Towards Ethical Data Ecosystems:
A Literature Study

Minna M. Rantanen
Department of Management and Entrepreneurship
University of Turku
Turku, Finland
minna.m.rantanen@utu.fi

Sami Hyrynsalmi
Computing Sciences
Tampere University
Pori, Finland
sami.hyrynsalmi@tuni.fi

Sonja M. Hyrynsalmi
Department of Future Technologies
University of Turku
Turku, Finland
smnyla@utu.fi

Abstract—While the importance of data is growing as the fuel of the new data economy, also the role of the data ecosystems is growing. The new data ecosystems enables the use, reuse and enrichment of big data sets by or together with third parties. However, in the context of technology management, the governance of these kinds of data ecosystems raises ethical questions and issues that should be acknowledged by researchers and practitioners. This study reviews the extant literature regarding the given advice about ethical considerations. The method of systematic literature study is used to collect the primary articles (N=20). The selected articles are analyzed and themed according to reoccurring themes: privacy, accountability, ownership, accessibility, and motivation. The results show the discussion is fragmented and concrete ethical guidelines are lacking. Thus, this study requires more work for governing data ecosystems in an ethical way.

Index Terms—ecosystem governance, data economy, ethical governance, data ecosystem, ecosystem ethics, systematic literature study

I. INTRODUCTION

One of the most quoted present-day idiom in data economy states that data has become the fuel—the new oil—of the modern knowledge economy. With the development of new data analyzing methods, tools and processes, new businesses and industrial domains are expected to born as well as the old ones are expected to be renewed. I.e., data is important to all aspects of the economy.

Furthermore, the modern networked business enables a widespread use and reuse of data. In addition, by combining different data sources, one can enrich data in order to create and capture value from the new information. For instance, an electricity company could enrich its data of households' electricity consumption with general building information such as the building year, type of dwelling and the size. Based on those information, data outliers can be identified and further services, such as a new thermal insulation, could be offered.

Thus, the *data ecosystems* are emerging as new interesting options for all kinds of companies. A data ecosystem is a set of actors working together in data and other shared resources [1]. In practice, variety of data ecosystems are have been formed for example by industries, governments and public-private partnerships.

Working together does not refer to only organizations, but also to working with data subjects. In past years awareness about data rights has risen due to for example General Data Protection Regulation in the European Union [2]. Alas, various data scandals have had their impact in the fears of individuals about sharing data. Thus, respecting the rights of data subjects and creation of trust are enablers of data ecosystems.

From the technology management's point-of-view, the question of data ecosystem governance becomes crucial. The data ecosystem could potentially provide sustainable business and competitive advantage. However, ecosystems are not viable without ethical considerations [3], and in the case of data ecosystems, all governance models and actions should take ethical issues into account—as the emergence of new big data opportunities have changed even the research ethics guidelines [4]. The starting point for this research is that an ethical data ecosystem requires an ethically justified governance model to support its upkeeping.

Therefore, this study focuses to survey the current status of data ecosystem governance models' ethical discussion. The aim of this work is to set building blocks for further work defining ethically justified governance models for data ecosystems. That is, we focus on the following question:

RQ: What are the current ethical aspects considered when discussing governance of data ecosystems in academic literature?

To map the extant literature, we use systematic literature study method, as guided by Kitchenham and Charters [5], to collect existing primary studies on the phenomenon. For this review, we excluded the consideration of sharing and distributing research data for other researchers as our focus is on the systems used by customers, both consumers and companies.

This study contributes to emerging discussion on ecosystem ethics [3] as well as on an emerging area of data ecosystem and its governance [1]. While the objective of this review is to give guidelines for the development of data ecosystem governance models and emphasize the ethical questions of data management.

The rest of this study is structured as follows. Section II

reviews the central concepts and related work for this study. Section III presents the research process used and Section IV gives the results. Discussion is presented in Section V with ideas for future work and, finally, Section VI closes the study.

II. Background

During the last decade, *ecosystems* emerged as a concept to characterize complex socio-technical systems [6]. Inspired by a complex biological system contain symbiotic relationships between different kinds of species [7], Moore [8] defined a 'business ecosystem' as an economic community co-evolving around an innovation. Since Moore's influential work, a series of different kinds of ecosystems have sprang into extant literature, such as 'software ecosystem' [9], 'mobile application ecosystem' [10] and 'innovation ecosystem' [11].

Thus, it is not surprise also the emerging complex network of interrelated data producers and consumers have been characterized with the concept of 'data ecosystem' [12]–[14]. According to the definition of a data ecosystem, by Olveira and Lóscio [1], it is:

> "...a set of networks composed by autonomous actors that directly or indirectly consume, produce or provide data and other related resources (e.g., software, services and infrastructure). Each actor performs one or more roles and is connected to other actors through relationships, in such a way that actors collaboration and competition promotes data ecosystem self-regulation."

The study by Olveira and Lóscio [1] furthermore emphasizes that the terminology in the field is not mature and different concepts are used interchangeably. On one hand, e.g. 'open data' refers to data that is free to use, reuse and distribute, yet also it has some of data ecosystem characteristics [15]. The concept of 'big data' means, in its simplest definition, a set of data that is too large or complex for traditional software to handle. In a more broad view, the concept is used also to refer all modern data science activities. [16], [17]

On the other hand, the concept such as 'data economy' refers generally to new economy based on data, technologies, tools and products [16]; thus, covering also the data ecosystem aspects. However, as it was stated, these concepts are used often to refer to same phenomenon, even though from different angles.

Nevertheless, as these data ecosystems are artificial constructions, they can be controlled to some extent. Baars and Jansen [18] defines *ecosystem governance* as *"[p]rocedures and processes by which a company controls, changes or maintains its current and future position in [an ecosystem] on all different scope levels"*. In more general way, ecosystem governance can be defined as a use of different leverages to ensure health, well-being and functionality of an ecosystem.

In the case of the data ecosystems, a keystone organization—the one being responsible for keeping the ecosystem functional [19]—or organizations are able to adjust and renew the ecosystem by defining new procedures, rules and processes. For example, a keystone organization might prevent of using data for certain cases for keeping the whole ecosystem healthy.

Nevertheless, our focus turns on ethical approaches for data ecosystem governance. While the literature on data ecosystems and ecosystem governance are still emerging areas, data ecosystems have some special characteristics that differentiate them from other kinds of artificial ecosystem constructs. One of the key difference is that the use, reuse and collection of data might often, but not always, consider information that is either personal or valuable for a company or an individual person.

However, to the best of authors knowledge, the previous work in this area is limited and often focused on a single domain, such as health data ecosystem, for example. In the following sections, we will review what is the current knowledge in the extant literature and discuss about different research areas.

III. Research process

Literature studies have been widely used to summarize, synthesize and characterize the existing evidence presented in the literature on a selected theme. As decisions, which studies to include and which not, is crucial for the results, systematic literature studies have been promoted as a method to objectively, systematically and in repeatable way to collect the primary studies [5].

There are two different high-level categories for the systematic literature studies [20]: systematic literature reviews are used to collect all primary evidence on a selected phenomenon for a meta-analysis. Systematic mapping studies, sometimes also referred as scoping studies, are instead used to characterize the state of the extant research in a selected theme.

This study's focus is to map the existing work on ethical governance of data ecosystems. Thus, we follow more general rules of systematic mapping studies and aim to include all relevant primary studies for the review.

We follow the systematic literature review guidelines defined by Kitchenham and Charters [5]. While their guidelines are specifically defined to support evidence-based software engineering, their method is well-applicable also for other domains of engineering sciences. The research process, used in this study, contains the following steps:

1. **Prestudy.** Firstly, we studied whether there are enough material for a literature study. In the prestudy phase, we found a dozen primary studies addressing the phenomenon at hand. Thus, there is enough primary data available for a systematic review. In addition, we noted that the primary studies have been published in various publication forums; thus, the manual search process (i.e., manually going through the selected publication forums) might not be suitable for this kind of a study. Therefore, we decided to use electronic search.

2. **Search term formation.** Secondly, we form the search term to be used in the electronic searches. Based on experimenting as well as the prestudy, we ended up using the following search term:
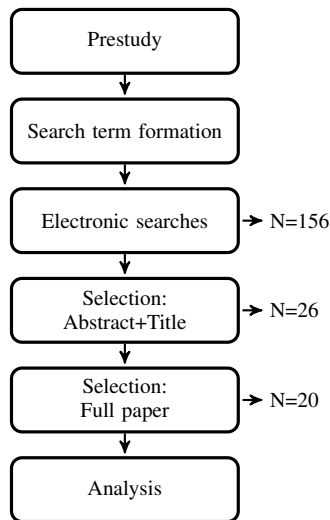
Fig. 1. The research process and numbers of articles in different steps.

Fig. 2. The yearly publications of selected papers

```
(governance* OR government) AND
(ethic*) AND ("data ecosystem" OR
"data economy" OR "big data" OR
"open data")
```

The final search term consists of three parts. In the first part, we included the concepts of 'governance' and 'government' in order to capture the discussion of handling, managing and governing data ecosystems. While the terms refer clearly different conceptualizations, they are often used interchangeably. In the second part, we include the ethics for capturing the ethical aspect of the governance. The final part of the search term defines the different synonyms for data ecosystems. The terms 'data economy', 'big data' and 'open data' are often used to discuss about the same phenomenon although there are different emphasizes on all concepts.

3. **Electronic searches.** Thirdly, we use Elsevier's Scopus publication database for this systematic literature study. Scopus was selected as it indexes most of the engineering and social science publication fora; therefore, a majority of relevant publications should be indexed by the selected search engine. We focus the searches on the abstract, title and keywords years. Based on the testing with full content, too many false positive results were returned. The search was done in the beginning of February 2019.

4. **Selection.** Fourthly, we select primary studies for this analysis by using the defined inclusion and exclusion criteria. The inclusion criteria used are the following: (i) Studies addressing governance of large data system (being referred either as 'big data' or 'open data' etc.); (ii) Studies addressing ethical aspects of governance; (iii) Peer-reviewed articles; and (iv) Articles written in English.

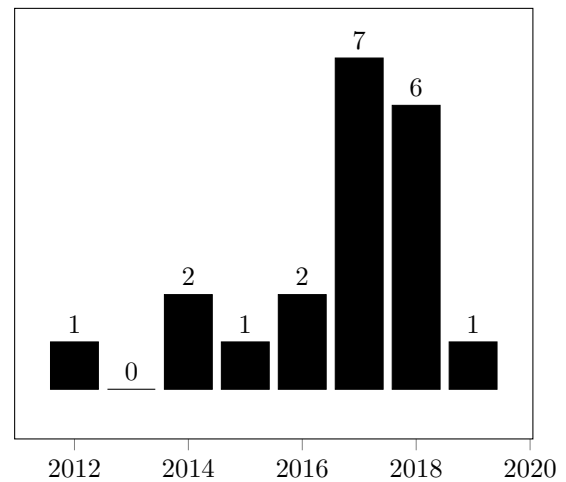Similarly, we used the following exclusion criteria: (i) Commentaries, book reviews, prefaces, summaries of a panel discussion, presentation notes, etc. non-peer reviewed content; and (ii) Studies discussing on sharing research data (as a data ecosystem) to other researchers. The selection is done in two phases: 1) First, we evaluated the papers based on their titles and abstracts. If a paper cannot be judged to be included or excluded based on these attributes, it was included for the next step. 2) Second, the full study was evaluated. In both evaluation parts, the same aforementioned inclusion and exclusion criteria were used.

Two authors of this study went through the results of electronic searches and independently applied the inclusion and exclusion criteria. Their results were compared and a common agreement was sought for. The full paper review was performed by a single researcher while another researcher verified and agreed with the results.

5. **Analysis.** Fifthly, the primary articles are studied. We use thematic analysis [21]. Articles are read through, annotated and repeating thematic areas synthesized. The reporting is based on the synthesized results.

The full research process is illustrated in Figure 1. In addition, the studies included in the different steps are given in the figure.

IV. RESULTS

A. Descriptive statistics

The search from the Scopus database returned 156 primary studies. After the application of the inclusion and exclusion criteria, the final dataset contained 20 studies. The selected papers are reported in Table I. A large number of studies were dropped during the selection step as they focused on guiding the ethical governance of research data. While this is an important theme, it is outside of our consideration of data systems providing continuous and stable services for customers.

From the selected primary studies, 19 are journal articles and 1 conference paper. The selected papers are published between 2012 and 2019 as shown in Figure 2. The discussion

on the topic seems to grow during the last few years. It is worth noting that that not all of the publications from the year 2018 have been indexed in the beginning of 2019 and, thus, the number of publications from the year 2018 is expected to grow in future.

Only *Journal of Bioethical Inquiry* appears twice in the publication fora. Thus, the initial assumption of widespread discussion, which guided to the use electronic searches, seems to be valid. Furthermore, this hints that the discussion on the topic seems to be diverse and it has not yet narrowed towards a common target or a forum.

B. Frequent ethical themes

In the selected primary studies ethical aspects of handling, managing and governing data ecosystems were diverse. Topics of the articles can be divided to four different subject areas: 1) health or bio-medical data [22], [23], [27]–[29], [31], [34], [40], [41], 2) sharing big data [30], [32], [33], [38], 3) enabling research [26], [35]–[37], [39] and 4) utilizing open data [24], [25].

However, there were some ethical aspects that were discussed in almost all of the articles. Privacy was discussed in all but one article, access to data in 18 articles, accountabilities or responsibility in 17 articles and data ownership in 12 articles. These themes resemble the four ethical issues of the information age presented by Mason [42]: privacy, accuracy, property and accessibility (PAPA).

But also other ethical themes were repeatedly present. A vast amount of the articles pondered the ethical stance from perspectives of different actors of data ecosystem mainly considering the relationship of data subject and other data users. These recurring themes resemble motivation category of ethical issues presented by Conger et al. [43], which is expansion of Mason's [42] original PAPA.

Other recurring ethical themes in the articles were: trust, transparency and security, which were mentioned in 15 articles each. In general it seems that trust is seen as enabler of data economy whereas transparency is seen as a mean to create trust. Security was often mentioned but was not the main theme in any of the articles. The relationships between articles and selected themes, discussed in depth in the articles, are shown in Table II. Aspects only mentioned, but not deeply analyzed in the articles, are omitted from the table for simplicity.

In this analysis of the primary studies about the governance of data ecosystems we utilize categorizations of Mason [42] and Conger et al. [43]. We analyze how authors have acknowledged ethical issues in five subject areas that were distinguished from primary studies: privacy, accountability, ownership, accessibility and motivation. It must be noted that boundaries of these categories are indistinct and there are some issues that overlap. For example informed consent can be equally seen as issue of ownership and privacy, but it also is an issue that relates to motivation.

1) Privacy: Information privacy is generally understood as a right to be free from secret scrutiny. Assuring privacy means that one should have a right to determine whether, when, how and to whom one's personal information is to be revealed [42], [44]. Thus, privacy is inherently linked to accessibility and ownership of data.

Lately privacy issues related to information technology have had a lot of interest from policymakers. For example, the European Union is enforcing one's right to privacy through General Data Protection Regulation act, which has affected the ways that personal data can be collected and used. Currently the European Union is also a proposal for ePrivacy regulation act that aims to regulate for example use of user information in online marketing. [2], [45]

Growing interest towards privacy and privacy policies was also visible in the primary studies about data ecosystem governance. Privacy is mentioned in almost all of the articles as something that must be ensured in governance. Especially in articles discussing health and bio medical data privacy and its safeguards are strongly emphasized.

Heitmueller et al. [23] see privacy and its trade-off with sharing is a challenge along side of balancing societal benefits and safeguarding proprietary rights. They argue that balancing collective interest and individual and organizational interests is a fundamental and universal challenge. They suggest that policymakers drafting policies about the use of big data in health care should carefully consider these trade-offs and how to balance these aspects.

Similarly, Heeney and Kerr [28], Vayena and Blasimme [34] Vayena et al. [40] ponder relationship of privacy and policies in health care or biomedical setting. Clearly, privacy is seen as one of the most important value of data governance in this setting. Nevertheless, it is also apparent that there is no self-evident solution of how to assure privacy with governance policies.

Mittelstadt and Floridi [27] discuss ethical implications of big data in biomedical context in form of literature review. They recognize privacy as one of the key areas of concerns in big data practices that is often linked to anonymisation and confidentiality. Their review shows that privacy is a complex issue that allows multitude of research aspects and in practice requires a lot of considerations.

In more general literature about big data and its use complexity of privacy issues is also visible. Kato, Takasaki, and Muramaki [30] propose a new privacy protection scheme for the data subject on the international cooperation information sharing platform. Their governance framework takes into account privacy and its self-regulation mainly from legal perspective in international information exchange between multiple parties. On the other hand, Saxena [32] focuses on privacy concerns created by big data in electronical government, thus taking more societal view on privacy and privacy risks.

In the context of research and data science Goroff, Polonetsky and Tene [35] call for a clear legal framework and ethical guidelines for use of administrative data in research. They recognize trade-off between privacy concerns and research potential, and highlight the need for actors of data ecosystem

	Authors	Title	Year	
1	Murtagh et al.	Navigating the perfect [data] storm	2012	[22]
2	Heitmueller et al.	Developing public policy to advance the use of big data in health care	2014	[23]
3	Sayogo et al.	Going beyond open data: Challenges and motivations for smart disclosure in ethical consumption	2014	[24]
4	Sieber & Johnson	Civic open data at a crossroads: Dominant models and current challenges	2015	[25]
5	Leonelli	Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems	2016	[26]
6	Mittelstadt & Floridi	The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts	2016	[27]
7	Heeney & Kerr	Balancing the local and the universal in maintaining ethical access to a genomics biobank	2017	[28]
8	Holm & Ploug	Big Data and Health Research—The Governance Challenges in a Mixed Data Economy	2017	[29]
9	Kato, Takasaki & Muramaki	Proposal of a New Privacy Protection Scheme for the Data Subject on the International Cooperation Information Sharing Platform	2017	[30]
10	Mählmann et al.	Big Data for Public Health Policy-Making: Policy Empowerment	2017	[31]
11	Saxena	Privacy concerns in integrating big data in "e-Oman"	2017	[32]
11	Smith	Towards an ethical application of customer feedback data	2017	[33]
13	Vayena & Blasimme	Biomedical Big Data: New Models of Control Over Access, Use and Governance	2017	[34]
14	Goroff, Polonetsky & Tene	Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data	2018	[35]
15	Jones & Ford	Population data science: advancing the safe use of population data for public benefit	2018	[36]
16	Marivate & Moorosi	Exploring Data Science for public good in South Africa: Evaluating factors that lead to success.	2018	[37]
17	McQuillan	People's Councils for Ethical Machine Learning	2018	[38]
18	Stockmann	Toward Area-Smart Data Science: Critical Questions for Working With Big Data From China	2018	[39]
19	Vayena et al.	Digital health: meeting the ethical and policy challenges	2018	[40]
20	Sánchez & Sarría-Santamera	Unlocking data: Where is the key?	2019	[41]

cooperation in reaching balance between them. Similarly, Jones and Ford [36] put an emphasis on privacy-by-design in population data science.

In general, privacy is seen as complex challenge for data ecosystem governance. It needs to be considered carefully when drafting policies, but also when deciding upon functionalities of data ecosystems. Though primary studies acknowledge this complexity and importance, practical guidelines are lacking.

2) Accountability: Ethical issues in relation to accountability are often seen as questions of responsibility in case of harmful events. Mason [42] originally used term accuracy, that he described with as responsibility over authenticity, fidelity and accuracy of information. Conger et al. [43] separate group responsibility and personal responsibility to differentiate between shared and individual's accountability.

Responsibility issues have become more complex as the information system ecosystems have become more interconnected and barriers between individual systems more obscure. Thus, issues about accountability are relevant also in case of data ecosystems and their management. That is, for example transparency can be included in this category.

Accountability or responsibility were mentioned in almost all articles, but only five articles discussed them in depth. In articles focusing on health and biomedical data Vayena et al. [40] focus on identification of responsible parties, whereas Sánchez and Sarría-Santamera [41] discuss roles and responsibilities as vital parts of governance. Both link responsibility to transparency and trust. They suggest that transparency in governance is a mechanism for creating trust and fostering cooperation. Also in articles handling open data [24], [25], accountability often appears parallel with transparency and trust.

In data research only Leonelli [26] focused on responsibilities and accountability. She discusses participative, reflexive management of data practices, which she sees as a way to intertwine technical solutions with ethical considerations. Her model is based on ongoing training on (research) ethics and participative ethical assessment in which individuals involved in data processing exchange ideas about potential ethical implications. Thus, Leonelli [26] highlights the importance of ethics and accountability, but also complexity and situationality of them.

Primary studies see accountability and transparency as means to achieve trust that is fundamental to data ecosystems viability. However, the relationship of these ethical themes remains only intuitively justified and too few practical advice are given on how to incorporate these themes on governance of data ecosystems.

3) Ownership: Ownership or property issues in relation to information and data ethics involve questions such who owns the data and what are fair prices for its exchange. Besides data ownership issues, ownership issues include to owning of hardware and transmission channels and how access should be allocated. [42]

Increasing value of data and information with lowering cost of data acquisition should have increased the interest about ownership issues in data ecosystem governance. However, in primary studies ownership issues were mentioned in 12 articles and discussed in depth in only one of them.

Mittelstadt and Floridi [27] distinguish ownership as one of the key areas of big data and also note that these issues will be under further scrutiny in the future. Despite this, other primary articles did not handle ownership issues profoundly. Also issues of intellectual property were not discussed in the primary articles.

TABLE II
FREQUENTLY APPEARING ETHICAL THEMES DISCUSSED IN DEPTH AND THEIR RELATIONS TO THE SELECTED PAPERS ARE ILLUSTRATED. THE SELECTED PAPERS ARE CATEGORIZED BY THEIR MAIN RESEARCH DOMAIN INTO FOUR GROUP: HEALTH OR BIO-MEDICAL DATA, SHARING BIG DATA, UTILIZING OPEN DATA AND ENABLING RESEARCH. THE MAIN THEMES ARE BOLDED.

	Study	**Privacy**	**Accountability**	**Ownership**	**Access**	**Motivation**	Consent	Security	Trust	Transpar.
Health	1	[22]				X					
	2	[23]	X			X					
	6	[27]	X		X			X			
	7	[28]	X			X		X			
	8	[29]						X			
	10	[31]				X					
	13	[34]	X	X			X			X	
	19	[40]									
	20	[41]		X			X			X	
Big data	9	[30]	X					X			
	11	[32]	X								
	17	[38]					X				
	11	[33]					X			X	
Open	3	[24]		X		X					X
	4	[25]		X		X	X				X
Science	5	[26]		X						X	
	14	[35]	X			X	X				
	15	[36]	X			X	X				
	16	[37]				X					
	18	[39]				X					
	Sum		9	5	1	10	7	5	0	4	2

However, 17 articles mention informed consent and in five of them it is one of the main themes. Informed consent relates to ownership, since it can be understood as individuals right to control data use. In the articles handling health and biomedical data informed consent was considered as one of the major ethical themes of big data [27], and as practical challenge for governance models and regulations [28], [29], [34].

It must be acknowledged, that informed consent is "a standard tool" in this domain as Heeny and Kerr [28] note. Explicit informed consent is also the only reasonable solutions from legalistic perspective, if one wishes to share information between different parties [30].

Thus, research on ownership from ethical perspective is desperately needed in the relation data ecosystems and their governance. It must be remembered that ethical issues of ownership include more topics than just permission to handle somebody's personal data.

4) Accessibility: Accessibility can be generally understood as a question about what information does a person or an organization have a right or privilege to obtain, under which conditions, and with what safeguards. [42]. Thus access is an issue that strongly relates to property and ownership. Mason [42] relates access also to literacy, since it is crucial to gaining information. This view emphasizes access as an ability, not as a right [43].

Unethical access can be seen as unauthorized, unnecessary use of the access to read, copy or otherwise making use of databases, programs or other computer resources [43]. This links access issues also to security, although it is often presented as its own domain or category.

Accessibility or access to information was often mentioned in the selected papers. Only two articles did not mention issues of access or its restrictions. Ten articles focused on the topic as one of their main themes.

In articles about health and biomedical data ecosystems access and security are reoccurring theme. Heitmueller et al. [23] represent taxonomy of data types and levels of access. Although, their article focuses on health care they also consider open data commons and other forms of data than health information.

Murtagh et al. [22] consider a ideal access in emerging data economy. They argue for ubiquitous, properly designed, and widely recognized and used IDs that would facilitate global accessibility. They see proper identification as security measure, but also as a catalyst for change in the ways that we use, share and analyze data.

Heeney and Kerr [28] recognize accessibility as important ethical factor in biobank governance. They argue for the need of more universal maintaining of access, in order to achieve more sustainable data ecosystem around genomic data.

Also Mählmann, Reumann and Evangelatos [31] argue that the full potential of big data can only be realized if data are being made accessible and shared, but not without basing data governance on moral codes. They see big data as an essential for public health policy-making, although they admit that respecting rights and liberties of individuals form a challenges that must be considered carefully.

Naturally, open data articles address access issues in depth. Sayogo et al. [24] see greater access to data as a challenge, that should be solved to achieve greater sustainability and transparency. Similarly, Sieber and Johnson [25] see open data

as means to increase participation in government functioning and decision-support.

Similar arguments for more open access to gain greater good are at the core of majority of the science related papers [35]–[37], [39], since without access to data benefits of research are not achieved. However, also in these articles importance of careful ethical considerations is acknowledged.

These reflections about access to data highlight the value of data and data ecosystems, but also ignore the aspect that access is an ability not a right. Perhaps focusing on the accessibly as an ability would be more fruitful when considering governance of data ecosystems, since it would further separate security and access as ethical issues. However, this does not mean that security controls should be forgotten.

5) Motivation: Conger et al. [43] recognized that motivation is an important factor when researching ethical issues. They state that motivation extents Mason's [42] categories to include beneficiaries of unethical acts, thus emphasizing discussion about stakeholders. They use concept motivation instead of stakeholders to highlight the behavioural aspects of computer ethics, and thus emphasize questions of beneficiaries, justice and personal motivation as ethical factors.

As apparent based on previous themes all articles at least mentioned some motivational aspects of data ecosystems. However, beneficiaries of unethical acts were not discussed.

Some authors did, however, consider the multitude of stakeholders or actors of data ecosystem and their conflicting motivations in relation to data economy and its governance. Discussions about actors and motivations were distributed through all four subject areas.

In relation to health and biomedical data ecosystems Vayena et al. [40] contemplate on conditions of fair innovation in digital health, big data ecosystem and its actors ranging citizens and patients to policy makers and researchers.

Sánchez and Sarría-Santamera [41] discuss motivations of different actors in data ecosystem along side of governance requirements. What separates their paper from the mass is that they also briefly discuss profit-sharing systems. Financial benefits or business models seem otherwise to be an untouched topic in ethical governance of data ecosystems.

In articles that focus value of research it seems that balancing motivations of the researchers, public, organizations and individuals is a common theme. Goroff et al. [35] discuss competing motivations and restrictions, Jones and Ford [36] ponder on core concepts and major challenges for data-intensive research from ethical, legal and societal perspectives. Also in context of open data Sieber and Johnson [25] consider in depth conflicting motivations.

Similarly, Smith [33] focuses on tension between customer motivations and organizational motivations in relation to use of customer feedback data. Competing motivations are also in core of the article of McQuillan [38] who considers use of people's councils to balance and to contest judgments done with machine learning technologies.

As data ecosystems by definition are socio-technical systems, these kinds of vast ethical considerations about actors

motivations are appreciated. However, there clearly is a need for more meticulous ethical analysis of beneficiaries and possible casualties of data ecosystems and their governance. After all an ecosystem should be beneficial to all actors so that it can stay viable.

V. DISCUSSION

A. Key findings

This study surveyed the extant literature in order to identify the ethical aspects considered in the discussion of data ecosystems governance and management. Based on the analysis of ethical themes found in the selected academic literature about data ecosystem governance, we summarize our observations in the following five points.

Firstly, there is a relatively small number of reoccurring ethical themes. In total nine different themes stand out. Furthermore, these themes can all be summarized under five themes of privacy, accountability, ownership, accessibility and motivation. Overall, this indicates that the studied articles, in one hand, agree well with the critical themes and attributes of data ecosystem governance. On the other hand, it shows that the work in the field has been fragmented in content-wise. Only a few of the studies addressed all of the themes; most of the primary studies focused only in a few themes and bypassed the big picture.

Secondly, more work about ethical governance of data ecosystem has been made about health and bio-medical data than any other subject area of primary studies. Since ethical issues of handling large data masses—that also contain sensitive data, but which can also be further utilized—are fundamentally connected to health care and its governance, this is not surprise. This does, however, remind that when considering ethical governance of data ecosystems the cumulative body of knowledge in health related fields should be remembered.

Furthermore, it is worth to be aware that the data ecosystems are going to grow also in the other areas than in the health care—current research is lacking interests towards these areas. In addition, it is worthy to note that our results show also that the work in the ethical data ecosystem governance is fragmented in domain-wise. There is lack of studies addressing general guidelines and governance models for data ecosystems.

Thirdly, although all five ethical themes or even all nine themes can be found in many articles, profound analysis is fragmented. It is understandable that focuses vary, but in order to draft ethical guidelines to data ecosystem all of these ethical themes should be considered.

Fourthly, although all of the primary studies focus on ethical aspects of data ecosystem governance it is remarkable how superficial ethical analyses are in general and how authors have managed to avoid use of ethical theories. These reoccurring themes have been contemplated on the field of ethics as well as many of its branches that apply ethical theories technological settings. Thus, referring to actual ethics or even applied ethics would be highly recommended and beneficial when discussing ethical aspects of technology.

Fifthly, there is a considerable gap between ethical considerations and practical data ecosystem governance. Although ethical issues are often mentioned, there is limited amount of guidelines of how implement them into governance practices. To govern data ecosystems in an ethically justified ways we need to consider practical ways to control, change and maintain these ecosystems.

All in all, it is clear that ethical issues of data ecosystem governance are fundamental to viability of these systems which could renew the modern knowledge economy and create new kind of business. However, more profound and multidisciplinary analysis is needed in order to fully understand the ethical basis of data ecosystem and how to implement it in governance practices.

B. Future work

While our survey shows that lots have been done for creating and guiding ethical data ecosystems, our survey also emphasizes that there are still lots of avenues needing further work. Based on our review in the field, we are proposing the following avenues for further inquiries.

First, what seems to be distinctive to the studies literature is the lack of comprehensive guidelines to the practitioners. Whereas the literature is already rich in nuances, future work should aim to generate more practical guidelines for the ethically justified data ecosystem governance. Also comparisons to existing practices in the field should be studied from ethics perspective. The identified major and minor themes works well as a starting point for these kinds of inquiries.

Second, a majority of the literature studied focuses on certain well-limited domains such as health and bio-medical data ecosystems. There is a remarkable shortage of work addressing more general type data ecosystems. Thus, in future work, more emphasize should be given to general type models as most of the domain, nevertheless, share similarities. However, to find out how different domain-specific data ecosystems differ, is a topic for further work.

Third, as mentioned earlier, more thorough and reflective research on ethical value basis of data ecosystems and their governance is needed. This calls for multidisciplinary cooperation in order to aggregate existing information that could be applied in emerging domain of data ecosystems and their governance.

C. Limitations

There are some limitations that should be noted. First, as always with systematic literature studies, the selection of the search term has remarkable effects on the results. While we have done our best to tackle pitfalls in the formulation of the search term (e.g., the decision to include several different alternatives for data ecosystems), the study is still limited by the selected search terms. For example, our searches could not find papers using different synonyms for the studied concepts. Thus, this limitation is something to bear in mind.

Second, the selection of the publications is also important for systematic literature studies. We decided to use electronic search and Scopus publication database. While the decision is justified with a large number of indexed fora by Scopus, this still limits the results of this mapping into those included into Scopus. In future works, also other publication databases should be considered to be included.

VI. Conclusions

This study reviewed existing academic knowledge about ethical governance of data ecosystems. A data ecosystem refers to a network, which consume, produce or provide data and other resources. Data ecosystem governance, however, refers to the usage of different leverages in order to guarantee the functionality, sustainability and well-being of the ecosystem.

We used the systematic mapping study to collect the primary studies discussing the phenomenon. The final set of studies included 20 articles published during 2012–2019. The results show that there is active discussion on themes of privacy, accountability, ownership, accessibility, and motivation. However, the survey also showed that the discussion is fragmented both in content as well as in domains. Health and bio-medical data are over-represented while other domains are infrequently discussed. Also content-wise, the discussion is fragmented and a comprehensive view is missing.

Nevertheless, this study contributes to the emerging areas of ecosystem ethics and data ecosystems as well as works as a starting point for future development towards ethical data ecosystem governance models. The identified ethical themes should be taken into account in current and future work considering data ecosystems.

References

[1] M. I. S. Oliveira and B. F. Lóscio, "What is a data ecosystem?" in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, ser. dg.o '18. New York, NY, USA: ACM, 2018, pp. 74:1–74:9.

[2] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L119, pp. 1–88, May 2016.

[3] J. Koskinen, M. M. Rantanen, K. K. Kimppa, and S. Hyrynsalmi, "Ecosystem ethics: An ethical analysis of orchestrators' ultimate power and the dilemma of ecosystem ruling," in *Proceedings of the 9th International Workshop on Software Ecosystems*, ser. CEUR Workshop Proceedings, S. Hyrynsalmi, A. Suominen, C. Jud, and J. Bosch, Eds., vol. 2053. Aachen, Germany: CEUR-WS, Nov. 2017, pp. 43–54.

[4] M. Zimmer, "Addressing conceptual gaps in big data research ethics: An application of contextual integrity," *Social Media + Society*, vol. 4, no. 2, p. 2056305118768300, 2018.

[5] B. A. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering. version 2.3." Keele University, Keele, Staffs, United Kingdom, EBSE Technical Report EBSE-2007-01, July 2007.

[6] M. Seppänen, S. Hyrynsalmi, K. Manikas, and A. Suominen, "Yet another ecosystem literature review: 10 + 1 research communities," in *2017 IEEE European Technology and Engineering Management Summit (E-TEMS)*, ser. E-TEMS 2017. IEEE, Oct. 2017, pp. 1–8.

[7] J. F. Moore, *The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems*. New York, NY, USA: Harper Business, 1996.

[8] ——, "Predators and prey: A new ecology of competition," *Harvard Business Review*, vol. 71, no. 3, pp. 75–86, May-June 1993.

[9] J. Bosch, "From software product lines to software ecosystems," in *Proceedings of the 13th International Software Product Line Conference*, ser. SPLC '09. Pittsburgh, PA, USA: Carnegie Mellon University, 2009, pp. 111–119.

[10] S. Hyrynsalmi, M. Seppänen, and A. Suominen, "Sources of value in application ecosystems," *The Journal of Systems and Software*, vol. 96, pp. 61–72, October 2014.

[11] D.-S. Oh, F. Phillips, S. Park, and E. Lee, "Innovation ecosystems: A critical examination," *Technovation*, vol. 54, pp. 1–6, August 2016.

[12] L. Reggi and S. Dawes, "Open government data ecosystems: Linking transparency for innovation with transparency for participation and accountability," in *Electronic Government*, H. J. Scholl, O. Glassey, M. Janssen, B. Klievink, I. Lindgren, P. Parycek, E. Tambouris, M. A. Wimmer, T. Janowski, and D. Sá Soares, Eds. Cham: Springer International Publishing, 2016, pp. 74–86.

[13] Y. Demchenko, C. de Laat, and P. Membrey, "Defining architecture components of the big data ecosystem," in *2014 International Conference on Collaboration Technologies and Systems (CTS)*, May 2014, pp. 104–112.

[14] P. E. Bourne, J. R. Lorsch, and E. D. Green, "Perspective: Sustaining the big-data ecosystem," *Nature*, vol. 527, no. 7576, pp. S16–S17, Nov. 2015.

[15] D. Lee, "Building an open data ecosystem: An irish experience," in *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, ser. ICEGOV '14. New York, NY, USA: ACM, 2014, pp. 351–360.

[16] L. Cao, "Data science: A comprehensive overview," *ACM Comput. Surv.*, vol. 50, no. 3, pp. 43:1–43:42, Jun. 2017.

[17] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution that Will Transform how We Live, Work, and Think*, ser. An Eamon Dolan book. Houghton Mifflin Harcourt, 2013.

[18] A. Baars and S. Jansen, "A framework for software ecosystem governance," in *Software Business*, ser. Lecture Notes in Business Information Processing, M. A. Cusumano, B. Iyer, and N. Venkatraman, Eds., vol. 114. Berlin, Germany: Springer Berlin Heidelberg, 2012, pp. 168–180.

[19] M. Iansiti and R. Levien, "Strategy as ecology," *Harvard Business Review*, vol. 82, no. 3, pp. 68–78, March 2004.

[20] B. Napoleão, K. R. Felizardo, É. F. de Souza, and N. L. Vijaykumar, "Practical similarities and differences between systematic literature reviews and systematic mappings: a tertiary study," in *Proceedings of the 29th International Conference on Software Engineering & Knowledge Engineering*, 2017, pp. 85–90.

[21] G. Guest, K. M. MacQueen, and E. E. Namey, *Applied thematic analysis*. Thousand Oaks, CA: SAGE, 2012.

[22] M. Murtagh, G. Thorisson, S. Wallace, J. Kaye, I. Demir, I. Fortier, J. Harris, D. Cox, M. Deschênes, P. Laflamme, V. Ferretti, N. Sheehan, T. Hudson, A. C. Thomsen, R. Stolk, B. Knoppers, A. Brookes, and P. Burton, "Navigating the perfect [data] storm," *Norsk Epidemiologi*, vol. 21, no. 2, Apr. 2012.

[23] A. Heitmueller, S. Henderson, W. Warburton, A. Elmagarmid, A. S. Pentland, and A. Darzi, "Developing public policy to advance the use of big data in health care," *Health Affairs*, vol. 33, no. 9, pp. 1523–1530, 2014.

[24] D. S. Sayogo, J. Zhang, T. A. Pardo, G. K. Tayi, J. Hrdinová, D. F. Andersen, and L. F. Luna-Reyes, "Going beyond open data: Challenges and motivations for smart disclosure in ethical consumption," *JTAER*, vol. 9, pp. 1–16, 2014.

[25] R. E. Sieber and P. A. Johnson, "Civic open data at a crossroads: Dominant models and current challenges," *Government Information Quarterly*, vol. 32, no. 3, pp. 308 – 315, 2015.

[26] S. Leonelli, "Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2083, p. 20160122, 2016.

[27] B. D. Mittelstadt and L. Floridi, "The ethics of big data: Current and foreseeable issues in biomedical contexts," *Science and Engineering Ethics*, vol. 22, no. 2, pp. 303–341, Apr 2016.

[28] C. Heeney and S. M. Kerr, "Balancing the local and the universal in maintaining ethical access to a genomics biobank," *BMC Medical Ethics*, vol. 18, no. 1, p. 80, Dec 2017.

[29] S. Holm and T. Ploug, "Big data and health research—the governance challenges in a mixed data economy," *Journal of Bioethical Inquiry*, vol. 14, no. 4, pp. 515–525, Dec 2017.

[30] N. Kato, H. Takasaki, and Y. Murakami, "Proposal of a new privacy protection scheme for the data subject on the international cooperation information sharing platform," in *eHealth 360°*, K. Giokas, L. Bokor, and F. Hopfgartner, Eds. Cham: Springer International Publishing, 2017, pp. 23–28.

[31] L. Mählmann, M. Reumann, N. Evangelatos, and A. Brand, "Big data for public health policy-making: Policy empowerment," *Public Health Genomics*, vol. 1, pp. 3212–320, 2017.

[32] S. Saxena, "Privacy concerns in integrating big data in "e-oman"," *Journal of Information, Communication and Ethics in Society*, vol. 15, no. 4, pp. 385–396, 2017.

[33] R. Smith, "Towards an ethical application of customer feedback data," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec 2017, pp. 3955–3957.

[34] E. Vayena and A. Blasimme, "Biomedical big data: New models of control over access, use and governance," *Journal of Bioethical Inquiry*, vol. 14, no. 4, pp. 501–513, Dec 2017.

[35] D. Goroff, J. Polonetsky, and O. Tene, "Privacy protective research: Facilitating ethically responsible access to administrative data," *The ANNALS of the American Academy of Political and Social Science*, vol. 675, no. 1, pp. 46–66, 2018.

[36] J. K. Helen and F. D. Vincent, "Population data science: advancing the safe use of population data for public benefit," *Epidemiol Health*, vol. 40, no. 0, p. e2 018 061–0, 2018.

[37] V. Marivate and N. Moorosi, "Exploring data science for public good in south africa: Evaluating factors that lead to success," in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, ser. dg.o '18. New York, NY, USA: ACM, 2018, pp. 6:1–6:6.

[38] D. McQuillan, "People's councils for ethical machine learning," *Social Media + Society*, vol. 4, no. 2, p. 2056305118768303, 2018.

[39] D. Stockmann, "Toward area-smart data science: Critical questions for working with big data from china," *Policy & Internet*, vol. 10, no. 4, pp. 393–414, 2018.

[40] E. Vayena, T. Haeusermann, A. Adjekum, and B. Alessandro, "Digital health: meeting the ethical and policy challenges," *Swiss medical weekly*, vol. 148, p. w14571, 2018.

[41] M. C. Sánchez and A. Sarría-Santamera, "Unlocking data: Where is the key?" *Bioethics*, pp. 1–10, 2019.

[42] R. O. Mason, "Four ethical issues of the information age," *MIS Quarterly*, vol. 10, no. 1, pp. 5–12, 1986.

[43] S. Conger, K. D. Loch, and B. L. Helft, "Ethics and information technology use: a factor analysis of attitudes to computer use," *Information Systems Journal*, vol. 5, no. 3, pp. 161–183, 7 1995.

[44] H. J. Smith, T. Dinev, and H. Xu, "Information privacy research: An interdisciplinary review," *MIS Q.*, vol. 35, no. 4, pp. 989–1016, Dec. 2011.

[45] "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning the respect for private life and the protection of personal data in electronic communications and repealing directive 2002/58/ec (regulation on privacy and electronic communications)," 2017. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017PC0010