining	2	1
18	transactions controlled semantic model	1	0
19	trust	1	0

Note: Based on the occurrences and total link strength in Table 5, there is dearth of theories in data economy research stream

theory (Yildirim et al., 2021) and attitudes (Tenopir et al., 2020; Baždarić et al., 2021). Figure 2 displays the prominent theories used in data economy research.

Because of data dimensions, Zhao and Fan (2018) differentiated open government data into tangible, human and intangible. The study found that culture plays a crucial role in its open government data capacity. Trust has been a dominant theory in the information systems and data economy research stream (Meijer et al., 2014). Trust implies the willingness of parties (OD generators and users) to rely on each party’s ability to generate and use OD transparently. This transparency, therefore, implies that the security and privacy of relevant stakeholders must be protected.

Our literature search also reveals that the dynamic capability theory has been a dominant theoretical underpinning in the

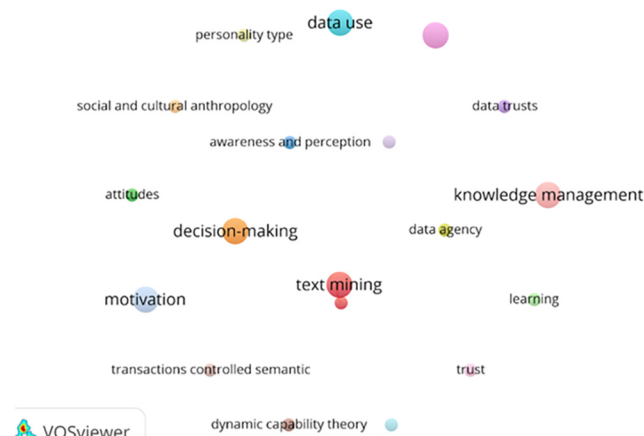
data economy research stream. The core proposition of the dynamic capabilities’ theory holds that firms must be able to develop their short-term competitive positions into long-term competitive advantages to survive in the face of rapidly changing business climates (O’Connor, 2008). The use and application of OD to solve different challenges also trigger uncertainties to traditionally held norms. Lee and Yoo (2019) align the dynamic capabilities theory with opening innovation and argue that survival means that firms must develop the ability to identify opportunities and threats and explore skills necessary to detect and harness market opportunities.

4.1.2 Context

The application of big data transcends many contexts. As our review shows, scholars have adopted OD in industrial applications (Huang et al., 2021), entrepreneurship (Aridi et al., 2021); blockchain (Hu et al., 2021), vehicle design (Urquhart et al., 2021) and health (Geneviève et al., 2021; Ochs et al., 2021). Our findings make it quite challenging to state which domains enjoy the most research output. This unclarity is because, as argued by (Alencar et al., 2014), the application of open is multisectoral, so there is seemingly a bandwagon attempt by various stakeholders to leverage the opportunities present in its adoption (Janssen et al., 2012; Zuidervijk and Janssen, 2014).

As the contexts of OD applications are diverse, its benefits percolate across these contexts. Janssen et al. (2012) categorized these benefits into three thematic areas: political and social, economic and operational and technical. Per political and social benefits, they argue that OD has, among other things, ushering in more transparency, trust in government and public engagement. It has stimulated economic growth and competitiveness, new product and service design and open innovation for economic benefits. Operational and technical benefits include improvement in public policies, external policy checks by the public and the ability to reuse data.

Figure 2 Theoretical foundations of data economy research



4.1.3 Characteristics (C)

Our literature analysis also reveals the characteristics of the data economy structure and its composition. As shown in Figure 3, four themes generated by the Biblioshiny app emerged and are broadly categorized into vertical – development degree (density) and horizontal – relevance (centrality). The four thematic areas identified by our literature are motor themes (upper right quadrant), niche themes (upper left quadrant), basic themes (lower right quadrant) and emerging or declining themes (lower left quadrant).

At the upper end of the significant themes are machine learning, artificial intelligence and the IoT, while data science is identified at the lower end. Several studies have identified the ever-evolving field of machine learning as critical to making sense of OD (Celi et al., 2019), just as machine learning improves itself through data sets. The clustering of machine learning and artificial intelligence confirms that artificial intelligence adopts machine learning to solve problems. In line with the basic themes, a plethora of studies have hinged the usability of OD on FAIR data principles and the ability of users to freely access these data sets (Janssen et al., 2012). Further studies would need to pay closer attention to the concepts categorized as niche and emerging or declining themes. For example, further exploration of wireless sensor networks, congestion control and FAIR data collection is necessary. Again, artificial neural networks and indoor positioning require further exploration based on their classification as niche themes.

4.1.4 Methodologies

Machine learning as a subfield of artificial intelligence data analysis techniques dominated the quantitative methodological approaches of extant studies (Figure 4). Machine learning approaches were text mining and natural language processing

Figure 3 Characteristics of data economy research

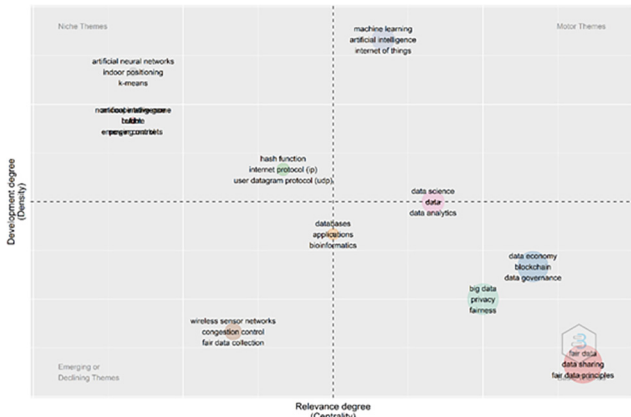
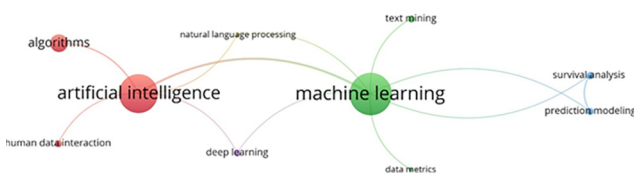


Figure 4 Characteristics of data economy research



(Desai, 2015). The role of deep learning in the data economy research stream has also been underscored because most of the studies adopted it (Kiarashinejad et al., 2020; Tabernik et al., 2020). Phan et al. (2017) applied deep learning in predicting human behavior with social health workers. Adopting the restricted Boltzmann Machine predicted human behavior accurately and generated explanations for these behavioral leanings better than the conventional method.

Again, Musci et al. (2018) employed survival analysis in a longitudinal study to determine marijuana use among elementary school pupils in a US city. They found that genetics play a key role in first marijuana use. That differences in genetics also account for the effectiveness of classroom-based intervention in delaying drug use among pupils. del Pozo Cruz et al. (2020) also applied survival analysis to estimate mortality risks among adults. They found that poor diet quality and activity profile increased the likelihood of mortality rate. Predictive modeling was also used in some studies. For instance, Kaur and Kumari (2020) applied predictive modeling in an Indian study to classify diabetic and nondiabetic patients. Using the Boruta wrapper features selection algorithm performs better than manually selecting the attributes, especially with little medical knowledge.

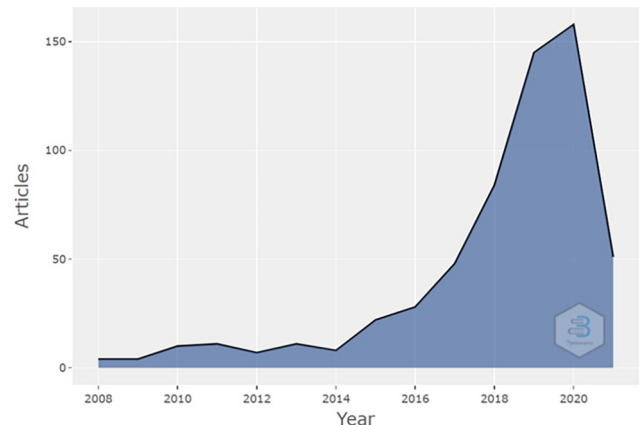
4.1.5 Scientific production of data economy

As presented in Figure 5, our analysis shows that the scientific production of articles in the data economy may have commenced over a decade but received a boost in 2016. The growth of the data economy in terms of scientific output has remained on the rising curve, suggesting that the emerging field has the potential to grow significantly on an annual basis. In 2020, there were slightly over 150 articles published. When the data was collected in the middle of 2021, over 50 articles on data economy were already published, which shows that the total number may surpass previous years.

4.1.6 Thematic evolution of data economy

Furthermore, this study conducted the thematic evolution of data economy and FAIR data. Thematic evolution of a field is the analysis that seeks to unravel a set of themes that have evolved across subunits over a period (Chen et al., 2019). In other words, we consider a theme to have evolved from A to B if there exist common keywords within a thematic network.

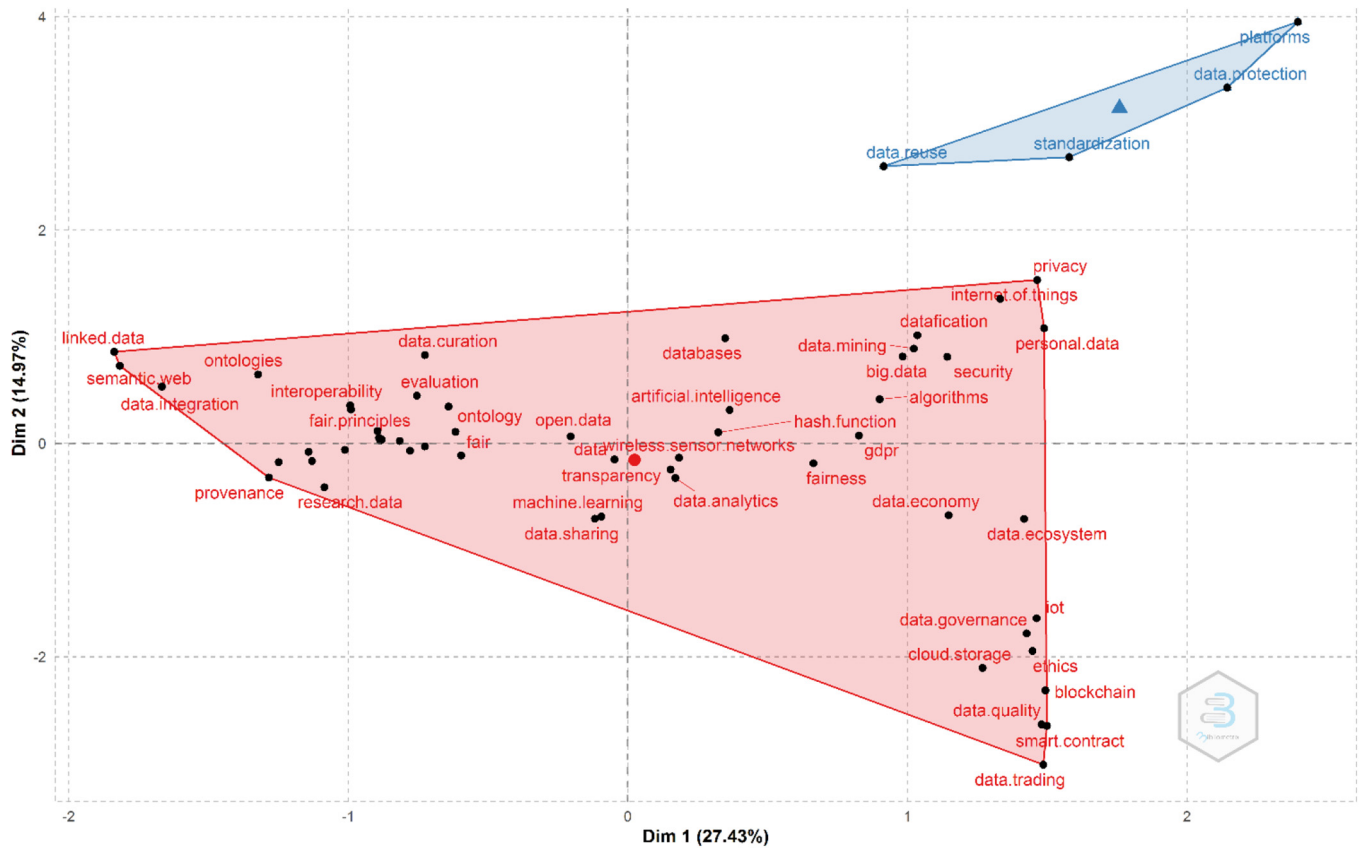
Figure 5 Scientific production in the domain of data economy







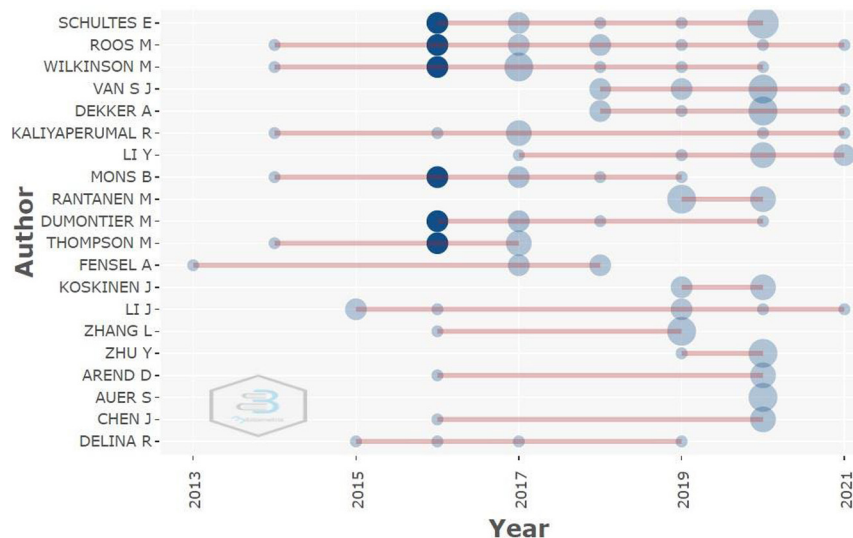
**Figure 8** Conceptual structure map based on correspondence analysis (CA)



on the author’s fractionalized number of articles. According to [Aria and Cuccurullo \(2017\)](#), fractional authorship measures an individual author’s contributions to a published article. Based on this premise, our analysis revealed the top 20 authors’ productivity over the years (2013 to mid-2021). As shown in [Figure 9](#), the line delineates an author’s productivity timeline; the round

bubble shapes represent the author’s production. The shape’s size depicts the number of articles produced by the author per year. The color intensity of the bubble represents the total number of citations the author has received per year. The first bubble on the left-hand side of the line depicts the author’s first production year. Consequently, Roos M. and Kaliyaperumal R. both have the most extended timeline

**Figure 9** To 20 prolific authors publishing in the field of data economy between 2013 and mid-2021



of article production (2014–2021), followed by Wilkinson M. and Li J. having six years of scientific production (from 2014 to 2020 and 2015 to 2021, respectively).

It can be seen from the analysis that the work of Fensel and her colleague was the only article on data economy in 2013, which perhaps was presented at the IEEE international conference on big data (Tomic and Fensel, 2013). Besides the author's productivity timelines, the analysis based on the authors' fractionalized number of articles revealed that Schultes E. came first among the top 20 authors. Notably, Schultes began to publish articles on the data economy in 2016, with 11 articles between 2016 and 2020. However, Schultes's work received a value of 3.11 of the fractionalized frequency with many citations in 2016.

### 4.3 Social structure of data economy

*RQ3.* What social synergy and collaborations exist in the domain of the data economy?

Social synergy is a concept proposed by Barbara Marx Hubbard after the biting global financial crisis of 2008. Social synergy indicates the mindset of a group's creative presence and productivity through knowledge sharing. It means researchers can combine efforts to attempt more considerable academic challenges with the impact of technology and offer solutions. The earlier study by Baraibar-Diez et al. (2020) established an increasing interest in social impact research.

A social structure indicates social actors with common interests linked together by connections. Studying the social structure gives more insights than studying an isolated actor. The data economy social structure graphical representation depicts the author as nodes and their relationships as edges. The author's cocitation network in Figure 10 shows three unique social network analysis centrality measures. The measures are betweenness, closeness and page rank to understand the most critical nodes and edges and how they interact in a data-economy bibliometric network. Data economy authors' cocitation network reveals how pairs of papers are cited together in the source articles. When many authors cocite pairs of papers, then research clusters are

formed. Cocitation of data economy pairs of papers formed two clusters, as shown in Figure 10. One is represented by the blue color, while the red represents the other segment. The red segment is more densely clustered than the blue segment, but the blue segment is more influential than the red cluster. Wilkinson is the most influential author in the network, with the most significant node in the blue segment and the author's role is central in the network. The node size determines the total number of citations, while the thickness of the edges between nodes determines the number of times the sources are cited together.

Regarding betweenness in network analysis, which shows the number of paths between two nodes, Wilkinson topped the list in the network with 219.97. Also, Wilkinson scored 0.014 in the closeness centrality measurement, which indicates the length of the path from one node to the other in a network. Further, Wilkinson scored 0.034 in page rank, which unravels the nodes that extend their influence beyond the direct connections into the more comprehensive network. Wilkinson's influence transcends the blue segment and extends to the red segment. Aside from Wilkinson, Johnson is the second influential author in the blue segment with 40.27 betweenness, 0.013 closeness and 0.020-page rank. Sharma had the highest betweenness with 39.07, 0.011 closeness and a 0.020-page rank in the red segment.

Collaboration in academic paper writing is an act of working with one or more authors to produce an academic paper in the form of articles, conference proceedings or book chapters. Common goals, familiar research areas and authors' high reputations could motivate factors to collaborate. Collaboration could also be a mentor–mentee process whereby senior and experienced authors collaborate with junior authors to bring them up. Collaboration facilitates training by observation and training by doing to leverage an expert. Collaboration is good, but it could translate to a positive or negative experience. It is essential to research collaborators and ensure collaborators are a great fit before they tie into their reputation.

Furthermore, it is vital to establish a reputation model for necessary quality assurance. Figure 11 showcases the data

**Figure 10** Authors cocitation network analysis

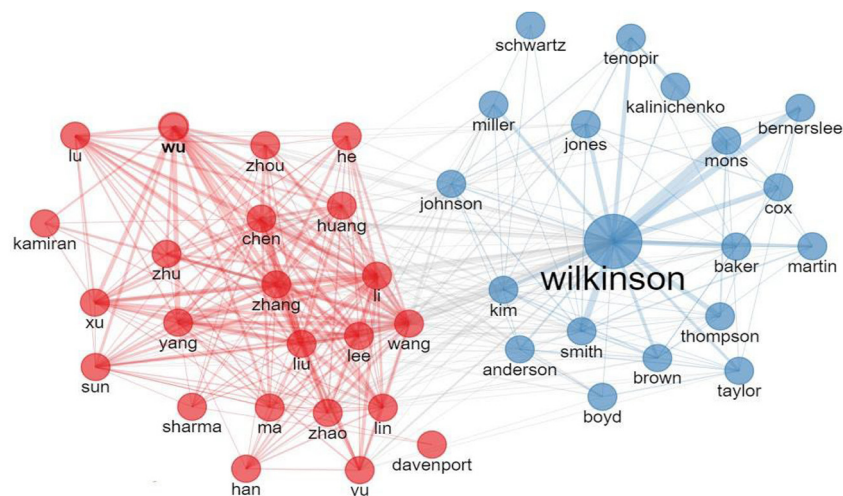
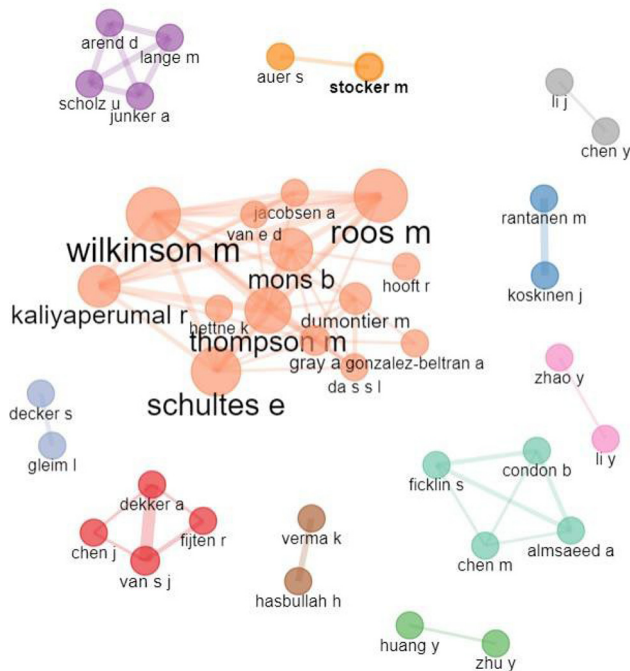


Figure 11 Author's collaboration analysis

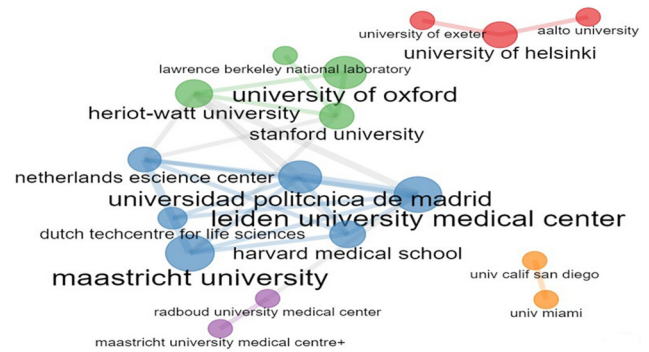


economy authors' collaboration. The figure shows 12 clusters of the authors. The Wilkinson collaboration cluster is the biggest and most densely connected to the other clusters of the 12 clusters. Most of the clusters only have a single collaborator with a single path, and there were no betweenness values recorded for the relationship. However, they have closeness and page rank values. For example, Van has three collaborators, but the author's collaboration is stronger with Dekker than with the two other collaborators. Van has a betweenness of 0.6, a closeness of 0.0007 and a page rank of 0.037.

On the other hand, Dekker has a betweenness of 0.4, a 0.0007 closeness and a page rank of 0.035. On the contrary, Chen and Fijten did not have betweenness values. Similarly, Rantanen and Koskinen have a strong collaboration. The most significant collaboration cluster has a lot of influential authors, and Mons had the highest betweenness of 12.68, 0.0009 closeness and 0.034-page rank. Schultes is close to Mons, with 12.59 betweenness, 0.009 closeness and a 0.034-page rank. Wilkinson led the in-betweenness of the author's co-citation network with 219.97 but had 7.10 betweenness, 0.0009 closeness and 0.04 for page rank. Collaboration is growing among the authors of the data economy, but there is room for improvement.

The author's collaboration is essential, as is the collaboration between institutions. Collaboration between different institutions across borders facilitates knowledge and resource sharing. Figure 12 shows five distinct clusters of institutions' collaborations in the data economy. The orange cluster consists of the University of Miami and the University of California, San Diego (Cluster 5). This cluster has direct collaboration and has no betweenness values, but Miami has 0.004 closeness and a 0.06-page rank, while California has 0.004 closeness and a 0.06-page rank. Like the Miami collaboration, Maastricht

Figure 12 Institutions collaboration analysis



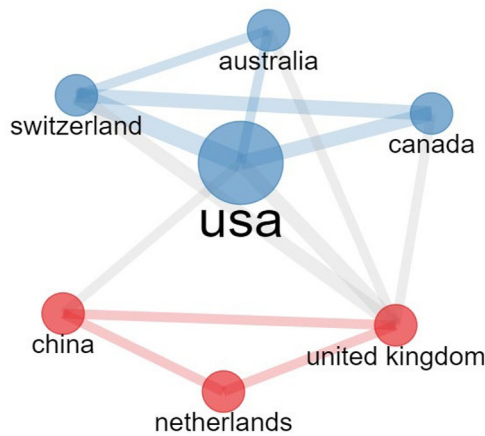
University Medical Center has a single-path collaboration with Radboud University Medical Center in purple (Cluster 4). Maastricht has no betweenness record but has a 0.009- and 0.027-page rank.

In contrast to Maastricht, Radboud has 10 betweenness, 0.009 closeness and 0.05-page rank. Cluster 3 in green consists of Stanford University, Heriot-Watt University, the University of Oxford and Lawrence Berkeley National Laboratory. Oxford and Lawrence did not have betweenness scores, but Oxford has a 0.009 closeness and a 0.04-page rank, while Lawrence has a 0.009 closeness and a 0.03-page rank. Stanford and Heriot-Watt's betweenness are very close (11.66 and 11.46) and their page rank is (0.08 and 0.09). They both have a closeness of 0.009. Stanford and Heriot-Watt are the focal collaborators in Cluster 3. Out of the five clusters, Cluster 2 is the biggest, with six institutions spanning the European and American continents and four institutions from The Netherlands dominating the cluster.

Leiden University Medical Center had the highest betweenness of 21.74, 0.0010 closeness and 0.11-page rank, but Leiden was followed by The Netherlands eScience Center with a 3.76 betweenness, 0.0010 closeness and 0.08-page rank. Universidad Politecnica De Madrid had a similar result to The Netherlands eScience with 3.74 betweenness, 0.0010 closeness and 0.089-page rank. There is network activity outside Clusters 2 and 3; Cluster 2 collaborates with Cluster 3 with light and strong edges, which indicates the level of collaboration. The University of Helsinki is the focal collaborator in Cluster 1, with Aalto University and Exeter. The University of Helsinki had one betweenness, 0.005 closeness and a 0.086-page rank. Aalto University and the University of Exeter did not have betweenness, and they had similar closeness and page ranks of 0.005 and 0.046. The University of Helsinki had both national and international collaborations.

Figure 13 depicts the focal countries in international collaboration, but the VOSviewer Figure 14 shows the comprehensive clusters with focal countries and their networks. The result also reveals the emerging countries in the field of data economy. There are countries like Estonia and Poland. Denmark, Czech Republic, Slovenia, Portugal, Norway, France, Cyprus, Greece, Sweden, Finland, Ireland, Belgium and Bulgaria in Europe. Further, Brazil and Venezuela in Southern America. Canada in Northern America. Also, Singapore, the Philippines, India, Malaysia, Taiwan, Japan, Iran and Saudi Arabia are part of the Asia Continent. More

**Figure 13** International collaboration network



countries like Kenya and South Africa from Africa. Figures 13 and 14 from Biblioshiny and VOSviewer complement themselves and compensate for the missing links in both figures.

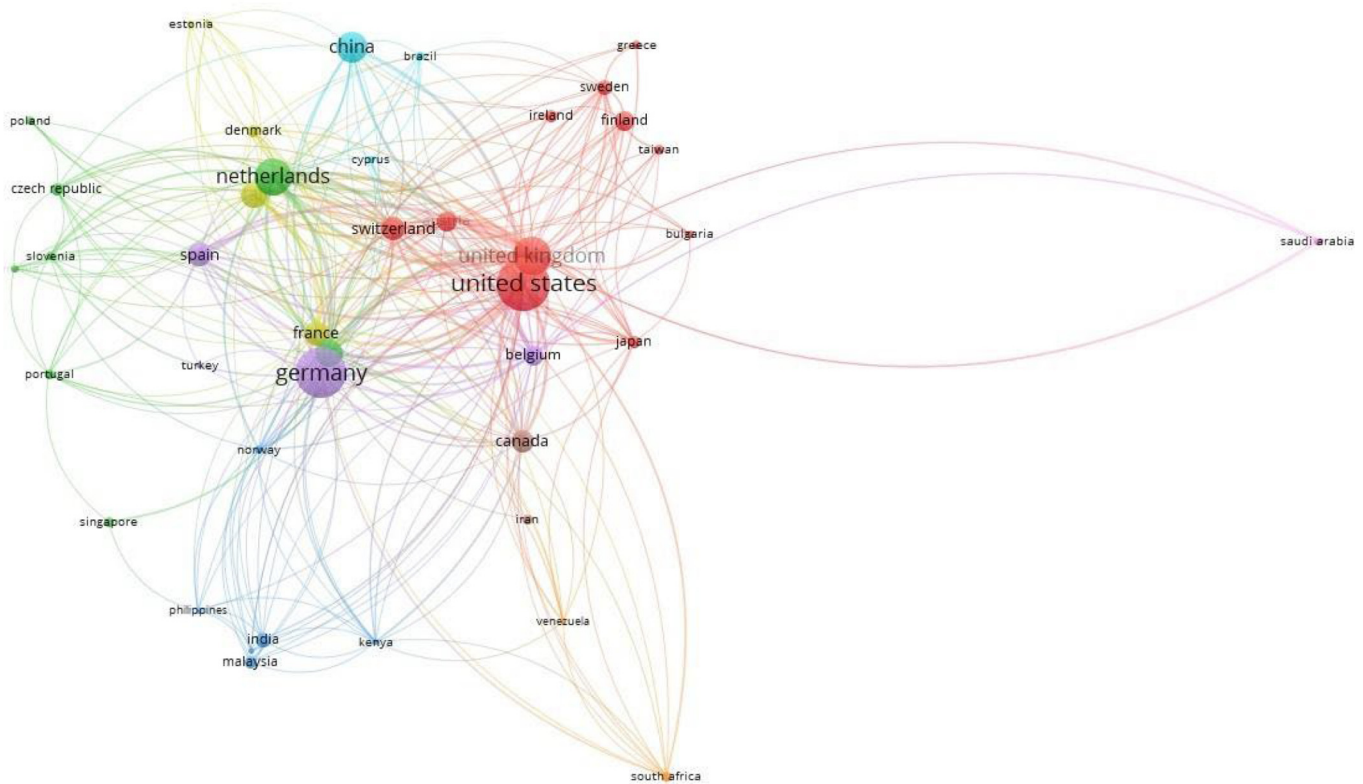
The international collaboration of data economy covers Europe, Asia, North America and Australia. Figure 14 shows two interconnected clusters. Clusters 1 with red color consists of the UK with 5.18 of betweenness, 0.17 closeness and 0.22-page rank, China with 0.4 betweenness, 0.11 closeness and 0.11-page rank, and The Netherlands with no betweenness, 0.1 closeness and 0.08-page rank. The UK had the highest betweenness, closeness and page rank with a strong collaboration path and collaboration across the borders. In the

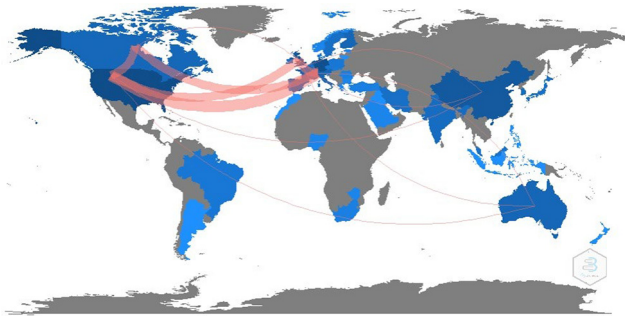
second cluster with blue color, the USA had the highest betweenness with 2.05, closeness with 0.15 and 0.20 for the page rank. Canada and Australia had no betweenness values and the same closeness values of 0.11 but different page ranks. Canada page rank of 0.13 and Australia page rank of 0.010. Switzerland had 0.38 betweenness, 0.13 closeness and 0.18-page rank. The betweenness, closeness and page rank of Cluster 1 are higher than those in Cluster 2. There is more pronounced international collaboration in Cluster 1 than in Cluster 2.

The collaboration world map in Figure 15 reveals six countries collaborating with single and multiple countries. Argentina had a single collaboration with New Zealand. Moreover, Australia collaborates with Argentina, Brazil, France, Greece, Ireland, Japan, New Zealand, Norway, Poland, Singapore, South Africa, Sweden and Switzerland. Out of the 13 countries, Australia collaborated with Switzerland remarkably (2).

Brazil collaborates with six countries, Argentina, Greece, New Zealand, Norway, Poland and South Africa. Also, Canada collaborates with Argentina, Australia, Brazil, France, Greece, Ireland, Japan, New Zealand, Norway, Poland, South Africa, Sweden and Switzerland. Canada had the highest collaboration with Switzerland in 13 countries. Canada's collaboration with Switzerland is similar to Australia's collaboration with Switzerland, only that Canada is higher by 1. China also collaborated with Argentina, Australia, Brazil, Canada, France, Greece, Ireland, Italy, Japan, The Netherlands, New Zealand, Norway, Poland, South Africa, Sweden and Switzerland. China had the highest collaboration with 16 countries. China had the highest collaboration with The Netherlands (2). Lastly,

**Figure 14** Emerging countries on data economy (VOSviewer visualization)



**Figure 15** Global authors collaboration map

France had a collaboration with Argentina. The bibliometric result of data economy shows that the collaborations between some countries are very low, while collaboration does not exist in some countries.

#### 4.4 Study implications

The review contributes to education, research, managerial practice and policymaking on the data economy.

For educators, we established that research and development, human capital creation and creativity would form the new drivers of economic growth in the data-driven economy. It will also result in rapid institutional responses to economic changes through labor market adjustments, depreciation of capital investments, shortening product life cycles and extensive investment planning. The study provides educators with an example of relevant topics for inclusion in teaching and learning resources. The data economy has come to stay as a research domain. Also, this study suggests that a full-fledged program should be developed at certificate, diploma, degree and master's levels. Presently, Laurea University of Applied Sciences (UAS), Finland, has started "Introduction to Data Economy" as one online credit course as part of their UAS Master's program. These programs will create more awareness and enlightenment for the data economy. Further, it will serve as an opportunity to produce seasoned graduates of the data economy that can impact academia and industry. It will also be a means of eradicating job shortages of data economists.

For academic researchers, this study may help researchers globally to discover a suitable niche of data economy and develop it for enlightenment, economic growth, individual empowerment and broad societal benefits. Further, it may help the team leaders and research institutes to form clusters of research areas as dimensions of data economy such as business renewal and new business models for data economy, data economy legal issues, data economy ethical issues, technology integrated with data economy, innovative data economy, health data economy and many more.

This study established how FAIR data, OD, data management, data sharing, data economy, machine learning and big data are interlaced with different dimensions. The complex thematic evolution of the data economy portrays a multidisciplinary data economy. This result is consistent with the studies of Choi and Pak (2007) and Choi and Anita (2008). The authors argued that multidisciplinary, interdisciplinarity and transdisciplinarity are different concepts. They further

posit that it is not necessary to interchange them in research. This study shows that data can resolve complex world problems as a multidisciplinary field of research, and this assertion is consistent with the study of Choi and Pak (2007). Also, this study shows the impact of authors, institutions and countries' collaboration to advance the data economy research domain.

This result is consistent with the study of Choi and Pak (2007). The study proposed eight motivating strategies that can influence multidisciplinary teamwork: team, enthusiasm, accessibility, motivation, workplace, objectives, role and kingship (teamwork). These strategies are proposed to promote the multidisciplinary collaboration promoters and bar the barriers to research collaboration. Further, this study integrates data economy conceptual structure, intellectual structure and social structure to advance the research domain of data economy and show their integration as interactive and holistic.

Publication of data economy is not evenly distributed (Milletler, 2019). This global research gap will motivate researchers in countries with scanty literature on data economy to focus on expanding the existing contribution in the research domain of data economy. In addition, this study gives insights into metrics for making contributions to the data economy. It gives direction to how researchers can make a theoretical, methodological, contextual and conceptual impact on the data economy (Akhavan et al., 2016).

Recent studies elucidate the essential theory in evaluating the quality and contribution of research. This study established the theoretical integration of data economy. Apart from theoretical integration, there is a need for data economy theory, as Milletler (2019) mentioned that the standard economic theories are deficient in explaining the phenomenon of data economy. There is a need for advanced theoretical and conceptual frameworks for the data economy. In addition, this shows a need for data and methodological triangulation for the research process and the epistemological development of a research question.

For data managers, the increasing use of data across the main segments of the economy implies that data has become a new form of capital for the current knowledge economies. Therefore, the vast amounts of available data can be used and reused to benefit society and increase growth opportunities. The industrialization of learning in the data economy will also accelerate innovation because of the availability of data and analysis tools. The continuous growth of the data-driven economy further implies that human capital will be continuously discounted in the subsequent years. This impact follows the continuous deployment of artificial intelligence and machine learning in data collection and analysis. Its increasing use will increase the capabilities of machine intelligence to the extent that it can match human intelligence, making machines a viable substitute for human capital.

It is suggested that the research should focus on theoretical, innovative methodologies and best practices of data economy. It is also essential to integrate applied research for the data economy for impactful managerial implications. This study proposes a comprehensive business model that harmonizes the ethical, legal, technology and societal issues. This proposed business model will proffer solutions to the existing teething problems of the data economy. To get desired results from the

proposed project, professionals from ethics, legal, technology, environmentalists and experts from different fields that stake in the data economy should be part of the project. It should be a business model that can work across different fields of expertise without minor modification. Organizing workshops and seminars will contribute to the proposed project's success.

There is a growing impact of funding for data economy at the continent, international, national and local levels for policymakers. An example is the European Commission funding for data economy and the Foundation for Economic Education (Liikesivistysrahasto) in Finland, which has data economy as one of its bold themes more than two years ago. Because of the impact of the data economy research domain, the results of our study may simplify the decision-making of funding for the data economy in the future.

## 5. Conclusions

This study attempted to answer three research questions and combined two databases (WoS and Scopus), spanning 13 years focusing on FAIR data and data economy. The results from the bibliometric analysis generate interesting insights for the research community, data stakeholders and society at large. The article production from 2008 to 2021 shows exponential growth, and the first spike took place in 2016 and 2020. The growth is becoming steady, and the result certifies Roos and Kaliyaperumal as the authors with the most extended productivity timeline. Interestingly, Schultes excelled within the shortest time frame with high citations.

Also, the results show FAIR data concepts with dimensions of open science, data management, machine learning, big data, data sharing and data economy. This result indicates the importance of ethical issues related to data. The FAIR principle must predominate whether an organization is thinking of OD or a data economy. The study also reveals the importance and intervention of emerging technologies in FAIR and the data economy context. Further, the social structure of the study shows that Wilkinson is an influential author and has a cluster of collaborators in their network.

Regarding institution collaboration, institutions in The Netherlands have the most significant cluster, while Standard and Heriot-Watt top the list in their cluster and the University of Helsinki. The USA plays a central role in two clusters of countries and connects other countries from Europe, Australia and Asia. International collaboration also cuts across countries and portrays intercontinental collaboration. Our results show multilateral and bilateral collaborations. All the metrics examined signaled the advancement of data economy literature.

Further, the compound annual growth rate (CAGR) the Bibliometrix app generates for the literature on FAIR data and data economy is in tandem with the exponential growth discussed earlier. The CAGR is based on  $[\text{number of articles (final year)}/\text{number of articles (initial year)}]^{(1/n)} - 1$ . The computation is rooted in the study period: the number of articles in the initial year and the accumulated literature. The annual growth rate for this bibliometric study is 7.97%. This result shows the demand, value and performance of data-related literature.

The CAGR will help researchers, journals and other data stakeholders properly understand the data literature life cycle, either growth stage, maturity or decline that needs renewal. The CAGR percentage in this study shows that the data literature is at the growth stage, and this growth needs sustainability. Data relevance is progressing, and it is a key strategic asset for companies and academia now and in the future. This study combined FAIR data and data economy. It contributed to the literature on big data, information discovery and delivery by shedding light on the importance of the data economy's conceptual, intellectual and social structure.

This bibliometric study is a road map for future researchers, but the study is not without limitations. First, the study was limited to 2008–2021 but did not consider the earlier years of the data economy. Though it is an emerging field, this study did not account for the scanty literature before 2008. Though the literature was extracted from 1970, the study cut off 38 years of literature to sanitize the data. Because English is a broader acceptable means of communication, the study excluded other languages than English. Some of the non-English articles would have contributed to this study. This study scope is also limited to FAIR data and data economy. The future researcher can work around these limitations by extending the results of this study.

Based on this study, the researchers can expand on the intersection of sharing, platform and data economy. This research shows a dearth of theory building and testing in the research domain of the data economy. Future researchers should test existing theories and show a plausible explanation of the phenomenon of their investigation. Also, future researchers need to synthesize a wide range of literature with higher-level thinking skills to build theory around the data economy. It is also essential to combine methods to strengthen the existing methodology for data economy research. This study used bibliometric methods to unravel the proper position of scholars' contribution to the data economy. Future researchers can use the insights in this study to embark on empirical and longitudinal studies. This study, without any doubt, will open further discussion on the data economy.

## References

- Agata (2020), "Everything you need to know about the data economy", *Opendatasoft.com*, available at: [www.opendatasoft.com/blog/everything-you-need-to-know-about-the-data-economy](http://www.opendatasoft.com/blog/everything-you-need-to-know-about-the-data-economy) (accessed 16 July 2021).
- Agbo, F.J., Olaleye, S.A., Sanusi, I.T. and Dada, O.A. (2021c), "A review of thematic growth of international journal of education and development using ICT", *International Journal of Education and Development Using Information and Communication Technology*, Vol. 17 No. 2, pp. 17-36.
- Agbo, F.J., Oyelere, S., Suhonen, J. and Tukiainen, M.S. (2021b), "Scientific production and thematic breakthroughs in smart learning environments: a bibliometric analysis", *Smart Learning Environments*, Vol. 8 No. 1, pp. 1-25.
- Agbo, F.J., Sanusi, I.T., Oyelere, S.S. and Suhonen, J. (2021a), "Application of virtual reality in computer science education: a systemic review based on bibliometric and

content analysis methods”, *Education Sciences*, Vol. 11 No. 3, p. 142.

Akhavan, P., Ebrahim, N.A., Fetрати, M.A. and Pezeshkan, A. (2016), “Major trends in knowledge management research: a bibliometric study”, *Scientometrics*, Vol. 107 No. 3, pp. 1249-1264.

Alencar, P., Cowan, D., McGarry, F. and Palmer, R.M. (2014), “Multi-sectoral collaborative open data applications”, *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, IEEE*, pp. 64-73.

Allen, A.L. (2016), “Protecting one’s own privacy in a big data economy”, *Harvard Law Review Forum*, Vol. 130, p. 71.

Aria, M. and Cuccurullo, C. (2017), “Bibliometrix: an R-tool for comprehensive science mapping analysis”, *Journal of Informetrics*, Vol. 11 No. 4, pp. 959-975.

Aridi, A., Hayter, C.S. and Radosevic, S. (2021), “Windows of opportunities for catching up: an analysis of ICT sector development in Ukraine”, *The Journal of Technology Transfer*, Vol. 46 No. 3, pp. 701-719.

Baraibar-Diez, E., Luna, M., Odriozola, M.D. and Llorente, I. (2020), “Mapping social impact: a bibliometric analysis”, *Sustainability*, Vol. 12 No. 22, p. 9389.

Baždarić, K., Vrkić, I., Arh, E., Mavrinac, M., Marković, M.G., Bilić-Zulle, L., Stojanovski, J. and Malički, M. (2021), “Attitudes and practices of open data, preprinting, and peer-review – a cross sectional study on Croatian scientists”, *Plos One*, Vol. 16 No. 6, p. e0244529.

Bonti, A., Saini, A., Pham, T., Abdelrazek, M. and Pinto, L. (2021), “DSURVEY: ABlockchain-enhanced survey platform for the data economy”, *CS & IT Conference Proceedings*, Vol. 11, No. 1.

Celi, L.A., Citi, L., Ghassemi, M. and Pollard, T.J. (2019), “The PLOS ONE collection on machine learning in health and biomedicine: towards open code and open data”, *Plos One*, Vol. 14 No. 1, p. e0210232.

Chen, X., Lun, Y., Yan, J., Hao, T. and Weng, H. (2019), “Discovering thematic change and evolution of utilizing social media for healthcare research”, *BMC Medical Informatics and Decision Making*, Vol. 19 No. S2, pp. 39-53.

Choi, B.C. and Anita, W.P. (2008), “Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 3. discipline, inter-discipline distance, and selection of discipline”, *Clinical & Investigative Medicine*, Vol. 31 No. 1, pp. E41-E48.

Choi, B.C. and Pak, A.W. (2007), “Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 2. promoters, barriers, and strategies of enhancement”, *Clinical & Investigative Medicine*, Vol. 30 No. 6, pp. E224-E232.

Curry, E. and Sheth, A. (2018), “Next-generation smart environments: from system of systems to data ecosystems”, *IEEE Intelligent Systems*, Vol. 33 No. 3, pp. 69-76, doi: [10.1109/mis.2018.033001418](https://doi.org/10.1109/mis.2018.033001418).

del Pozo Cruz, B., McGregor, D.E., del Pozo Cruz, J., Buman, M.P., Palarea-Albaladejo, J., Alfonso-Rosa, R.M. and Chastin, S.F. (2020), “Integrating sleep, physical activity, and diet quality to estimate all-cause mortality risk: a combined compositional clustering and survival analysis of the national health and nutrition examination survey 2005–

2006 cycle”, *American Journal of Epidemiology*, Vol. 189 No. 10, pp. 1057-1064.

Desai, A. (2015), “A review on knowledge discovery using text classification techniques in text mining”, *International Journal of Computer Applications*, Vol. 111 No. 6.

Ding, L., Lebo, T., Erickson, J.S., DiFranzo, D., Williams, G.T., Li, X., Michaelis, J., Graves, A., Zheng, J.G., Shangguan, Z. and Flores, J. (2011), “TWC LOGD: a portal for linked open government data ecosystems”, *Journal of Web Semantics*, Vol. 9 No. 3, pp. 325-333, doi: [10.1016/j.websem.2011.06.002](https://doi.org/10.1016/j.websem.2011.06.002).

Dourish, P. and Gómez Cruz, E. (2018), “Datafication and data fiction: narrating data and narrating with data”, *Big Data & Society*, Vol. 5 No. 2, p. 2053951718784083.

Dunning, A., De Smaele, M. and Böhmer, J. (2017), “Are the FAIR data principles fair?”, *International Journal of Digital Curation*, Vol. 12 No. 2, pp. 177-195, doi: [10.2218/ijdc.v12i2.567](https://doi.org/10.2218/ijdc.v12i2.567).

Edwards-Schachter, M. and Wallace, M.L. (2017), “Shaken, but not stirred’: sixty years of defining social innovation”, *Technological Forecasting and Social Change*, Vol. 119, pp. 64-79.

Elgendy, N., Elragal, A. and Päiväranta, T. (2021), “DECAS: a modern data-driven decision theory for big data and analytics”, *Journal of Decision Systems*, Vol. 31 No. 4, pp. 337-373.

Elvy, S.A. (2017), “Paying for privacy and the personal data economy”, *Columbia Law Review*, Vol. 117, p. 1369.

Geneviève, L.D., Martani, A., Perneger, T., Wangmo, T. and Elger, B.S. (2021), “Systemic fairness for sharing health data: perspectives from Swiss stakeholders”, *Frontiers in Public Health*, Vol. 9.

Glassberg, S.E. (2018), “How to build great data products”, *Harvard Business Review*.

Hu, D.L.Y., Pan, L., Li, M. and Zheng, S. (2021), “A blockchain-based trading system for big data”, *Computer Networks*, Vol. 191, p. 107994.

Huang, Y.C., Tremouilhac, P., Nguyen, A., Jung, N. and Bräse, S. (2021), “ChemSpectra: a web-based spectra editor for analytical data”, *Journal of Cheminformatics*, Vol. 13 No. 1, pp. 1-9.

Huijboom, N. and Van den Broek, T. (2011), “Open data: an international comparison of strategies”, *European Journal of ePractice*, Vol. 12 No. 1, pp. 4-16.

Janssen, M., Charalabidis, Y. and Zuiderwijk, A. (2012), “Benefits, adoption barriers and myths of open data and open government”, *Information Systems Management*, Vol. 29 No. 4, pp. 258-268, doi: [10.1080/10580530.2012.716740](https://doi.org/10.1080/10580530.2012.716740).

Kassen, M. (2013), “A promising phenomenon of open data: a case study of the Chicago open data project”, *Government Information Quarterly*, Vol. 30 No. 4, pp. 508-513, doi: [10.1016/j.giq.2013.05.012](https://doi.org/10.1016/j.giq.2013.05.012).

Kaur, H. and Kumari, V. (2020), “Predictive modelling and analytics for diabetes using a machine learning approach”, *Applied Computing and Informatics*, Vol. 18 Nos 1/2, pp. 90-100.

Kiarashinejad, Y., Abdollahramezani, S. and Adibi, A. (2020), “Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures”, *Npj Computational Materials*, Vol. 6 No. 1, pp. 1-12.



- Kobayashi, S., Kane, T.B. and Paton, C. (2018), “The privacy and security implications of open data in healthcare”, *Yearbook of Medical Informatics*, Vol. 27 No. 1, pp. 41–47.
- Lahoti, P., Gummadi, K. and Weikum, G. (2019), “iFair: learning individually fair data representations for algorithmic decision making”, *2019 IEEE 35Th International Conference On Data Engineering (ICDE)*, doi: [10.1109/icde.2019.00121](https://doi.org/10.1109/icde.2019.00121)
- Lai, N.Y.G., Wong, K.H., Yu, L.J. and Kang, H.S. (2020), “Virtual reality (VR) in engineering education and training: a bibliometric analysis”, *Proceedings of the 2020 The 2nd World Symposium on Software Engineering*, pp. 161–165.
- Lammi, M. and Pantzar, M. (2019), “The data economy: how technological change has altered the role of the citizen-consumer”, *Technology in Society*, Vol. 59, p. 101157, doi: [10.1016/j.techsoc.2019.101157](https://doi.org/10.1016/j.techsoc.2019.101157).
- Lee, K. and Yoo, J. (2019), “How does open innovation lead competitive advantage? A dynamic capability view perspective”, *Plos One*, Vol. 14 No. 11, p. e0223405.
- Mai, J.E. (2016), “Personal information as communicative acts”, *Ethics and Information Technology*, Vol. 18 No. 1, pp. 51–57.
- Mayer-Schoenberger, V. and Cukier, K. (2013), *Big Data. A Revolution That Will Transform How we Live, Work, and Think*, John Murray Publishers, London.
- Meijer, R., Conradie, P. and Choenni, S. (2014), “Reconciling contradictions of open data regarding transparency, privacy, security and trust”, *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 9 No. 3, pp. 32–44.
- Mejias, U.A. and Couldry, N. (2019), “Datafication”, *Internet Policy Review*, Vol. 8 No. 4.
- Milletler, B. (2019), “Data economy: radical transformation or dystopia?”, *Frontier Technology Quarterly*.
- Molloy, J. (2011), “The open knowledge foundation: open data means better science”, *PLoS Biology*, Vol. 9 No. 12, p. e1001195, doi: [10.1371/journal.pbio.1001195](https://doi.org/10.1371/journal.pbio.1001195).
- Monino, J. (2016), “Data value, big data analytics, and decision-making”, *Journal of the Knowledge Economy*, Vol. 12 No. 1, pp. 256–267, doi: [10.1007/s13132-016-0396-2](https://doi.org/10.1007/s13132-016-0396-2).
- Murray-Rust, P. (2008), “Open data in science”, *Nature Precedings*, p. 1, doi: [10.1038/npre.2008.1526.1](https://doi.org/10.1038/npre.2008.1526.1).
- Musci, R.J., Fairman, B., Masyn, K.E., Uhl, G., Maher, B., Sisto, D.Y., Kellam, S.G. and Ialongo, N.S. (2018), “Polygenic score × intervention moderation: an application of discrete-time survival analysis to model the timing of first marijuana use among urban youth”, *Prevention Science*, Vol. 19 No. 1, pp. 6–14.
- Nobre, G. and Tavares, E. (2017), “Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study”, *Scientometrics*, Vol. 111 No. 1, pp. 463–492, doi: [10.1007/s11192-017-2281-6](https://doi.org/10.1007/s11192-017-2281-6).
- Ochs, C., Büttner, B. and Lamla, J. (2021), “Trading social visibility for economic amenability: data-based value translation on a “health and fitness platform”, *Science, Technology, & Human Values*, Vol. 46 No. 3, pp. 480–506.
- O’Connor, G.C. (2008), “Major innovation as a dynamic capability: a systems approach”, *Journal of Product Innovation Management*, Vol. 25 No. 4, pp. 313–330.
- Olaleye, S.A., Sanusi, I.T. and Dada, O.A. (2021), “Bibliometric structured review of mobile information systems”, *International Conference on Human-Computer Interaction, Springer, Cham*, pp. 284–297.
- Olaleye, S.A. (2020), “Visualizing cultural emotional intelligence literature: a bibliometric review 2001–2020”, in Laine, P., Némethová, I. and Wiwczarowski, T. (Eds), *Intercultural Competence at Work*. Publications of Seinäjoki University of Applied Sciences B. Reports, Seinäjoki: Seinäjoki ammattikorkeakoulu, Vol. 160, pp. 142–156, available at: <https://urn.fi/URN:NBN:fi-fe20201215100768>
- Oliveira, M., Barros Lima, G. and Farias Lóscio, B. (2019), “Investigations into data ecosystems: a systematic mapping study”, *Knowledge and Information Systems*, Vol. 61 No. 2, pp. 589–630, doi: [10.1007/s10115-018-1323-6](https://doi.org/10.1007/s10115-018-1323-6).
- Opher, A., Chou, A., Onda, A. and Sounderrajan, K. (2016), *The Rise of the Data Economy: driving Value through Internet of Things Data Monetization*, IBM Corporation, Somers, New York.
- Paul, J. and Feliciano-Cestero, M.M. (2021), “Five decades of research on foreign direct investment by MNEs: an overview and research agenda”, *Journal of Business Research*, Vol. 124, pp. 800–812.
- Paul, J., Merchant, A., Dwivedi, Y.K. and Rose, G. (2021), “Writing an impactful review article: what do we know and what do we need to know?”, *Journal of Business Research*, Vol. 133, pp. 337–340.
- Phan, N., Dou, D., Wang, H., Kil, D. and Piniewski, B. (2017), “Ontology-based deep learning for human behavior prediction with explanations in health social networks”, *Information Sciences*, Vol. 384, pp. 298–313.
- Rantanen, M., Hyrnsalmi, S. and Hyrnsalmi, S. (2019), “Towards ethical data ecosystems: a literature study”, *2019 IEEE International Conference On Engineering, Technology And Innovation (ICE/ITMC)*, doi: [10.1109/ice.2019.8792599](https://doi.org/10.1109/ice.2019.8792599)
- Reichman, O., Jones, M. and Schildhauer, M. (2011), “Challenges and opportunities of open data in ecology”, *Science*, Vol. 331 No. 6018, pp. 703–705, doi: [10.1126/science.1197962](https://doi.org/10.1126/science.1197962)
- Ruckenstein, M. and Schüll, N.D. (2017), “The datafication of health”, *Annual Review of Anthropology*, Vol. 46 No. 1, pp. 261–278.
- Sagr, M., Ng, K., Oyelere, S.S. and Tedre, M. (2021), “People, ideas, milestones: a scientometric study of computational thinking”, *ACM Transactions on Computing Education (TOCE)*, Vol. 21 No. 3, pp. 1–17.
- Shan, S., Luo, Y., Zhou, Y. and Wei, Y. (2019), “Big data analysis adaptation and enterprises’ competitive advantages: the perspective of dynamic capability and resource-based theories”, *Technology Analysis & Strategic Management*, Vol. 31 No. 4, pp. 406–420.
- Sharma, D., Taggar, R., Bindra, S. and Dhir, S. (2020), “A systematic review of responsiveness to develop future research agenda: a TCCM and bibliometric analysis”, *Benchmarking: An International Journal*, Vol. 27 No. 9, pp. 2649–2677.
- Singh, S., Akbani, I. and Dhir, S. (2020), “Service innovation implementation: a systematic review and research agenda”, *The Service Industries Journal*, Vol. 40 Nos 7/8, pp. 491–517.
- Sumbal, M.S., Tsui, E. and See-to, E.W. (2017), “Interrelationship between big data and knowledge management: an exploratory study in the oil and gas sector”,

- Journal of Knowledge Management*, Vol. 21 No. 1. doi: [10.1108/JKM-07-2016-0262](https://doi.org/10.1108/JKM-07-2016-0262)
- Tabernik, D., Šela, S., Skvarč, J. and Skočaj, D. (2020), "Segmentation-based deep-learning approach for surface-defect detection", *Journal of Intelligent Manufacturing*, Vol. 31 No. 3, pp. 759-776.
- Tanhua, T., Pouliquen, S., Hausman, J., O'Brien, K., Bricher, P., De Bruin, T., Buck, J.J., Burger, E.F., Carval, T., Casey, K.S. and Diggs, S. (2019), "Ocean FAIR data services", *Frontiers in Marine Science*, Vol. 6, doi: [10.3389/fmars.2019.00440](https://doi.org/10.3389/fmars.2019.00440).
- Tenopir, C., Rice, N.M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R. and Sandusky, R.J. (2020), "Data sharing, management, use, and reuse: practices and perceptions of scientists worldwide", *Plos One*, Vol. 15 No. 3, p. e0229003.
- Tomic, S.D.K. and Fensel, A. (2013), "OpenFridge: a platform for data economy for energy efficiency data", *2013 IEEE International Conference on Big Data, IEEE*, pp. 43-47.
- Urquhart, L., Sailaja, N., Lindley, J., Mcauley, D. and Forrester, I. (2021), Human data interaction through design: an exploratory from theory to practice using design as a vehicle. *Proceedings of ACM SIGCHI 2020 (Extended Abstracts)*, 8-13 May, Yokohama, Japan/Online.
- Van Dijck, J. (2014), "Datafication, dataism and dataveillance: big data between scientific paradigm and ideology", *Surveillance & Society*, Vol. 12 No. 2, pp. 197-208, doi: [10.24908/ss.v12i2.4776](https://doi.org/10.24908/ss.v12i2.4776)
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J. (2016), "The FAIR guiding principles for scientific data management and stewardship", *Scientific Data*, Vol. 3 No. 1, pp. 1-9.

- Wise, J., de Barron, A.G., Splendiani, A., Balali-Mood, B., Vasant, D., Little, E., Mellino, G., Harrow, I., Smith, I., Taubert, J. and van Bochove, K. (2019), "Implementation and relevance of FAIR data principles in biopharmaceutical R&D", *Drug Discovery Today*, Vol. 24 No. 4, pp. 933-938, doi: [10.1016/j.drudis.2019.01.008](https://doi.org/10.1016/j.drudis.2019.01.008)
- Yildirim, M., Okay, F.Y. and Özdemir, S. (2021), "Big data analytics for default prediction using graph theory", *Expert Systems with Applications*, Vol. 176, p. 114840.
- Zech, H. (2016), "A legal framework for a data economy in the European digital single market: rights to use data", *Journal of Intellectual Property Law & Practice*, Vol. 11 No. 6, pp. 460-470, doi: [10.1093/jiplp/jpw049](https://doi.org/10.1093/jiplp/jpw049).
- Zhao, Y. and Fan, B. (2018), "Exploring open government data capacity of government agency: based on the resource-based theory", *Government Information Quarterly*, Vol. 35 No. 1, pp. 1-12.
- Zuboff, S. (2015), "Big other: surveillance capitalism and the prospects of an information civilization", *Journal of Information Technology*, Vol. 30 No. 1, pp. 75-89, doi: [10.1057/jit.2015.5](https://doi.org/10.1057/jit.2015.5)
- Zuiderwijk, A. and Janssen, M. (2014), "Open data policies, their implementation and impact: a framework for comparison", *Government Information Quarterly*, Vol. 31 No. 1, pp. 17-29, doi: [10.1016/j.giq.2013.04.003](https://doi.org/10.1016/j.giq.2013.04.003).
- Zuiderwijk, A. and Janssen, M. (2014), "The negative effects of open government data-investigating the dark side of open data", *Proceedings of the 15th Annual International Conference on Digital Government Research*, pp. 147-152.

### Corresponding author

Friday Joseph Agbo can be contacted at: [friday.agbo@uef.fi](mailto:friday.agbo@uef.fi)