DESIDOC Journal of Library & Information Technology, Vol. 42, No. 2, March 2022, pp. 119-129, DOI : 10.14429/djlit.42.2.17503

Data Capital: A Systematic Literature Review

Arief Ramadhan

Computer Science Department, Bina Nusantara University, Jakarta - 11480, Indonesia
E-mail: arief.ramadhan@binus.edu

ABSTRACT

Data capital is one of the newly emerging concepts in information technology. It emerged in early 2015, and currently, we have an opportunity to explore it in more detail. Data capital is a way to treat data as an asset that is used to increase revenue, gain profit or generate new income. There has never been a proper review of data capital. Therefore, this study aims to conduct a systematic literature review of academic articles on data capital. We use the Kitchenham method and combine it with the narrative synthesis as prescribed by Cochrane. The obtained results show that there is still very limited academic research related to the concept of data capital. The reviewed articles focus on five issues, i.e., ethical, legal, infrastructure, data capital application, and data capital conceptualisation. This study is the first attempt to conduct a detailed systematic literature review in the data capital context. It provides a broad insight into data capital. It improves our understanding of how data capital is currently developing, why it is important, and the impact of data capital incorporation. Any issues that have been researched in the data capital context are also revealed in this study including the type of research performed, approach used, and method chosen. These insights will encourage other researchers to delve into data capital research.

Keywords: Data capital; Data assets; Information technology; Systematic literature review

1. INTRODUCTION

There are many decisions that industry players can make. However, the most important are usually decisions related to increasing revenues and decreasing costs; these decisions determine whether the company survives or not. Therefore, the role of data becomes very important. Biddulph[1] calls data as a weapon that must be used to its full potential. If the organisation does not use its data properly as a competitive weapon, then it may operate less efficiently, lose market share, and lose millions in new revenue streams and profits[1]. That kind use of data is associated with new terminology, i.e., data capital.

The term data capital became popular in 2015. At that time, Sonderegger[2], Oracle's Big Data strategist, immediately predicted that data capital would "replace big data as the big topic of boardroom conversation". Big Data is more focused on how to manage data that has 3V properties, i.e., Variety, Velocity, and Volume. However, data capital is more focused on how to increase revenue using data, and it does not have to be big data. Small and simple data can also be a capital for obtaining value. Data in the form of ordinary transactions owned by a retailer can be more valuable than the goods sold by the retailer[1].

In data capital terminology, data is a type of capital that is on par with other capital such as financial and human capital[3]. Because it is a capital, then like any other capital, data capital "can be used to produce goods or services, and it can have long term value for the company"[3]. Data can also be considered as raw material for creating something valuable, for example, to create digital services[3]. This new form of capital is essential for businesses to survive and thrive in the digital age[4].

As a capital, data can be used to increase company profits, for example in terms of exploring consumer insights and demand forecasting[1]. Giant companies (e.g., Google, Amazon, LinkedIn, and Alibaba) have been built on this very foundation[5]. Therefore, data must be considered as an important asset. Some companies in the field of Information and Technology (IT), even though they do not have many physical assets but possess many data assets, have multiple advantages compared to companies that only focus on physical assets[3]. Some examples of companies that have many data assets are Airbnb, Facebook, and Netflix[3].

An example of data as capital is illustrated as follows[5]. Suppose we sell our house through a realtor service. The realtor includes our home sales information on the realtor website. Suppose that a month later, our house is sold. Not long after, various promotional flyers suddenly arrive at our house. The promotional flyers are from various moving services. Then, a question arises, how do these companies know that we will soon move from our home?

The simple answer is that someone periodically extracts the data that are on the realtor's website. Every time an extraction is performed, the new data are compared with the previous ones. When there is a change (e.g., the data on a previously listed house are no longer available), the house is marked as sold. All data on the houses that have been sold are then neatly

Received : 11 September 2021, Revised : 01 December 2021
Accepted : 12 January 2022, Online published : 28 February 2022

arranged and sold to various moving service companies. That is how the moving services know which homes have been sold and are likely to need help when moving.

Another example is when a senior high school student graduates from his/her school, soon, various promotional flyers arrive at his/her house. The flyers come from various universities or colleges that invite students to apply to them. The question is how do all these educational institutions know the student's address and how do they know that the student has successfully graduated from high school? Similar to the previous example, the data on students who have graduated are capitalised and sold to universities or colleges for further use by them to send promotional flyers to targeted students.

The two abovementioned examples show how data can be capitalised to generate financial returns; this is called data capital. Data capital is slightly different from data mining. Data capital can use data mining, but it may also not use data mining. As defined by Tan *et al.*[6], "Data mining is the process of automatically discovering useful information in large data repositories". There is the keyword "automatic" in the definition. Usually, the "automatic" process is performed using several sophisticated algorithms for predictive modeling cluster analysis, anomaly detection, or association analysis[6]. Furthermore, Tan *et al.*[6] stated that not all information discovery tasks are considered to be data mining. For example, looking up individual records from database management systems or finding a particular web page using a query or search engine is not data mining. Data capital does not necessarily use a sophisticated algorithm. As shown above, the data can be capitalised by merely periodically retrieving the realtor's website and manually comparing the data. According to the explanation by Tan *et al.*[6], the process in the abovementioned example uses data as capital through ordinary information discovery without data mining.

The approach to treating data as capital has become more widespread in the industrial world[7]. Several large IT solution companies (e.g., Accenture and Dell Inc.) offer data capital solutions among their services[8,9]. However, very little academic research has focused on data capital. There has never been a review of data capital. Therefore, in this paper, we conduct a systematic literature review of scientific articles to elucidate the status of research on data capital.

2. SCOPE AND OBJECTIVES

Our research focuses on data capital. Taking into account the various explanations in the Introduction section, we view data capital as a way to treat data as an asset that is used to increase revenue, gain profit or generate new income.

Our research objective is to conduct a systematic literature review of academic articles on data capital. We want to highlight the progress of research on data capital. We identify several issues related to data capital addressed in academic research. The type of research, research approaches, and research methods are explored. The impact of incorporating data capital is also reviewed.

3. METHODOLOGY

This systematic literature review differs from the traditional review because systematic literature review (i.e., systematic review) adopts a replicable, scientific, and transparent process[10]. Systematic literature review has been accepted as a valid research methodology[11]. By performing a systematic review, the review is more thorough, fair, and is perceived as being fair[12]. A systematic review uses an exhaustive search in designated databases[13] to identify studies that fit the search keywords. Then, the pre-specified inclusion criteria are used to justify whether the identified studies are relevant or not to be included in the next process[11]. Because the data capital terminology is relatively new, we will focus on broad research questions in our systematic review. Therefore, we adopt the mapping studies-type systematic literature review approaches used by Kitchenham *et al.*[14]. We also combine the Kitchenham method with the narrative synthesis strategy of Cochrane[15], as follows.

3.1 Research Questions

The research questions that are addressed by this systematic literature review are

RQ1: How is the progress of research in data capital?
RQ2: What issues related to data capital are addressed?
RQ3: What type of research do data capital researchers address?
RQ4: What research approaches do data capital researchers use?
RQ5: What research methods do data capital researchers use?
RQ6: What are the impacts of incorporating data capital?

3.2 Search Process

To obtain broad and comprehensive search results, we searched in 9 reputable electronic research indexing databases, i.e., ACM Digital Library, Emerald Insight, IEEE Xplore, ScienceDirect, SAGE Publications, Nature, SpringerLink, Scopus and Web of Science (WoS). Because the term data capital has become popular since 2015, we limited the search to articles starting from 2015. We limited our search to only academic and peer-reviewed articles to ensure their quality. Specifically, we only considered articles from scientific journals or conference proceedings. In addition, we searched only to the titles and abstracts to ensure that only relevant articles were identified.

If a company wants to develop an enterprise architecture for its data capital, then there are several issues that need to be considered, i.e., data equality and data liquidity[4]. One of the principles of data capital is that data comes from activity[4]. Thus, each company must be able to record all activity in the form of data. This is especially relevant for activities related to each process in the business value chain[3]. This principle involves datafying and datafication activities[3]. Therefore, we used datafy, datafication, data equality, and data liquidity as keywords in the search process. These keywords along with the phrase "data capital" were used to form the search string. In addition, for the word "datafy", its present continuous tense form (i.e., "datafying") was also added to the search string. Thus, the search string used in the search process was as follows: "data capital" OR "datafy" OR "datafying" OR "datafication" OR "data equality" OR "data liquidity".

Table 1. Search results from each database

Database name	Search string	Search results	Description
Emerald Insight	title:”data capital” OR (title:”datafy”) OR (title:”datafying”) OR (title:”datafication”) OR (title:”data equality”) OR (title:”data liquidity”) OR (abstract:”data capital”) OR (abstract:”datafy”) OR (abstract:”datafying”) OR (abstract:”datafication”) OR (abstract:”data equality”) OR (abstract:”data liquidity”)	13	It was not filtered for only journals or conference proceedings. Therefore, the total search results still include articles in other forms, such as book chapters, expert briefings, and others.
IEEE Xplorer	(“Document Title”:”data capital”) OR (“Abstract”:”data capital”) OR (“Document Title”:”datafy”) OR (“Abstract”:”datafy”) OR (“Document Title”:”datafying”) OR (“Abstract”:”datafying”) OR (“Document Title”:”datafication”) OR (“Abstract”:”datafication”) OR (“Document Title”:”data equality”) OR (“Abstract”:”data equality”) OR (“Document Title”:”data liquidity”) OR (“Abstract”:”data liquidity”)	9	It was set so that searches were carried out only on titles and abstracts and only on journals and conference proceedings.
ScienceDirect	“data capital” OR “datafy” OR “datafying” OR “datafication” OR “data equality” OR “data liquidity”	30	It was not filtered for only journals or conference proceedings. In addition, in ScienceDirect, when it is set to search for Title and Abstract, it will automatically search for keywords from the article. So, it was not only set for the title and abstract.
SAGE Publications	for [[Title “data capital”] OR [Title “datafy”] OR [Title “datafying”] OR [Title “datafication”] OR [Title “data equality”] OR [Title “data liquidity”]]	51	It was not filtered to only journals or conference proceedings. Therefore, the total search results still include articles in other forms, such as book review, introduction, and others.
	for [Abstract “data capital”] OR [Abstract “datafy”] OR [Abstract “datafying”] OR [Abstract “datafication”] OR [Abstract “data equality”] OR [Abstract “data liquidity”]	175	
Nature	“data capital” OR “datafy” OR “datafying” OR “datafication” OR “data equality” OR “data liquidity”	17	It was not set to only titles and abstracts and was not filtered for only journals or conference proceedings.
Springer Link	“data capital” OR “datafy” OR “datafying” OR “datafication” OR “data equality” OR “data liquidity”	258	It was not set to only titles and abstracts.
ACM Digital Library	[Title: “data capital”] OR [Title: “datafy”] OR [Title: “datafying”] OR [Title: “datafication”] OR [[Title: “data equality”] AND [Title: “data liquidity”] AND [Abstract: “data capital”] AND [Abstract: “datafy”] AND [Abstract: “datafying”]] OR [Abstract: “datafication”] OR [Abstract: “data equality”] OR [Abstract: “data liquidity”] AND [Publication Date: (01/01/2015 TO *)]	15	It was not filtered for only journals or conference proceedings.
Scopus	TITLE (“data capital”) OR ABS (“data capital”) OR TITLE (“datafy”) OR ABS (“datafy”) OR TITLE (“datafying”) OR ABS (“datafying”) OR TITLE (“datafication”) OR ABS (“datafication”) OR TITLE (“data equality”) OR ABS (“data equality”) OR TITLE (“data liquidity”) OR ABS (“data liquidity”) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015)) AND (LIMIT-TO (DOCTYPE , “ar”) OR LIMIT-TO (DOCTYPE , “cp”)) AND (EXCLUDE (PUBYEAR , 2022)) AND (LIMIT-TO (OA , “all”))	244	It was set so that searches were carried out only on titles and abstracts and only on journals and conference proceedings.
Web of Science	(TI=(“Data Capital” or “Datafication” or “Datafying” or “Datafy” or “Data liquidity” or “Data equality”)) OR AB=(“Data Capital” or “Datafication” or “Datafying” or “Datafy” or “Data liquidity” or “Data equality”))	234	It was set so that searches were carried out only on titles and abstracts and only on journals and conference proceedings.

3.3 Study Selection

During this stage, we screened the identified articles. We exclude an article if it is a duplicate. We also used several inclusion and exclusion criteria. We read the article and move it to the next stage if it met all inclusion criteria without satisfying any exclusion criteria.

3.3.1 The Inclusion Criteria

- There are issues in the context of data as capital addressed in the article, i.e., treating data as an asset and using them to increase revenue, gain profit or generate new income
- The article is published in a scientific journal or conference proceedings
- The article is written in English
- The article is a full paper (not a PowerPoint presentation or extended abstract)

3.3.2 The Exclusion Criteria

- There is no full access to the article
- The article does not have full text (e.g., only abstract exists)
- There is no information about the name of the scientific journal or conference
- The search keywords did not match the text in the title or abstract of the article
- The article is not a primary study (i.e., a secondary study or tertiary study).

3.4 Quality Assessment

In this stage, the full text of the selected article was read. A quality assessment checklist was used to determine the quality of the article. We read the entire article carefully and in detail. The items in the quality assessment checklist were adopted from Zuiderwijk et al.[16] and established based on suggestions from Kitchenham and Charters[12]. The quality assessment checklist is as follows:

- The research questions or research objectives are stated clearly
- The research approach can be inferred
- The research method is described in detail
- The research questions are answered or the research objectives are attained
- There are clear statements of the findings.

3.5 Data Extraction

To address the research questions, several data categories were derived from the research questions. There are data that can be found/determined and extracted easily, for example, title, author's name, year of publication, and name of journal/conference. Several other data require more effort to be extracted. For example, to address RQ4, we extracted the research approach used of an article by observing the form of data processed by the article. If the data processed by the article is a number, then we set that the article uses a quantitative approach, if the data processed by the article is text, then we set it as qualitative, and if both, we set it as a mixed approach. As another example, to address the RQ5, we determined the methods by thoroughly reading the Introduction, Research Methodology, Discussion, and Conclusion sections of the article. We synthesised the sentences that are residing in those sections to conclude the research methods.

3.6 Data Synthesis

In this stage, we adopted several narrative synthesis strategies from Cochrane[15]. We described each of the included studies in the form of summarisation, group the studies, tabulate the results, transform the data, and translate the data. We then delivered our result in accordance with each research question. The results obtained from the search process, study selection, quality assessment, and data extraction stages were also used to answer the research questions. We explain how these results answer the research questions.

4. RESULTS AND DISCUSSION

4.1 Search Process Results

The search was carried out on 9 databases, which were described in Section 3.2. The search string is then adjusted to comply with the rules of each database. We use the advanced search tool on each database. In addition, the search year is set from 2015. The search results from each database are shown in Table 1.

4.2 Study Selection Results

First, we excluded duplicate articles, i.e. articles those appeared in more than one database. Most of the duplication occurs because the articles are already indexed in the original publisher (e.g. Springer), as well as indexed in Scopus and the Web of Science. Another example is the search results on SAGE Publications, due to the limited availability of advanced search facilities, the search process needs to be done twice, based on the Title and based on the Abstract, so that there may be duplication of search results for certain articles in SAGE Publications. In total, there were 357 duplications. So, there are 689 articles that we selected based on inclusion and exclusion criteria.

In this stage, we made a selection based on the inclusion and exclusion criteria described in Section 3.3. Several inclusion and exclusion criteria were actually implemented since the search process in sub-section 4.1. For example, exclusion criteria no. 1 was applied using the advanced search facility of the database so that the search results in Table 1 eventually passed exclusion criteria no. 1. As another example, inclusion criteria no. 2 and exclusion criteria no. 4 were directly applied during the search process on several databases so that the search results were more focused. The description column in Table 1 describes which databases directly applied inclusion criteria no. 2 and exclusion criteria no. 4.

In this section, we only focused on the inclusion criteria and exclusion criteria that were not applied in sub-section 4.1. Thus, we are fully focused on inclusion criteria no. 1, 3, and 4. For the exclusion criteria, we focused on exclusion criteria no. 2, 3, and 5. Inclusion criteria no. 2 and exclusion criteria no. 4 still need to be done for some databases, because, as shown in Table 1, not all databases were filtered based on inclusion criteria no. 2 and exclusion criteria no. 4.

Many articles are excluded because the search keywords did not match with the text in the title or abstract of the article (exclusion criteria no. 4). There were 249 articles excluded because of exclusion criteria no. 4. Thus, the remaining 440 articles will be processed.

For inclusion criteria no. 2, many articles are excluded because they are in the form of a book chapter, expert briefing, book review, comment, opinion, or introduction document. There are 42 articles that are excluded based on inclusion criteria no 2. So there are 398 articles left.

Furthermore, based on inclusion criteria no 3, there are several articles that are not written in English. There are 30 articles that are excluded by this inclusion criteria. So, there are remaining 368 articles. Furthermore, there are 9 other articles that are excluded based on other criteria. So, there are 359 articles left.

If we pay attention, there are still a lot of articles that have not been filtered, that are 359 articles. We analyse that it happened because of the use of the keyword “datafication” during the search. If we try to remove the word “datafication” during the search process, the results will be significantly reduced. Therefore, in relation to the first inclusion criteria, we were very strict at focusing on the context of data capital, which treats data as an asset that is used to increase revenue, gain profit, or generate new income.

To ensure relatedness, we use the search facility in the PDF reader application to search for certain words related to data capital in the remaining articles. The search was carried out on the entire contents of the text. The words we were looking for were “asset”, “capital”, “revenue”, “income”, and “profit”. If one of those words is found in the article, then we read the article carefully. We read carefully whether the discussion in the article is really related to data capital. If there are only one or two sentences that mention data as capital, it will be excluded. Articles that discusses data in other contexts will also not be selected. Based on this inclusion criteria, we obtained 17 articles that were really related to data capital data, they discussed data as capital in depth. So we excluded 342 articles. The number of articles that were screened due to each inclusion and exclusion criteria at this stage is shown in Table 2.

Table 2. Number of articles excluded based on the inclusion and exclusion criteria

No. of inclusion criteria	# of articles excluded
1	342
2	42
3	30
4	2
No. of exclusion criteria	**# of articles excluded**
1	0*
2	1
3	1
4	271
5	5

*was done in the search process

Based on the selection process that is depicted in Table 2, there were 17 articles were obtained that could be included in the next stage.

4.3 Quality Assessment Results

In this stage, we read all 17 articles. We conducted an assessment based on the quality assessment checklist that is shown in Section 3.4. After studying the entire article, we did not have any major quality concerns. Apart from that, we had several minor concerns, especially related to the research method. Only five articles stated the research method clearly. The other articles did not present the research method in detail, for example, a lack of information about the type of method used. However, all the included articles were qualified enough to be used for further analysis and were considered to pass the quality assessment.

4.4 Data Extraction Results

The data extraction process and the kinds of data being extracted followed what is described in Section 3.5. The results of this data extraction will be further elaborated on when answering all research questions in the next sub-section. The extracted data for all 17 articles are shown in Table 3. To make the table more concise, the data for the names of the authors, the title of the articles, the publication name, and the year of publication are combined under “Article Information”, which is linked with the associated reference in the References section.

4.5 Data Synthesis Results

4.5.1 RQ1: How is the Progress of Research in Data Capital?

As shown in Table 3, this systematic literature review was conducted from 2015, there were only 17 research articles that were highly related to data capital. We use the transform data narrative synthesis strategy from Cochrane[15] by drawing the distribution of each article by year of publication, as shown in Fig. 1. Most of the articles were published in 2020, which was approximately 29.42 per cent. In general, there was an increasing trend of research related to data capital year over year.

Even though there is an increase in the number of research related to data capital, it is still very rare. Although this small number is very worrying, on the other hand, this shows the opportunities for the emergence of new research in this field. It is necessary to increase the awareness of researchers related to data capital. Moreover, the growth of data in the future will greatly support research related to data, especially data capital. Several research topics on data capital can be considered, for example, the enablers and inhibitors on future intention to adopt data capital, the implementation model of data capital, and the effectiveness of data capital to increase company revenue.

4.5.2 RQ2: Are Issues Related to Data Capital Addressed?

After reading all 17 articles thoroughly and by adopting the narrative synthesis strategy from Cochrane[15], we can group the articles into five main issues, ie, ethical issues, legal issues, infrastructure issues, data capital application issues, and issues

Table 3. Data extraction results

Article information	Publication type	Research type	Research approach	Research methods
Gumbus and Grodzinsky[18]	Journal	Basic	Qualitative	Bibliographic Revision
Siow et al.[17]	Conference Proceedings	Applied	Qualitative	Descriptive explanation
Jesse[19]	Journal	Basic	Qualitative	Bibliographic Revision
Couldry and Yu[20]	Journal	Basic	Qualitative	Discourse deconstruction
Sadowski[21]	Journal	Basic	Qualitative	Bibliographic Revision
Breidbach and Maglio[22]	Journal	Basic	Qualitative	Midrange theorizing approach that consists of content analysis, a general metatheoretical classification scheme, and applying the framework into the context
Tang et al.[23]	Journal	Basic	Qualitative	Bibliographic Revision
Dencik[24]	Journal	Basic	Qualitative	Bibliographic Revision
Hracs and Webster[25]	Journal	Basic	Qualitative	Bibliographic Revision
Hagen[26]	Journal	Basic	Mixed-methods	Survey, Descriptive Statistics, Intervies
Maasø and Hagen[27]	Journal	Basic	Qualitative	semi-structured interviews, text interpretations
Lehtiniemi[28]	Journal	Basic	Qualitative	Bibliographic Revision
Younes[29]	Journal	Basic	Qualitative	Bibliographic Revision
Larsson[30]	Journal	Basic	Qualitative	Bibliographic Revision
Mejias and Couldry[31]	Journal	Basic	Qualitative	Bibliographic Revision
Goodridge et al.[32]	Journal	Basic	Mixed Methods	Bibliographic Revision, Economical Analysis
Lun[33]	Journal	Basic	Qualitative	Bibliographic Revision

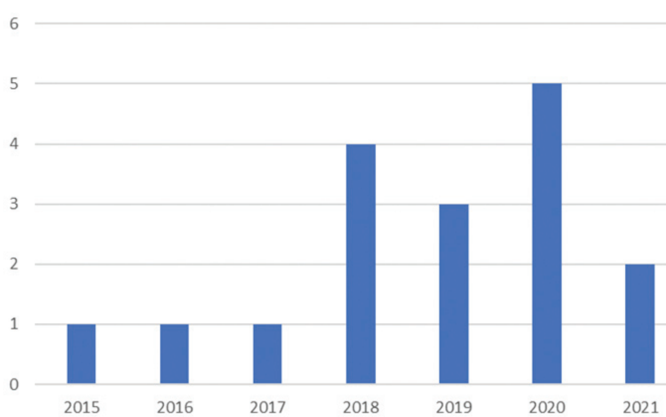

Figure 1. Distribution of articles by year of publication.

regarding the conceptualisation of data capital in more depth. Regarding ethical issues, five articles highlighted this issue. Legal issues were raised by one article. Infrastructure issues were the main discussion in the two articles. Data capital application in certain fields is discussed by 5 articles. However, the other four articles enriched the concept of data capital. These results are shown in Annexure I.

As shown in Annexure I, most of the issues raised are at the conceptual or theoretical level. Only two articles focused on technical aspects. Thus, future research of data capital related to technical aspects can be improved. In addition, no academic study has specifically discussed how to construct data capital in step by step manner. No specific study discussed the model or framework for data capital. Thus, this opens up opportunities for new research related to the construction, models, and frameworks of data capital.

4.5.3 RQ3: What Type of Research Do Data Capital Researchers Address?

To answer this question, we use the grouping narrative synthesis strategy from Cochrane[15]. We borrow the definition from Vogt[34] that basic research is research that is undertaken with the primary goal to advance knowledge and theoretical understanding rather than solving some practical problems. Meanwhile, applied research is research that was undertaken with the intention of applying the results to some specific problems[34].

Based on Table 3, there is only one article that is considered applied research, which is the study conducted by Siow *et al.*[17]. Meanwhile, the other 16 articles are considered basic research, which focuses on increasing knowledge related to data capital.

The result shows that more applied research is needed to clarify how data capital is applied in the real world. Research related to infrastructure, technology adoption, and governance of data capital in organisations are still very open to being carried out. Several information technologies related to Industry 4.0 are worth being discussed in relation to data capital. There are many technologies related to industry 4.0 to consider[35-36]. In addition, several future research projects that focus on discussing real examples of applying certain algorithms are also important, for example, how to utilise machine learning or artificial intelligence to leverage the data to create new income or increase revenue in the context of data capital. Some new emerging methods, such as deep learning[37,38] can be attempted.

4.5.4 *RQ4: What Research Approaches Do Data Capital Researchers Use?*

We also use the grouping narrative synthesis strategy from Cochrane[15] to answer this question. We have examined that 15 of the articles reviewed in this study are conducted based on text analysis. This is a characteristic of qualitative research[39]. The authors of reviewed articles explore existing textual theories or knowledge, then analyse, interpret, and relate them to answer their respective research objectives. As implied by Recker[40], the quality of qualitative research is highly dependent on interpretation from the researcher as an instrument. Thus, it can be concluded that the approaches used in the 15 articles that we reviewed were qualitative (Table 3).

While the other two articles combine a qualitative approach with a quantitative approach. Recker[40] calls the merging as Mixed Method. In these two studies, apart from analyzing text, numerical data was also used in their analysis. Therefore, the two articles were included in the Mixed Method category. This provides opportunities for new research related to data capital using a quantitative approach to add to the repertoire of data capital research.

4.5.5 *RQ5: What Research Methods Do Data Capital Researchers Use?*

Only five articles mentioned the research method they used. The other 12 articles did not mention the research method they used. Thus, we carefully read the remaining 12 articles that did not mention their research methods. Therefore, we adopted one translating data narrative synthesis strategy from Cochrane[15]. We used thematic analysis to identify areas in common between articles. We attempted to find patterns in how they conducted and explained their research.

From the results of our analysis, we found that the 12 articles used the same pattern: they reviewed previous research or documents and then linked them all. We followed a strategy from Fernández-Rovira *et al.*[41], to include those types of research in the bibliographic revision method group. This is also in line with what was stated by PAHO[42] in which bibliographic revision is when researchers try to revise and cite already published research or knowledge to place their research in context. In these studies, the methods used were mostly text analysis. Therefore, future research on data capital can explore methods other than text analysis. Methods based on statistical data processing or mathematical modeling are an interesting consideration.

4.5.6 *RQ6: What are the Impacts of Incorporating Data Capital?*

To answer this research question, we adopted one of the narrative synthesis strategies from Cochrane[15], which is summarisation. Several impacts can result from the implementation of data capital. The positive impact was put forward by Siow *et al.*[17], who stated that data capital makes the process of driving innovation faster and easier. Lun[33] revealed that good use of data will improve financial performance. Broadly speaking, Goodridge *et al.*[32] revealed that expanding the definition of investment in software and data increases various economic measures positively.

The negative impact was stated by Breidbach and Maglio[22], who said data capital can lead to several unethical practices. Gumbus and Grodzinsky[18] revealed that data capital can result in discrimination and injustice. Dencik[24] pointed out that the application of the concept of data capital in a data-driven system can introduce “inequalities, discriminate or exclude certain groups, or can dehumanise interaction and decision making around contentious and sensitive issues”. Prainsack[25] stated that the bad consequences of nudging based on data could outweigh the benefits.

Several controversies related to social justice have emerged by datafication, for example, datafication threatens the basic rights of the self[31] consumers’ privacy may be compromised by the extensive data monitored and analysed[29]. Another negative impact of data capital, especially in the music industry, is that it can result in only raising global superstars and mega-events, as well as making local and small events increasingly abandoned[27].

In addition to the impacts that show a positive tone or negative impact, there are several other impacts that are relatively neutral and focus more on the need for certain concepts, such as those about the concept, legal, or infrastructure for better application of data capital. Tang *et al.*[23] imply that any political determinism or interventions concerning inequalities in the exploitation of data resources have to be considered. A concept is needed that allows data producers to clearly derive profit from data[23]. Couldry and Yu[20] argued that due to the development of data collection, there needs to be a new regulation that focuses more on data collection and not just on data use because the current regulations do not focus on data collection. Lehtiniemi[28], Hagen[26], and Larsson[30] also revealed the need for the involvement of policymakers and copyright bodies to carry out regulations. Regarding regulations, Sadowski[21] also conveyed the necessary regulations regarding the data types that companies may collect, how they can collect data, where data will be sent and stored, and how much data can be possessed by companies. In addition, he also implies the need for a political response from the government regarding the new model of data ownership and data governance[21].

Jesse[19] said that the impact of data capital is that it requires a tight alignment between IT and operational strategy and that new organisational responsibilities need to be formed, the ways by which analytical experts and data scientists need to be connected with frontline managers and executive decision-makers to have an impact. Companies also have to evaluate their business models and make them more adaptable to the monetisation of the data[19].

Some of the impacts can open opportunities for future research to maximise the positive impacts, overcome the negative impacts, or implement other suggested impacts. Preliminary research using a dual-factor model[43,44] to find out the enablers and inhibitors on future intention to adopt data capital will be interesting.

5. LIMITATIONS

Although previous systematic literature reviews are replicable, scientific, transparent, thorough, and fair[11,13], choices still had to be made in relation to the research question, the database used, and the searching keywords. This study conducted a review related to the six questions. Future researchers may change the questions or add some other questions to obtain different results. This study has attempted to use a wide variety of the nine different main databases, i.e., the ACM Digital Library, Emerald Insight, IEEE Xplore, ScienceDirect, SAGE Publications, Nature, SpringerLink, Scopus and Web of Science (WoS). The choice of other databases might lead to different results, for example, using the Indian Citation Index (ICI).

The selected searching keywords may limit the number of articles collected in the search process. We used six searching keywords to catch as many related articles as possible. However, the use of different searching keywords might lead to different results.

To ensure that only strongly related articles were found, the search was only carried out on titles and abstracts. We used all the advanced facilities available in each database. Articles containing the searching keywords in the body but not in the title and abstract were not considered. A more extensive search that is also carried out besides only on the title and abstract may obtain different results, although the fitness of the articles found may not be sufficiently suitable.

We only considered scientific journals or conference proceedings in our search process. We omitted book chapters and gray literature, such as technical reports, dissertations, and master or graduate theses. Several good-quality book chapters and gray literature might exist that can enrich the results of this study. This study included English articles only. It is possible that several other related articles discuss data capital in different languages.

6. CONCLUSION

We performed a systematic literature review of relevant articles related to data capital to understand the current research trends and research gaps. We selected published articles based on predefined inclusion and exclusion criteria. Based on the review protocol, 17 articles were discussed in our study. Among them, the most were published in 2020 (5 articles). There were five main issues discussed by the articles, i.e., ethical issues, legal issues, infrastructure issues, data capital application issues and issues of conceptualizing data capital in more depth. Ethics issues were discussed in five articles, the legal issue was discussed by one article, infrastructure issues were discussed by two articles, data capital appalication discussed by five articles and conceptualisations of data capital were raised by four articles. A total of 16 out of the 17 articles reviewed were basic research studies. Fifteen of the reviewed articles are using qulitative research approach and the other two are using Mixed-Methods. There were five articles that mentioned the research method they used, while the rest could be identified and included in the bibliographic revision group. Based on the review results from all the included articles, we revealed several positive and negative impacts of data capital that can be considered for further research.

This study provides several contributions. To the best of our knowledge, this study is the first attempt to conduct a detailed systematic literature review in the context of data capital. It provides a broad insight into data capital. It improves our understanding of how data capital is currently developing, why data capital is important, and what impact incorporating data capital has. Any issue that have been researched in the context of data capital were also successfully revealed in this study, including the type of research carried out, the approaches used, and the methods chosen. These might appeal to other researchers to extend data capital research.

The findings from our study can serve as the starting point for further elaboration in the area of data capital. Future works can be performed to overcome the limitations of this study. A broader search can be carried out on the content of the articles (not just the title and abstract). Book chapters and several gray literatures can also be considered in future reviews. This study encourages researchers and practitioners to address the emerging data capital benefits and challenges to contribute to improving the knowledge about data capital.

REFERENCES

1. Biddulph, B. Data capital: The competitive weapon most retailers have yet to fully monetise. 2020. https://retailtouchpoints.com/topics/data-analytics/data-capital-the-competitive-weapon-many-retailers-have-failed-to-fully-monetize (Accessed on 28 July 2021).
2. Sonderegger, P. The rise of data capital, 2015. https://www.forbes.com/sites/oracle/2015/02/24/the-rise-of-data-capital/ (Accessed on 25 April 2021).
3. Oracle and MIT. The rise of data capital. 2016. http://files.technologyreview.com/whitepapers/MIT_Oracle+Report-The_Rise_of_Data_Capital.pdf(Accessed on 25 April 2021).
4. Jain, S. Data is the new capital, 2020. https://www.accenture.com/_acnmedia/PDF-129/Accenture-Data-is-the-New-Capital-POV.pdf (Accessed on 27 April 2021).
5. Alfieri, F. Use creativity and data capital as a revenue source. 2015. https://datafloq.com/read/use-creativity-data-capital-as-a-revenue-source/1021. (Accessed on 25 April 2021).
6. Tan, P.N.; Steinbach, M. & Kumar, V. Introduction to data mining. Boston, USA, 2006
7. VansonBourne. Protecting data capital in your organisation. 2019. https://advance.biz-tech-insights.com/whitepaper/Protecting-Data-Capital-in-Your-Organisation.pdf (Accessed on 23 July 2021).
8. Accenture. AI for data: Data capital management @ scale with AI. 2021. https://www.accenture.com/id-en/insights/technology/ai-data (Accessed on 23 July 2021).
9. Dell Inc. Unlock your data capital. 2018. https://www.delltechnologies.com/asset/en-id/products/storage/briefs-summaries/h17138-so-unlock-data-capital.pdf (Accessed on 23 July 2021).

10. Tranfield, D.; Denyer, D. & Smart, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br. J. Manage.*, 2003, **14**(3), 207-222. doi: 10.1111/1467-8551.00375
11. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Business Res.*, 2019, **104**, 333-339. doi: 10.1016/j.jbusres.2019.07.039.
12. Kitchenham, & B. Charters, S. Guidelines for performing systematic literature reviews in software engineering. Evidence-Based Software Engineering (EBSE) Technical Report, EBSE-2007-01, UK. 2007.
13. Linnenluecke, M.K. Conducting systematic literature reviews and bibliometric analyses. *Aust. J. Manage.*, 2020, **45**(2), 175-194. doi: 10.1177/0312896219877678.
14. Kitchenham, B.; Pretorius, R.; Budgen, D.; Brereton, O.P.; Turner, M.; Niazi, M. & Linkman, S. Systematic literature reviews in software engineering: A tertiary study. *Inf. Software Technol.*, 2010, **52**, 792-805. doi: 10.1016/j.infsof.2010.03.006.
15. Zuiderwijk, A.; Chen, Y.C. & Salem, F. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Gov. Inf. Q.*, 2021, **38**, 101577. doi: 10.1016/j.giq.2021.101577.
16. Cochrane. Cochrane consumers and communication review group: Data synthesis and analysis. Cochrane consumers and communication review group. 2013. https://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/public/uploads/AnalysisRestyled.pdf (Accessed on 29 November 2021).
17. Siow, E.; Wang, X. & Tiropanis, T. Facilitating data-driven innovation using VOICE observatory infrastructure. *In* The Proceedings of the Workshop on Data-Driven Innovation on the Web. 2016, pp. 1-3. doi: 10.1145/2911187.2914584.
18. Gumbus, A. & Grodzinsky, F. Era of Big Data: Danger of descrimination. *ACM SIGCAS Comput. Soc.*, 2015, **45**(3), 118-125. doi: 10.1145/2874239.2874256.
19. Jesse, N. Internet of things and big data: The disruption of the value chain and the rise of new software ecosystems. *AI & Soc.*, 2018, **33**, 229-239. doi: 10.1007/s00146-018-0807-y.
20. Couldry, N. & Yu, J. Deconstructing datafication's brave new world. *New Media Soc.*, 2018, **20**(12), 4473-4491. doi: 10.1177/1461444818775968.
21. Sadowski, J. When data is capital: Datafication, accumulation, and extraction. *Big Data Soc.*, 2019, **6**(1), 1-12. doi: 10.1177/2053951718820549.
22. Breidbach, C.F. & Maglio, P. Accountable algorithms? The ethical implications of data-driven business models. *J. Service Manage.*, 2020, **31**(2), 163-185. doi: 10.1108/JOSM-03-2019-0073.
23. Tang, C.; Plasek, J.M.; Zhu, Y. & Huang, Y. Data sovereigns for the world economy. *Humanit. Soc. Sci. Commun.*, 2020, **7**(184). doi: 10.1057/s41599-020-00664-y.
24. Dencik, L. Mobilizing media studies in an age of datafication. *Television & New Media*, 2020, **21**(6), 568-573. doi: 10.1177/1527476420918848.
25. Prainsack, B. The value of healthcare data: To nudge, or not? *Policy Studies*, 2020, **41**(5), 547-562. doi: 10.1080/01442872.2020.1723517
26. Hagen, A.N. Datafication, literacy, and democratisation in the music industry. *Popular Music Soc.*, 2021. doi: 10.1080/03007766.2021.1989558
27. Maasø, A. & Hagen, A.N. Metrics and decision-making in music streaming. *Popular Commun.*, 2020. doi: 10.1080/15405702.2019.1701675
28. Lehtiniemi, T. Personal data spaces: An intervention in surveillance capitalism? *Surveillance Society*, 2017, **15**(5), 626-639. doi: 10.1080/15405702.2019.1701675
29. Younes, A.S. Passive violation of consumers' privacy rights on the internet in the age of emerging data capital. *J. Content, Community Commun.*, 2019, **10**, 134-150. doi: 10.31620/JCCC.12.19/14
30. Larsson, S. Algorithmic governance and the need for consumer empowerment in data-driven markets. *Internet Policy Rev.*, 2018, **7**(2), 1-13. doi: 10.14763/2018.2.791
31. Mejias, U.A. & Couldry, N. Datafication. *Internet Policy Rev.*, 2019, **8**(4), 1-10. doi: 10.14763/2019.4.1428
32. Goodridge, P.; Haskel, J. & Edquist, H. We see data everywhere except in the productivity statistics. *Rev. Income Wealth*, 2021, **0**(0), 1-33. doi: 10.1111/roiw.12542
33. Lun, K.C. The datafication of everything – even toilets. Yearbook of Medical Informatics, 2018, **27**(1). doi: 10.1055/s-0038-1641199
34. Vogt, W.P. Dictionary of statistics and methodology. London, England, 2005
35. Santos, L.M.A.L.D.; da Costa, M.B.; Kothe, J.V.; Benitez, G.B.; Schaefer, J.L.; Baierle, I.C. & Nara, E.O.B. Industry 4.0 collaborative networks for industrial performance. *J. Manuf. Technol. Manage.*, 2020, **32**(2), 1245-1265. doi: 10.1108/JMTM-04-2020-0156.
36. Havle, C.A. & Üçler, Ç. Enablers for industry 4.0. *In* The Proceedings of the 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2018. doi: 10.1109/ISMSIT.2018.8567293.
37. Bengio, Y. Learning deep architectures for AI. *Found. Trends Mach. Learn.*, 2009, **2**(1), 1–127. doi: 10.1561/2200000006.
38. Deng, L. & Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Process.*, 2014, **7**(3-4), 1–199. doi: 10.1561/2000000039.
39. Creswell, J.W. & Creswell, J.D. Research design

qualitative, quantitative, and mixed methods approaches, fifth ed. Los Angeles, USA, 2018.
40. Recker, J. Scientific research in information systems. Springer, Berlin, Germany, 2013.
41. Fernández-Rovira, C.; Valdés, J.A.; Molleví, G. & Nicolas-Sans, R. The digital transformation of business. Towards the datafication of the relationship with customers. *Technol. Forecast. Soc. Change*, 2021, **162**, 120339. doi: 10.1016/j.techfore.2020.120339.
42. PAHO, 2015. How to start writing a scientific article. https://iris.paho.org/bitstream/handle/10665.2/35007/Howtowritesciarticle2015_eng.pdf?sequence=1&isAllowed=y (Accessed 26 July 2021).
43. Cenfetelli, R.T. Inhibitors and enablers as dual factor concepts in technology usage. *J. Assoc. Inf. Syst.*, 2014, **5**(11). doi: 10.17705/1jais.00059.
44. Bhattacherjee, A. & Hikmet, N. Enablers and inhibitors of healthcare information technology adoption: Toward

CONTRIBUTOR

Dr Arief Ramadhan is working as Senior Lecturer and Researcher in the Doctor of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia. His areas of interest include Information Technology, Library Information Systems, e-Business, e-Government, e-Learning and Data Analytics.

Annexure I

Issues related to data capital addressed in the reviewed articles

Main issue	Detailed issue	Reference No
Ethics	• Discrimination • Profiling based on race, age, and sexual orientation • Less-educated populations can be targeted by unethical companies to mislead them with scams of harmful offers. • Privacy issues (re-identification)	18
	• Service providers often conceal the real purpose underlying their value propositions and coerce customers into accepting them. • Impossibility of preserving the anonymity of individuals over time • Intellectual property rights breaches	22
	• Surveillance and infringements on individual privacy • Politicisation of datafication	24
	• Businesses are heavily reliant on consumer data • Some data capture and analysis of data are done with users' active consent while others have been doing so very sneakily • Consumers are compelled to release their personal information knowingly or unknowingly • The knowledge of consumer preference and behavior might be effective enough to yield higher value • The global data protection laws that are ostensibly present to protect our privacy, but do not guarantee our online anonymity	29
	• Personal data can be used as currency, with lack of consumer protection • Consumers are threatened with profiling and misuse of data • The need for consumer empowerment in terms of transparency and ill-functioning notions of consent or user agreements • The lack of transparency and the lack of consumers being informed over the data handling • Lack of transparency in data-driven markets. Consumers do not know where the data is being sold or where it is transferred • Consumers find it difficult to assess the bargain between data sharing and service access	30
Legal	• Very little regulation of the the rights (or wrongs) of collecting data from individuals • Existing laws only focus on data use and not on data collection. • Naturalisation of personal data collection • Risks of the costs of continuous surveillance	20
Infrastructure	• Data scattered in various places can be searched and accessed easily from one platform; then, analysis of the data can be carried out directly on that platform without having to move the data from its original place.	17
	• A company can utilise the Internet of Things (IoT) to extract sense from the data • Integration using the IoT has a vertical and horizontal dimension.	19

Main issue	Detailed issue	Reference No
Data capital application	• Healthcare data are becoming increasingly valuable assets • Medical device companies can extract value from patient data to predict the behaviors and needs of other patients • Reveal the important role of policymakers to regulate healthcare data • Reviewing healthcare data for nudging. The dangers of nudging may outweigh the benefits • Patient data should be used to create better institutions, and not to tackle "behavior" at the individual level	25
	• Commercial benefits in the music industry of data-usage • The data about music audience can be used to produce and promote music more flexible • Digital data has a huge part in shaping music-industry practices especially in artist-audience relationship • The winners of digital music industry competition are data-literate actors, and the rulers are service and data providers	26
	• Music industries rely on a huge volume of data to decide about what to promote, and how to promote it • The decision made by algorithms about what to be visible or invisible is the particular interest in music metrics	27
	• Healthcare institutions collect and leverage their data to gain business competitive advantage. It is also used to improve their operational efficiency and workflow and provide a safer and better quality of patient care • We need good quality data and not just big data, especially in healthcare • Everyone within the organisation has to be mindful of the data importance • The need for good data governance within the organisation	33
	• Personal Data Space allows users a new role in value creation • Users must be viewed not only as objects of data extraction but also suppliers of data • It is required a privacy-conscientiousness and features that allow users to limit the ways and purposes of data are used • By the permission form user, monetisation of data from PDS can be done indirectly • PDS reduces the possibility of data monopoly by data collectors. businesses would lack monopolistic control of data and analytics • Data stored in a PDS is a personal resource. It should be controllable for subjective and private benefit by the individual	28
Data capital concept enrichment	• Data collection concept versus datafication concept • Data as a commodity and data as capital • An unending accumulation of data as capital, which can be represented by the M-C-M'-C-M''-C-M'''-... formula • Data capital is the same as financial capital, which should be allowed to be managed across countries	21
	• Distinguish between the terms of data sovereignty and data ownership • Put forward the terms of data force and data producer	23
	• Review in detail about the definition and characteristics of datafication • Datafication is mostly done for economic value • Corporations are the main actors of datafication • Government also has an interest in datafication • Datafication consist of the human life transformation into data through the processes of quantification, and the generation of different kinds of value from data.	31
	• System of National Accounts (SNA) has recognised databases as one of the productive capital assets • Authors extend the definition of the asset boundary. They incorporate capital formation activity in data transformation and data analytics where both processes create produced (information and knowledge) assets • Proposed the data value chain framework that consists of Data-Building/Transformation Stage, Knowledge Creation stage, and Production/Operations stage • Of EU-28 countries, authors find that 57% of employees engaged in an expanded definition of (software and) data capital formation is already accounted • Activity in data capital formation outside the national accounts asset boundary is growing faster than national accounts measures in a number of countries	32