

Received March 7, 2021, accepted March 27, 2021, date of publication March 31, 2021, date of current version April 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070017

# Structural Characteristics of Stakeholder Relationships and Value Chain Network in Data Exchange Ecosystem

TERUAKI HAYASHI<sup>1</sup>, GENSEI ISHIMURA<sup>2</sup>, AND YUKIO OHSAWA<sup>1</sup>, (Member, IEEE)

<sup>1</sup>Department of Engineering, The University of Tokyo, Tokyo 113-8656, Japan

<sup>2</sup>Department of Information and Management for Innovation, i University, Tokyo 131-0044, Japan

Corresponding author: Teruaki Hayashi (hayashi@sys.t.u-tokyo.ac.jp)

This study was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (JP19H05577 and JP20H02384), and the Artificial Intelligence Research Promotion Foundation.

**ABSTRACT** In recent years, data have been treated as economic goods, and have begun circulating into new forms of the ecosystem through data exchange. The increasing expectation that data exchange will be used to create new businesses and economic value requires a strong understanding of the overall structure and the characteristics of the data exchange ecosystem. However, unlike other well-known business ecosystems, such as the financial markets or service ecosystems, the data exchange ecosystem is immature and the observable interactions in the ecosystem are quite limited. In this study, we proposed a novel framework for describing the relationships and their interactions as the stakeholder-centric value chain (SVC) by focusing on the interactions among stakeholders, i.e., the business players who have an interest in the data business—as the core components of the ecosystem. We examined 45 businesses using SVC from 120 data business participants and elucidated some of the business structures with data exchange using network analysis, which was the first comprehensive empirical study on the data-mediated business relationships among stakeholders in the data exchange ecosystem. We found that the integrated network of data businesses consists of many densely connected clusters with numerous low-frequency stakeholders and shows the disassortative and hierarchical hub-and-spoke structure. The results suggest that a few stakeholders monopolize links with many others and that a segregation of stakeholders appears in the ecosystem. Our approach and the results provide the important insights for all stakeholders in the data exchange ecosystem and those who consider entering the market.

**INDEX TERMS** Data ecosystem, data exchange, network analysis, stakeholder, value-chain.

## I. INTRODUCTION

Treated as economic goods, data have been circulated into new forms of economic activity and have begun to be exchanged and traded in the market in recent years [1]–[4]. Interdisciplinary business collaborations in the data exchange ecosystem have been appearing globally, and data transactions among businesses have attracted significant research attention [5], [6]. Moreover, the expectation that personal data can be valuable has increased [7]–[9], and several businesses are entering the data exchange market ecosystem.

The advent of the data exchange ecosystem has been discussed at various institutions and in the private sector [10], [11]. Therefore, there is a strong need for the method for

intuitive and easy understanding of overall structure and characteristics of the data exchange ecosystem. However, unlike other well-known business ecosystems such as the financial markets and service ecosystems, the data exchange ecosystem is still developing. More drastic changes have occurred, such as the introduction of laws and regulations represented by the General Data Protection Regulation in the European Union or the New York City Automated Decisions Systems Law, compared to those in other ecosystems [12], [13]. If personal data are leaked, it is difficult to determine who is responsible, what caused the leak, or the location of the bottleneck in the business. Deposition of data and control rights, as well as transfer to third parties, increase the complexity of stakeholders related to the business and make it difficult to control the value chains. The value chain, originated by Porter [14], is the set of activities and interactions

The associate editor coordinating the review of this manuscript and approving it for publication was Mehdi Hosseinzadeh<sup>1</sup>.

among stakeholders in the business model. With an overall understanding of the value chain in the ecosystem, it is possible to discuss the marketability of data in the data exchange. However, since the environment surrounding the data exchange is changing rapidly, observable interactions are quite limited. Moreover, without a unified scheme and sufficient dataset, it is not yet possible to understand the structure of the whole ecosystem.

To establish an appropriate unit of analysis and framework and create a comprehensive understanding of the data exchange ecosystem, we assume the stakeholder relationships and their interactions in the business models as the core components. In this study, we consider a stakeholder to be a player (an individual or a group) who has an interest in the data business and acts for the purpose of realizing some values through involving the business. For example, the stakeholder who generates the personal data is the user, and those who store these data are the data accumulators. The ecosystem is formed by the activities of people, and the relationships consist of the exchange of resources, such as money, services, or data. Therefore, stakeholder-centric business understanding and analysis are essential approaches which will allow us to assess the soundness of the emerging ecosystem of data exchange. Although the data exchange ecosystem is a kind of service ecosystem, the traditional methods for service ecosystem focus on the system performance or governance and lack the detailed value interactions between stakeholders [15], [16].

In this study, we proposed a framework for describing the stakeholder-centric value chain (SVC) and elucidated some of the business structures with data exchange through network analysis. Relationship networks, such as scientific collaboration, actor, and mobile phone calls, are empirically known to show power distributions, high assortativity, and nontrivial hierarchical modularity [17]. So, what about the stakeholder network of data businesses? If they display a power distribution, what functions and roles do hub stakeholders, who have many connections with other stakeholders, play in the ecosystem? In addition, there seem to be few interactions between data providing companies, meaning that there is a possibility that stakeholders are segregated by businesses. For example, it may be rare for mobile phone companies to exchange their customer information with each other. The structure and interaction among stakeholders by network analysis will be an important insight for those who engage in data businesses.

In summary, this is the first empirical study on the data-mediated business relationships among stakeholders, and the main contributions of this study are as follows:

1. There is no common framework for describing business models that allows the data exchange ecosystem to be understood in a unified manner, and no datasets that express the relationships among the entities in data businesses, such as stakeholders, data, or services. To solve this problem, we created the SVC framework

with the stakeholders as the nodes and the value of data and services as edges.

2. Our study included 120 participants who had interest in data businesses, and collected 45 business models by SVC through discussions with them. We found that the integrated network of data businesses consists of many densely connected clusters with numerous low-frequency stakeholders and that it shows the disassortative and hierarchical hub-and-spoke structure. The results suggest that a few stakeholders monopolize links with many others and a segregation of stakeholders appears in the ecosystem.

The remainder of this paper is organized as follows. Section 2 summarizes the relevant previous studies. Section 3 describes our approach and the experiment, and Section 4 discusses the results to improve understanding of the data businesses in the data exchange ecosystem and topics for future work. Finally, Section 5 provides concluding remarks.

## II. RELEVANT STUDIES

As represented by digital transformation, digitization and data collaboration are expected to become prevalent in society [18]. Unlike the conventional supply chain, the supply chain relationships in electronic marketplaces have been decentralized, with the roles and values shared equitably among stakeholders [19]. For example, in the business model for e-books, the computer industry, home appliances, publishers, and telecommunication companies dynamically work together to form a complex ecosystem [20]. E-commerce ecosystem and mobile markets are composed of heterogeneous networks consisting of multiple layers [16], [21]. A data exchange ecosystem is a data-specific and -mediated form of service ecosystem, which are self-organized in the long-term competition and cooperation among the business players (stakeholders). Since data exchange businesses constitute an emerging ecosystem, the roles of stakeholders in this ecosystem, including data marketplaces, are not yet fully understood. Although several studies have been conducted on data exchange and related marketplaces [2], [3], [22], there has been little systematic research focused on understanding the ecosystem.

Several researchers have attempted to tackle the challenge of understanding the characteristics of the data exchange ecosystem. For instance, Stahl *et al.* proposed six classification frameworks for electronic (online) data marketplaces, including the supplier, buyer, and platform(er) [23]. Deloitte LLP described the relationships among stakeholders in an open-data marketplace with such roles as data enablers, suppliers, and individuals [24]. Further, Quix *et al.* developed a business architecture and an exchange process for managing industrial data on a data exchange platform, which is a form of data exchange ecosystem, with 11 types of players, including the data owner, consumer, and broker service [25]. The Industrial Value Chain Initiative (IVI) offers tools that

support interdisciplinary data collaborations using the software IVI

Modeler,<sup>1</sup> which has 16 types of charts to describe business models. Lammi and Pantzar focused on citizen consumers and discussed how their previous roles changed into roles as data citizens from a sociological standpoint with the advent of the data economy [26]. Cao et al. [27] defined three players, i.e., data owners, collectors, and users, while Sooksatra et al. [28] assumed two players, i.e., users and service providers, in the data markets and proposed a method of coordinating the trading. The Data Trading Alliance proposed a data trading model that involved three stakeholders: the data provider, the data user, and the data trading market service provider [29]. There has also been research on legal protection between consumers and big data brokers [30], [31]. For leading data collaborations, the Innovators' Marketplace on Data Jackets provides a framework for discussion among stakeholders, specifically, data holders, users, and analysts [32].

Although many studies have been conducted to clarify and systematize the data exchange ecosystem, there are two limitations. First, finding a method to defining the roles of stakeholders in an actual data exchange based on the roles defined in previous studies. For example, in the analysis of a shared economy, consumers consume goods and also serve as producers [33]. In other words, consumers have multiple roles rather than a single role, and it depends on the relationships among stakeholders. Additionally, as demonstrated by previous studies, there has been no common definition of roles. The data exchange ecosystem involves heterogeneous stakeholders with different interests and relationships with the data businesses. Therefore, it is insufficient to determine the fixed roles of stakeholders in advance to obtain an overview of the emerging data ecosystem.

The second limitation is the lack of perspective on compensation for data and services. Numerous studies have been conducted on evaluating business transactions and network analysis [34], [35], assuming that the essence of economic activity is the exchange of money and goods between stakeholders. However, regarding the data ecosystem and its marketplaces, there has been little discussion of the position or consideration of data in the businesses. There have been assessments of pricing models [36]–[38], trading model [28], [29], and digital rights and privacy [39]–[42] in data exchange, but the purpose of data exchange and acquisition is the development of data-related services. Data trading is just one of the functions in the ecosystem, and without overall understanding of the value chain in the ecosystem, it is impossible to discuss the marketability of data.

For these reasons, we did not specify a rule describing the stakeholders in this study. Instead, we allowed them to be described in free format using the SVC framework, in the sense that the roles arise in relationships between stakeholders. Also, we targeted the interactions among stakeholders as

the value chain in the data businesses, rather than relationships based solely on specific functions such as data trading.

### III. OUR DATA EXCHANGE ECOSYSTEM APPROACH AND EXPERIMENT

#### A. SVC OF DATA

The SVC framework is an approximate unit of analysis used to clarify business models in the data exchange ecosystem focusing on the stakeholders and their relationships. In applications of the SVC framework, such as knowledge representation, the framework is based on a graphical representation that uses nodes, edges, and labels.

TABLE 1. Descriptive items of the SVC framework.





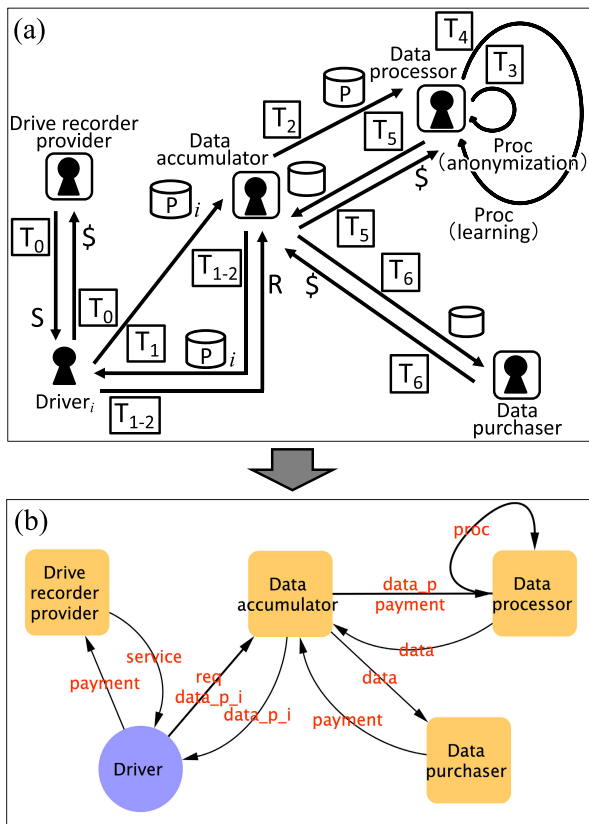
Element	Label	Icon	Description
Node	Individual		Label used when the stakeholder is an individual.
	Company		Label used when the stakeholder is a company or institution.
Edge	Request	R	Label indicating a request for data.
	Service	S	Label indicating the service offered including the product.
	Payment	\$	Label indicating the exchange of money (payment).
	Data		Label representing the data provided. "P" represents personal data, and the subscript <i>i</i> identifies an individual, if needed.
	Process	Proc()	Label representing data processing. The name of the algorithm can be described in parentheses.
	Timestep	$T_j$	The time order is represented by the subscript <i>j</i> . If a branch occurs, hyphens or underscores can be used.
Other	Comment		Additional information about nodes and edges.

Table 1 lists the elements and labels of the SVC framework. Each node represents a stakeholder in the data business. In most cases there are multiple users in the business models. In the SVC, to easily understand the types of stakeholders that interact in the data business, multiple users are represented by one node named "users." Each node can have a name, such as "drive data recorder provider," "driver," or "data accumulator." In addition, nodes have two types of labels, i.e., individual and company. Entities such as data and services exchanged among stakeholders are defined as labels at the edges of the directed graph. An edge has six types of labels, and the data they represent are further divided into three types, i.e., a collection of non-personal data, collection of personal data, and each individual's personal data. Many methods are available for data provision, such as downloading data stored on a website or obtaining data through application

<sup>1</sup><https://iv-i.org/wp/en/>

programming interfaces such that their details are expressed in the relevant comments.

There are various types of data processing, such as data cleansing, which in turn includes anonymization, visualization, and artificial intelligence techniques such as machine learning. The details of the data processing are provided in parentheses. Note that the data processing is described in a self-loop in the SVC. To consider the time-series information of data businesses, timesteps are attached to the edges as attributes, and if the nodes and edges require additional explanation, comments can be added to them. To describe data businesses, users can employ graphical icons (pictograms) to share and understand the structure of the business models easily.



**FIGURE 1.** Example of collecting data businesses and creating the stakeholder-centric network. (a) shows the graphical icons (pictograms) to share and understand the structure of the business models easily, and (b) is the description by the directed multi attributes graph model.

Figure 1(a) shows an example of the driving record data business model described using SVC. Previously, the drive recorder’s business model included the drive recorder provider selling products to the drivers and receiving compensation from the drivers. The data exchange ecosystem creates a different situation. At the first timestep  $T_0$ , the drive recorder provider sells the drive recorder to the driver as before (S and \$). The driving video data stored in the drive recorder is transferred to the data accumulator that operates the cloud service ( $T_{1-2}$ ). The data is treated as personal data because there are pedestrian faces included in the driver’s

information and video. For example, when a driver needs a driving record in case of an accident, the driver requests the data from the data accumulator (R) and receives his/her own data (the branch of timestep occurs as  $T_{1-2}$ ). In  $T_2$ , the data accumulator sends the personal data of the drivers to the data processor. The data processor anonymizes and processes the video at  $T_3$  and  $T_4$  (Proc) and receives the payment (\$) from the data accumulator in exchange for the data cleansed in  $T_5$ . At  $T_6$ , the data accumulator sells data to a third-party data purchaser and receives compensation (\$). Depending on the business model, the drive recorder provider may also serve as a data accumulator or data processor, but in this business model, each is divided into three stakeholders.

Note that the purpose of this study was to understand the business structures in the data exchange ecosystem. The main advantage of using the SVC framework is that it allows a diversity of stakeholders but restricts the relationships to those that are essential, which simplifies the analysis of the stakeholder relationships. Our targets were the statistical relationships among stakeholders in the businesses, rather than dynamic changes of the stakeholders or relationships due to those creative processes [43], [44]. Since these processes increase the complexity of stakeholder relationships, we did not model detailed and frequent interactions such as the contracts between data providers and receivers [25], [45].

### B. NETWORK MODEL AND DATA COLLECTION USING SVC

A combination of the relationships between stakeholders, which are the smallest units of knowledge, enables description of the stakeholder-centric network of data businesses. In addition, assuming that each data business is the sample, the structure of the data exchange ecosystem can be elucidated by integrating these data businesses using a network-based approach. To encode the data businesses described by the SVC through these networks efficiently, the data business network  $G$  was represented as  $(V, E, A, L)$ , which is a directed multi-attribute graph.  $G$  consists of nodes  $v(a) \in V$ , a set of stakeholders with attribute  $a$ , edges  $e_{ij}(l) \in E$ , and a set of relationships between the  $j$ -th and  $i$ -th nodes with the relationship label  $l$ . In this framework, the values of the attributes are not numerical, but rather are given as sets of relationship labels, where  $A$  represents two attribute values of a node and  $L$  is a set of relationship labels of the edges (Table 1). Each node  $v(a)$  has the name of a stakeholder and one attribute (individual/company), whereas each edge can have an unlimited number of attribute values depending on the number of its relationships.

Figure 1(b) shows the directed multi-attribute graph of Figure 1(a). In the network model diagram, companies and individuals are represented by square and circular nodes, respectively. The labels are listed by the edges, and the thickness represents the number of labels. For simplicity, the timesteps is not shown on the edges in Figure 1(b).

To collect the information on data businesses, we involved 120 participants consisted of business people and engineering students over 20 years old who were engaged or interested

in data businesses. Business people belonged to 18 different companies in Japan and were engaged in the actual data businesses, and the students were from the graduated school of engineering who have been studying the data businesses for over a year. We allowed them to form groups of two or three and briefed them on the SVC framework for 30 min. We then asked each group to discuss the outline of their company's or well-known data businesses, and described the ones that were agreed in the groups using the SVC framework in 30 min., which yielded 45 data business diagrams in total. After collecting the manually described diagrams of data businesses from the participants, we digitalized the diagrams into the network model and stored them in JSON format.

There were two parts to the analysis. First, to clarify the characteristics of the stakeholders and their relationships in the data business, we calculated their frequencies and compared their interactions by attribute types; to elucidate the structural characteristics of the data businesses and their patterns, we compared their distributions and the network motifs. Second, to reveal the structural characteristics of a collection of data businesses as an ecosystem, we integrated the data businesses by applying the network indices. Due to the complicated business model, the timestep information was missed for some businesses collected in the experiment and not present in our analysis. In this study, we used the information on nodes with attributes (individual/company) and names of stakeholders, and edges with six labels: request, service, payment, data (personal/non-personal data), and process.

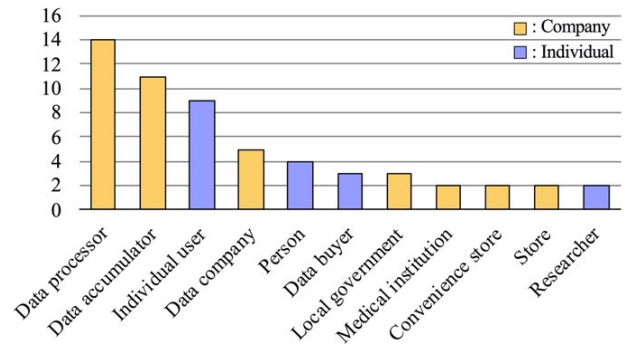
**TABLE 2. Characteristics values of the data businesses.**

	Values
Total number of stakeholders	214
Number of stakeholder types	155
Average/median numbers of stakeholders in each data business	4.76/4.00
Maximum/minimum number of stakeholders in each data business	10/2
Total number of relationships	465
Number of relationship types	6
Average/median numbers of relationships in each data business	10.33/10.00
Maximum/minimum numbers of relationships in each data business	24/1

**IV. RESULT AND DISCUSSION**

**A. STAKEHOLDERS AND DATA BUSINESSES**

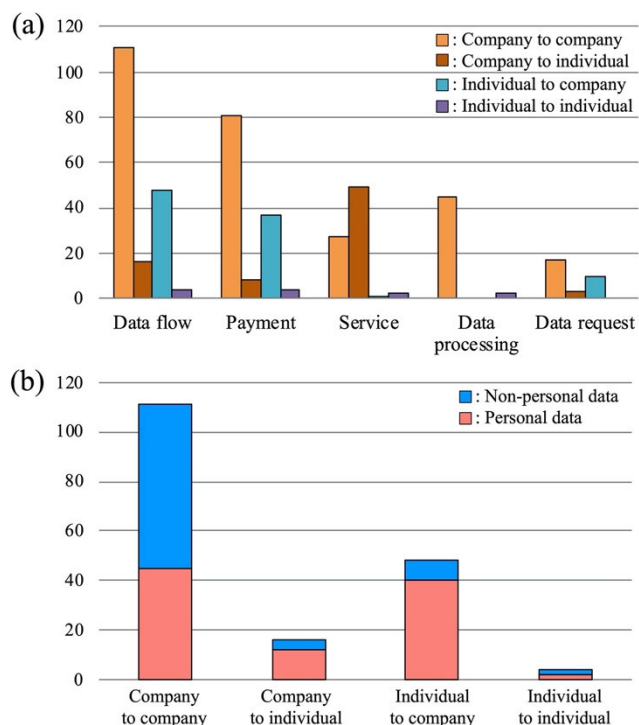
Table 2 lists the statistical information of the data businesses, and Figure 2 shows the frequencies of the 11 most



**FIGURE 2. Frequencies of 11 most frequent stakeholders among data businesses.**

frequent stakeholders. In this study, the number of relationship types was limited to six, but there were no restrictions on the number of stakeholder types. Consequently, the total number of emerging stakeholders was 214, and there were 155 types. The stakeholder types are a unique number of stakeholder names listed in the data businesses. For example, if “local government” appears in multiple data businesses, it counts as one type. The average number of stakeholders included in each data business was 4.76 and median was 4.00. In addition, the maximum number of appearances in the data business was 10, and the minimum number was two; therefore, there were not extremely many or few. According to Figure 2, data processor has the most appearances across the data businesses (14 times), followed by data accumulator (11 times) and individual user (9 times). The differences in color represent individuals and companies. We found 122 companies and 33 individuals in the data businesses, which shows that there were many interactions, especially between companies.

Regarding the relationships, in contrast, the average number of appearances in each data business was high at 10.33 compared with the number of stakeholders, reaching a maximum of 24 and minimum of 1 (Table 2). Figure 3(a) depicts five relationships of data flow (including personal/non-personal data), payment, service, data processing, and data request, which are classified into four types of interactions: company to company, company to individual, individual to company, and individual to individual. The data flow between companies is mainly due to the relationship between the data accumulator and processor, and the relationship between the stakeholders who have the roles of data seller and purchaser. The data flow from individuals to companies indicates that data generated by individual behaviors, such as purchasing behavior or movement history, are stored. Payment occurs the most between companies and from individuals to companies. Payments between companies are mainly for data sales or data processing, and payments from individuals to companies are for services such as drive recorders or healthcare applications. In contrast, data processing occurs through a self-loop from a data processor or individual data scientist. Data requests mainly occur between companies and



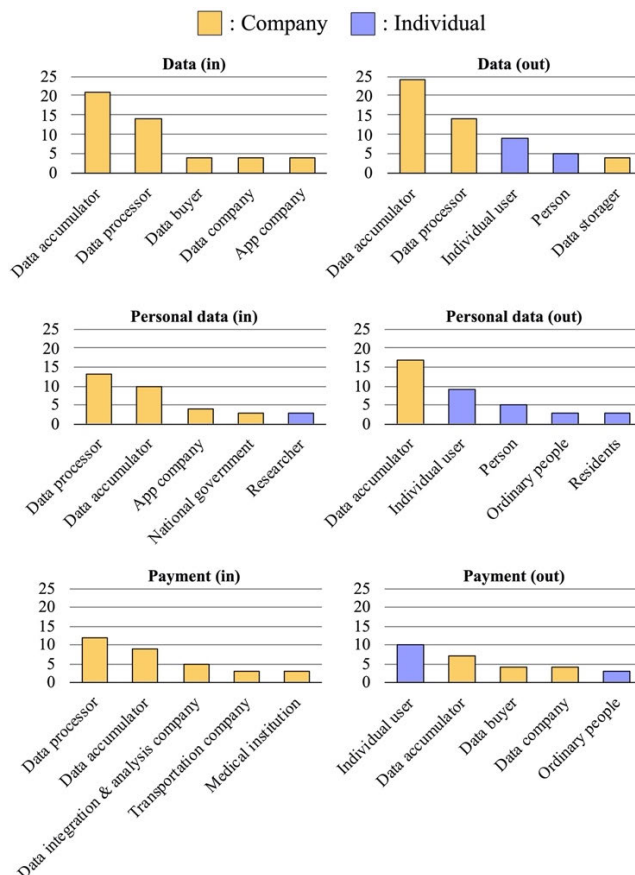
**FIGURE 3. (a) Frequencies of relationships between stakeholders. (b) Four types of data flows broken down into personal and non-personal data.**

from individuals to companies and are for the data that have been stored in the data accumulator.

Figure 3(b) illustrates the data flows, which have been divided into non-personal and personal data. There are many data exchanges between companies, which represent the interactions in which personal data are transferred to the data processor, anonymized or cleansed, and returned as non-personal data. The data flow from individuals to companies involves storing the individuals’ personal data, and the flow from companies to individuals consists of requests for their stored data. A few flows occur between individuals, stemming from interactions between individual data scientists. In addition, the data flows to third parties (data purchasers or data users) generally consist of non-personal data that have been anonymized by data processors.

Figure 4 shows the top five stakeholders who provide/receive data (personal and non-personal data), personal data, and payments. The analysis indicates the stakeholders who are at risk and who receive the most benefits across the data businesses. The top ranks are monopolized by two stakeholders: they both receive the most data (the data processor appeared 21 times, and the data accumulator, 14 times) and provide data (the data accumulator appeared 24 times and data processor 14 times).

Next, we discuss the stakeholders who handle personal data specifically. Although most stakeholders who receive personal data are companies, it is worth noting that users, ordinary people, and inhabitants occupy the top ranks among the stakeholders who provide personal data. These stakeholders produce and provide data in almost every business, and

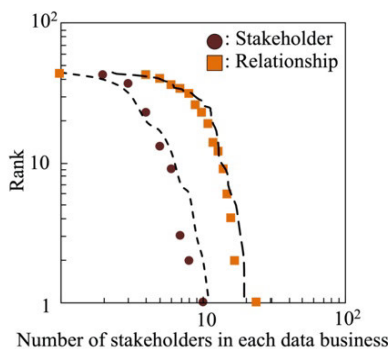


**FIGURE 4. Top five stakeholders who provide (out) and receive (in) data, personal data, and payments.**

most data businesses utilize the data generated by individual stakeholders. In contrast, the stakeholders who obtain the most personal data are the data processors and accumulators. Considering that the data processor receives most of the data from the accumulator, the data accumulator is the actor who collects the most personal data in the ecosystem. In contrast, handling personal data means not only monopolizing data, but also risking data leakage. If personal data are leaked from these stakeholders, data processors, or accumulators, the value chain of data exchange will be seriously damaged, and the entire ecosystem will suffer.

Third, we discuss the stakeholders who benefit the most in the ecosystem. The data processor and accumulator receive payments 12 and 9 times, respectively. In this case, the top five stakeholders are all companies. Since the data processor and accumulator also have potential data leakage risks, it is desirable for them to receive sufficient payment from the perspective of the ecosystem’s soundness. In contrast, the stakeholders who pay the most money are the individual users (10 times), which occurs for the services from the companies, and the data accumulator (7 times), who mainly makes payments for data processing to the data processor in exchange for anonymized data from the data accumulator.

To summarize the discussion of the data and payment flows, it can be said that the stakeholders who appear frequently and handle the data, such as the data processor and accumulator, are the hubs in the overall data businesses. There are 179 data flows in total, and 61 of the flows are around these two stakeholders. Thus, it can be said that a large portion of the data flows in the ecosystem are supported by these two stakeholders. In contrast, there are only 21 payment flows for the data processor and accumulator, whereas there are 130 payment flows in total. In this analysis, although we did not define the amount of each payment flow and could not evaluate the amount of money each stakeholder receives accurately, the payments are not concentrated on specific stakeholders, but rather occur throughout the ecosystem.



**FIGURE 5.** Distributions of the number of stakeholders and relationships in each data business. The dotted lines represent the Poisson distribution for stakeholders and Gaussian distribution for relationships given the same number of elements with the same average and standard deviation.

Figure 5 presents a log–log graph of the rank-frequency plots of the number of stakeholders and relationships in each data business. The probability of a stakeholder/relationship appearing  $m$  times is  $p(m)$ . However, the double logarithmic graph of the frequency distribution is considerably affected by noise. Therefore, we used a rank-frequency plot, which is equivalent to the complementary cumulative distribution function. Since the number of stakeholders in a data business is a non-negative integer with an unknown upper limit and the average and variance are approximately equal ( $\lambda = 4.76$ ,  $\sigma^2 = 2.82$ ), the frequency of occurrence of the stakeholders can be a Poisson distribution. In contrast, the variance of the relationship distribution is slightly larger than the average ( $\lambda = 10.33$ ,  $\sigma^2 = 20.41$ ), and the Shapiro–Wilk test [46] shows that the relationship distribution is Gaussian. In contrast, the number of stakeholders did not show a Gaussian distribution in the test. In Figure 5, the dotted lines represent the Poisson distribution of the stakeholders and Gaussian distributions of the relationships when the same number of elements, stakeholders, and relationships, with the same average and standard deviation are given, which show almost the same shapes as the distributions. In addition, the relationships between the numbers of stakeholders and relationships are almost linear, and if the number of stakeholders increases by 1, 2.17 relationships will appear on average.

The results show that the data businesses do not have extremely large numbers of stakeholders, but rather that they are close to the average value of 4.76. Furthermore, the number of relationships in each data business grows linearly rather than exponentially as the number of stakeholders increases. In other words, the stakeholders in each business do not have dense relationships with each other as in a star graph; instead, one or two hub stakeholders have many relationships with the others as in a hub-and-spoke network structure. Let us discuss these features in detail by considering the network motifs, which are the characteristic patterns that appear in networks [47]. Since the network targeted in this study is a directed graph, there are 13 types of relationships composed of three nodes, as shown in Table 3. Calculating the number of motifs across the data businesses revealed that the V-shape appears 642 times among the 679 motifs. In other words, the stakeholders in data businesses are not related mutually, but they exhibit a hub-and-spoke structure, which is a set of V-shaped structures centered around specific stakeholders that are hubs. In the stakeholder network of data businesses, the hubs are the stakeholders who have the function of data accumulation, e.g., data accumulators, medical institutions, or local governments.

**TABLE 3.** Thirteen types of network motifs of the directed graph.

Motif	Values	Motif	Values
	350		10
	128		4
	80		4
	41		3
	31		2
	14		0
	12		

**B. CHARACTERISTICS OF INTEGRATED STAKEHOLDER NETWORK**

This subsection provides an analysis of the structural characteristics of the integrated stakeholder network. The integrated network is a combination of 45 business model networks with a common stakeholder name. The data processor appeared in 14 data businesses, and the node of the data processor

TABLE 4. Characteristics values of the stakeholder network.

	Value
Number of nodes ( $ V $ )	155
Number of links ( $ E $ )	193
Average degree ( $\langle k \rangle$ )	2.49
Density ( $\rho$ )	0.0162
Average clustering coefficient ( $\langle C \rangle$ )	0.254
Assortativity ( $r$ )	-0.071
Number of components	9

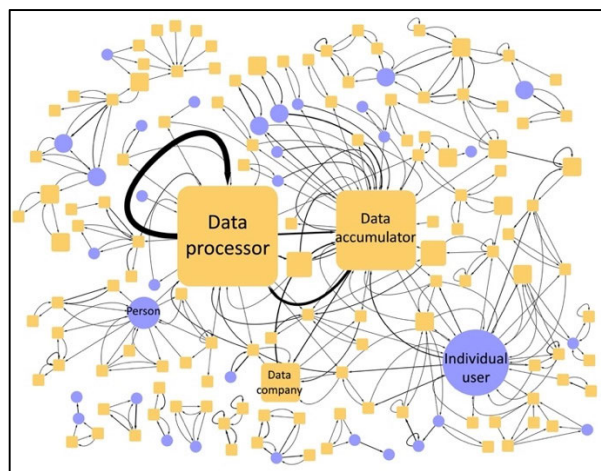


FIGURE 6. Diagram of integrated stakeholder network (the names of the five most frequent stakeholders are shown).

became the shared node in the network. The network of all the business models was divided into nine subgraphs (Figure 6), and the characteristic values are shown in Table 4. Note that self-loops were not included, and the values were calculated as an undirected graph. In the diagram, companies are represented by square nodes and individuals by circular nodes. The size of each node indicates the frequency of the relevant stakeholders. The relationship labels are embedded as the attributes of the edges, and the thickness represents the number of relationships. Figure 7 shows the degree distribution of the stakeholder network using rank-frequency plots.  $\gamma$  is a power-law index that is calculated considering the part of the distributions whose coefficient of determination  $R^2$  is  $\geq 0.97$ .

From a macroscopic perspective, the average degree ( $\langle k \rangle$ ) is low at 2.49, and  $\gamma$  is 2.49, indicating that there is a power distribution (Figure 7). The in- and out-degree also exhibit power distributions. The density  $\rho$  is very small at 0.0162, which shows that the stakeholder network is globally sparse. Furthermore, the clustering coefficient ( $\langle C \rangle$ ) is 0.286 and reflects the existence of hubs in local clusters. As illustrated in Figure 6, the network consists of many densely connected clusters with numerous low-frequency stakeholders. Thus, most stakeholders have small numbers of relationships, but

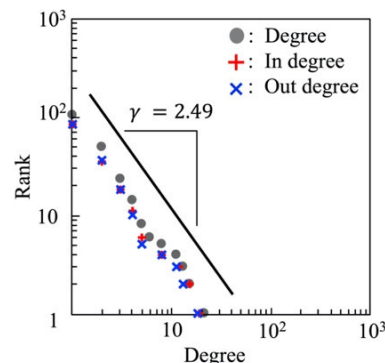


FIGURE 7. Degree distribution of integrated stakeholder network.

a few have extensive linkage with others. The fact that the degree distribution becomes a power distribution means that some stakeholders monopolize links with many other stakeholders. The stakeholders that appear frequently across the data businesses—data processors, data accumulators, and individual users—are located at central positions in the network. In other words, these hub stakeholders play central roles in the data businesses and may have more influence than others in the ecosystem.

It is also noteworthy that the assortativity of the network does not have a high positive value at  $r = -0.071$ . Assortativity indicates the degree of correlation between two neighboring nodes [48]. When  $r > 0$ , high-degree nodes tend to be linked with other high-degree nodes. In contrast, when  $r < 0$ , high-degree nodes tend to connect with low-degree nodes. Empirically, human-related networks such as networks of co-author relationships, actor relationships, or Facebook tend to have positive assortativity [17], [49]–[51]. Meanwhile, the stakeholder network of the data businesses is disassortative or neutral and has characteristics similar to those of engineering or natural networks (representing power grids or protein interactions, respectively [17]). The results suggest that even if the data business network is human-related, the nature of the relationships is different from those in other human-related networks. In the business model using electronic medical records, for example, the patients are mainly connected with medical institutions or doctors who collect data, and patients do not have any linkage with each other. Moreover, the data accumulator who functions as a data seller has a relationship with the data buyers such as the convenience store or the local government, which exhibits a hub-and-spoke structure. In the economic system, the data providers may not sell data to other data providers, and data analysts may have rivalry with one another. In other words, the business relationships among those who have the same roles in the market are unlikely to be linked, and the hub nodes in the network are connected to avoid other hubs. Therefore, the network tends to be disassortative, which has been a common observation in economic systems where trades occur between individuals or organizations with different skills and specialties [17], [52]. This feature also suggests that the network may have a hierarchical structure with roles



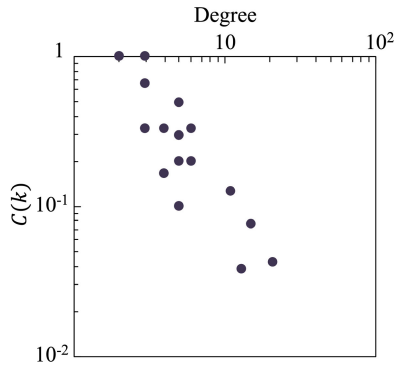


FIGURE 8. Scaling of clustering coefficient with degree.

not explicitly shown in the ecosystem, which means that the stakeholders with the same role seldom connect with each other.

To clarify the existence of the hierarchical structure in the network, we examined the relationship between the clustering coefficient and the degree of each node (Figure 8).  $C(k)$  is the clustering coefficient value of degree  $k$ , and we found that  $C(k)$  decreases with decreasing  $k$ , which shows the degree dependence of  $C(k)$ . This finding indicates that the clustering coefficient of the low-degree node is much larger than those of the hub nodes. In other words, the low-degree nodes are located in a dense local network, but the neighbors of the hub nodes are in a sparse network. This observation suggests that there is a hierarchical structure in the stakeholder network, which is composed of a combination of small modules with strongly connected local nodes. The quantitative basis for this nested hierarchical structure is  $C(k) \propto k^{-\beta}$  ( $0 < \beta < 2$ ) [53], [54], where  $\beta = 1.36$  in this study.

The disassortativity and hierarchical structure is led by a segregation of stakeholders in the ecosystem. When looking at the network, citizens, patients, and individuals who provide their personal data are not connected to each other, but rather have relationships only with data accumulators, medical institutions, and credit card companies whose function is data collection. Then, those stakeholders exchange data only with the stakeholders who have the data processing function, and the processed data flow to the data users, purchasers, and business operators who conduct their own data-driven businesses. The data processors have no relationship with the data providers or users, and the data users or purchasers have almost no direct connection with the individual data providers such as the citizens or patients. In other words, in the data business, similar stakeholders do not have a relationship. Although the email network is established by interactions among people, it displays scale-free, disassortative [17], and hierarchical characteristics [17], [55], which is similar to the stakeholder network of this study. This is because the entities (stakeholders in this study) with different business roles have closer communication (relationships such as data flows and services in this study) than the entities with the same role. It can be said that the hierarchical structure in data businesses is established by the relationships of stakeholder groups with different functions in the ecosystem.

### C. LIMITATIONS AND FUTURE WORK

In this study, we revealed some of the structural characteristics of data businesses in the data exchange ecosystem, focusing on stakeholder relationships. However, many issues need to be explored in the future. First, in the experiments, the edges had only relationship labels, and we did not define their capacities, such as payment and data capacities, or the times required for the tasks. Functioning as an ecosystem means that appropriate payments are made for the provision of data and services. Providing information on the edges will enable understanding of the detailed value chain and profit sufficiency by applying methods used to solve maximum flow problems [56].

The second issue is related to the robustness and resilience of the ecosystem. We found that the stakeholder network is disassortative and has a power distribution and hierarchical structure. Such networks are known to be robust to random node removal but vulnerable to hub removal [57], [58]. The lack of connection between hubs, as in the case of social networks [59], and the low network redundancy are due to the sparseness of the network. In other words, the hub stakeholders in the network have considerable influence on the ecosystem, and simultaneously, the stakeholder network easily disintegrates when these stakeholders are removed, as they are the Achilles heel of the ecosystem. Discussions of robustness and resilience are important for further understanding of the data exchange ecosystem.

Third, the roles and functions of the stakeholders must be determined. Since the kinds of roles possessed by the stakeholders in the data exchange ecosystem are not fully understood, we did not deliberately set roles for the stakeholders in this investigation. Through the relationships with other stakeholders, the data accumulators not only accumulate data, but also store data from various providers such as individuals and companies, and sell the data to data users and purchasers. Further, data users and purchasers utilize the data and provide services or products to other stakeholders. It is necessary to consider that one stakeholder has several roles when discussing the data businesses. The methods of block modeling and structural equivalence may be helpful in using functions to create new role classifications in the data exchange ecosystem based on the actual interaction patterns among stakeholders.

The fourth point that must be addressed is ecosystem dynamics. The data businesses dealt with in this study formed a static network, which is an approximation of a part of the ecosystem that changes from moment to moment. To understand the emerging and growing data exchange ecosystem, it is necessary to clarify the dynamics by applying simulations with a model-driven approach.

### V. CONCLUSION

Due to the rapid development of the data exchange ecosystem, elucidating the value chain of data businesses and stakeholder interactions in the ecosystem is important. Data exchange businesses constitute an emerging ecosystem, and

the analysis framework for understanding this ecosystem is lacking. Without a way to understand the business structure that is simple and intuitive, understanding the business's bottleneck in advance or when a security problem occurs is impossible. In this study, we proposed a unified description framework for understanding the flow of data values and human interactions in the data business, created a dataset for understanding the ecosystem, and elucidated a part of its relationship and interaction structure. Our approach and the results provide some important insights for all stakeholders in the data exchange ecosystem and those who consider entering the market. We believe the research presented here will facilitate the development of data businesses in the future. Additionally, in our future work, based on this empirical study, we will continuously collect datasets using SVC and develop the analysis protocol according to the limitations described in the previous section.

## ACKNOWLEDGMENT

The authors would like to thank Editage for their assistance in English language editing.

## REFERENCES

- [1] M. Balazinska, B. Howe, and D. Suciu, "Data markets in the cloud: An opportunity for the database community," *Proc. VLDB Endowment*, vol. 4, no. 12, pp. 1482–1485, Aug. 2011.
- [2] F. Schomm, F. Stahl, and G. Vossen, "Marketplaces for data: An initial survey," *ACM SIGMOD Rec.*, vol. 42, no. 1, pp. 15–26, May 2013.
- [3] F. Stahl, F. Schomm, and G. Vossen, "The data marketplace survey revisited," ERCIS Work., Münster, Germany, Tech. Rep., no. 18, 2014.
- [4] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, and W. Zhao, "A survey on big data market: Pricing, trading and protection," *IEEE Access*, vol. 6, pp. 15132–15154, 2018.
- [5] F. Stahl, F. Schomm, and G. Vossen, "Data marketplaces: An emerging species," in *Frontiers in Artificial Intelligence and Applications*, vol. 270. Amsterdam, The Netherlands: IOS Press, 2014, pp. 145–158.
- [6] M. Spiekermann, "Data marketplaces: Trends and monetisation of data goods," *Intereconomics*, vol. 54, no. 4, pp. 208–216, Jul. 2019.
- [7] C. Aperjis and B. A. Huberman, "A market for unbiased private data: Paying individuals according to their privacy attitudes," *SSRN Electron. J.*, pp. 1–17, Apr. 2012.
- [8] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," *IEEE Access*, vol. 2, pp. 1149–1176, 2014.
- [9] N. D. Line, T. Dogru, D. El-Manstrly, A. Buoye, E. Malthouse, and J. Kandampully, "Control, use and ownership of big data: A reciprocal view of customer big data value in the hospitality and tourism industry," *Tourism Manage.*, vol. 80, Oct. 2020, Art. no. 104106.
- [10] *Unlocking the Value of Personal Data*, World Econ. Forum, Cologny, Switzerland, 2013.
- [11] "Exploring the economics of personal data: A survey of methodologies for measuring monetary value," OECD Publishing, Paris, France, Tech. Rep., no. 220, 2013.
- [12] C. Rusche, "Data economy and antitrust regulation," *Intereconomics*, vol. 54, no. 2, pp. 114–119, Mar. 2019.
- [13] S. Abiteboul and J. Stoyanovich, "Transparency, fairness, data protection, neutrality: Data management challenges in the face of new regulation," *J. Data Inf. Qual.*, vol. 11, no. 3, pp. 1–9, Jul. 2019.
- [14] M. E. Porter, *Competitive Advantage: Creating and Sustaining Superior Performance*. New York, NY, USA: Simon Schuster, 2008.
- [15] A. P. Barros and M. Dumas, "The rise of Web service ecosystems," *IT Prof.*, vol. 8, no. 5, pp. 31–37, Sep. 2006.
- [16] X. Xue, Z. Chen, S. Wang, Z. Feng, Y. Duan, and Z. Zhou, "Value entropy: A systematic evaluation model of service ecosystem evolution," *IEEE Trans. Services Comput.*, early access, Aug. 14, 2020, doi: 10.1109/TSC.2020.3016660.
- [17] A. L. Barabási, *Network Science*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
- [18] I. O. Pappas, P. Mikalef, M. N. Giannakos, J. Krogstie, and G. Lekakos, "Big data and business analytics ecosystems: Paving the way towards digital transformation and sustainable societies," *Inf. Syst. E-Bus. Manage.*, vol. 16, no. 3, pp. 479–491, Aug. 2018.
- [19] T. Skjøtt-Larsen, H. Kotzab, and M. Grieger, "Electronic marketplaces and supply chain relationships," *Ind. Marketing Manage.*, vol. 32, no. 3, pp. 199–210, Apr. 2003.
- [20] Y. Yoo, O. Henfridsson, and K. Lyytinen, "Research commentary—The new organizing logic of digital innovation: An agenda for information systems research," *Inf. Syst. Res.*, vol. 21, no. 4, pp. 724–735, Dec. 2010.
- [21] K. Huang, J. Zhang, W. Tan, and Z. Feng, "Shifting to mobile: Network-based empirical study of mobile vulnerability market," *IEEE Trans. Services Comput.*, vol. 13, no. 1, pp. 144–157, Jan. 2020.
- [22] T. Hayashi and Y. Ohsawa, "Understanding the structural characteristics of data platforms using metadata and a network approach," *IEEE Access*, vol. 8, pp. 35469–35481, 2020.
- [23] F. Stahl, F. Schomm, G. Vossen, and L. Vomfell, "A classification framework for data marketplaces," *Vietnam J. Comput. Sci.*, vol. 3, no. 3, pp. 137–143, Aug. 2016.
- [24] *Open Growth: Stimulating Demand for Open Data in the U.K.*, Deloitte LLP, London, U.K., 2012.
- [25] C. Quix, A. Chakrabarti, S. Kleff, and J. Pullmann, "Business process modelling for a data exchange platform," in *Proc. 29th Int. Conf. Adv. Inf. Syst. Eng.*, 2017, pp. 153–160.
- [26] M. Lammi and M. Pantzara, "The data economy: How technological change has altered the role of the citizen-consumer," *Technol. Soc.*, vol. 59, pp. 1–8, Nov. 2019.
- [27] X. Cao, Y. Chen, and K. J. Ray Liu, "Data trading with multiple owners, collectors, and users: An iterative auction mechanism," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 2, pp. 268–281, Jun. 2017.
- [28] K. Sooksatra, W. Li, B. Mei, A. Alrawais, S. Wang, and J. Yu, "Solving data trading dilemma with asymmetric incomplete information using zero-determinant strategy," in *Proc. Int. Conf. Wireless Algorithms, Syst. Appl.*, 2018, pp. 425–437.
- [29] Data Trading Alliance. (2019). *A Private Industry Alliance Over Wide Industry Domain*. Accessed: Jun. 23, 2020. [Online]. Available: <https://www.internationaldataspaces.org/wp-content/uploads/2019/12/Day2-01-Yamahira-DataTradingAlliance.pdf>
- [30] C.-L. Yeh, "Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers," *Telecommun. Policy*, vol. 42, no. 4, pp. 282–292, May 2018.
- [31] Y. Nie and X. Han, "Research on consumers' protection in advantageous operation of big data brokers," *Cluster Comput.*, vol. 22, no. 4, pp. 8387–8400, Jul. 2019.
- [32] Y. Ohsawa, H. Kido, T. Hayashi, C. Liu, and K. Komoda, "Innovators marketplace on data jackets, for valuating, sharing, and synthesizing data," in *Knowledge-Based Information Systems in Practice*, vol. 30. New York, NY, USA: Springer-Verlag, 2015, pp. 83–97.
- [33] M. Antikainen, M. Lammi, and T. Hakanen, "Consumer service innovation in a circular economy—the customer value perspective," *J. Serviceol.*, vol. 3, no. 1, pp. 1–8, 2018.
- [34] V. Allee, "Value network analysis and value conversion of tangible and intangible assets," *J. Intellectual Capital*, vol. 9, no. 1, pp. 5–24, Jan. 2008.
- [35] T. Ohnishi, H. Takayasu, and M. Takayasu, "Hubs and authorities on Japanese inter-firm network: Characterization of nodes in very large directed networks," *Prog. Theor. Phys. Suppl.*, vol. 179, pp. 157–166, Mar. 2009.
- [36] W. Dai, C. Dai, K.-K.-R. Choo, C. Cui, D. Zou, and H. Jin, "SDTE: A secure blockchain-based data trading ecosystem," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 725–737, 2020.
- [37] Y. Shen, B. Guo, Y. Shen, X. Duan, X. Dong, and H. Zhang, "A pricing model for big personal data," *Tsinghua Sci. Technol.*, vol. 21, no. 5, pp. 482–490, Oct. 2016.
- [38] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "An online pricing mechanism for mobile crowdsensing data markets," in *Proc. 18th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2017, pp. 1–10.

- [39] S. Spiekermann, A. Acquisti, R. Böhme, and K.-L. Hui, "The challenges of personal data markets and privacy," *Electron. Markets*, vol. 25, no. 2, pp. 161–167, Jun. 2015.
- [40] C. Niu, Z. Zheng, F. Wu, X. Gao, and G. Chen, "Trading data in good faith: Integrating truthfulness and privacy preservation in data markets," in *Proc. IEEE 33rd Int. Conf. Data Eng. (ICDE)*, Apr. 2017, pp. 223–226.
- [41] S. D. C. D. Vimercati, S. Foresti, G. Livraga, and P. Samarati, "Empowering owners with control in digital data markets," in *Proc. IEEE 12th Int. Conf. Cloud Comput. (CLOUD)*, Jul. 2019, pp. 321–328.
- [42] C. Niu, Z. Zheng, F. Wu, X. Gao, and G. Chen, "Achieving data truthfulness and privacy preservation in data markets," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 1, pp. 105–119, Jan. 2019.
- [43] T. Hayashi and Y. Ohsawa, "Processing combinatorial thinking: Innovators marketplace as role-based game plus action planning," *Int. J. Knowl. Syst. Sci.*, vol. 4, no. 3, pp. 14–38, Jul. 2013.
- [44] E. Ruijter, S. Grimmelikhuijsen, M. Hogan, S. Enzerink, A. Ojo, and A. Meijer, "Connecting societal issues, users and data. scenario-based design of open data platforms," *Government Inf. Quart.*, vol. 34, no. 3, pp. 470–480, Sep. 2017.
- [45] J. Van Tassel, *Digital Rights Management: Protecting and Monetizing Content*. Waltham, MA, USA: Focal Press, 2006.
- [46] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, nos. 3–4, pp. 591–611, Dec. 1965.
- [47] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, Oct. 2002.
- [48] M. E. J. Newman, "Mixing patterns in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 67, no. 2, Feb. 2003, Art. no. 026126.
- [49] M. E. J. Newman and J. Park, "Why social networks are different from other types of networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 68, no. 3, Sep. 2003, Art. no. 036122.
- [50] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proc. Nat. Acad. Sci. USA*, vol. 101, pp. 5200–5205, Apr. 2004.
- [51] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the Facebook social graph," 2011, *arXiv:1111.4503*. [Online]. Available: <http://arxiv.org/abs/1111.4503>
- [52] S. S. Perera, M. G. H. Bell, M. Piraveenan, D. Kasthurirathna, and M. Parhi, "Topological structure of manufacturing industry supply chain networks," *Complexity*, vol. 2018, pp. 1–23, Oct. 2018.
- [53] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, Aug. 2002.
- [54] E. Ravasz and A.-L. Barabási, "Hierarchical organization in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 67, no. 2, Feb. 2003, Art. no. 026112.
- [55] B. H. Sims, N. Sinityn, and S. J. Eidenbenz, "Hierarchical and matrix structures in a large organizational email network: Visualization and modeling approaches," in *Social Network Analysis—Community Detection and Evolution* (Lecture Notes in Social Networks), R. Missaoui and I. Sarr, Eds. Cham, Switzerland: Springer, 2014.
- [56] S. Fang and L. Qi, "Manufacturing network flows: A generalized network flow model for manufacturing process modelling," *Optim. Method. Softw.*, vol. 18, no. 2, pp. 143–165, 2003.
- [57] A. Vázquez and Y. Moreno, "Resilience to damage of graphs with degree correlations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 67, no. 1, Jan. 2003, Art. no. 015101.
- [58] G. Paul, S. Sreenivasan, and H. E. Stanley, "Resilience of complex networks to random breakdown," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 72, no. 5, Nov. 2005, Art. no. 056130.
- [59] M. E. J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, no. 20, Oct. 2002, Art. no. 208701.



**TERUAKI HAYASHI** received the Ph.D. degree in engineering from The University of Tokyo. He is currently an Assistant Professor of systems innovation with the School of Engineering, The University of Tokyo, and the Vice Chairman of the Application Committee with the Data Trading Alliance. His specialization is in knowledge structuring for Data Utilization and Scenario Creation. He developed methods for supporting data exchange and understanding its ecosystem as his

core research and applied the results internationally to industry, government, and academia. He is also the coauthor of the book *Market of Data* (Kindaigakusha, 2017). He was awarded the Dean's Award by the School of Engineering, The University of Tokyo, in 2017, and an Excellence Award at the 23rd Annual Conference of the Japanese Society of Artificial Intelligence, in 2018.



**GENSEI ISHIMURA** received the Ph.D. degree in education from Hokkaido University, in 2018. He is currently a Professor with the Professional University of Information and Management for Innovation. Since 1999, he has been engaged in formulating the basic concept of construction of science museums, direction of exhibitions, and management consulting. He has been a specially appointed Associate Professor and a special Associate Professor with Hokkaido University, an Associate Professor with the Communication in Science and Technology Education and Research Program, and a special Associate Professor with the Tokyo Institute of Technology.



**YUKIO OHSAWA** (Member, IEEE) received the Ph.D. degree from the School of Engineering, The University of Tokyo, in 1995. He is currently a Professor of systems innovation with the School of Engineering The University of Tokyo. Before moving back to The University of Tokyo, he was a Research Associate with the School of Engineering Science, Osaka University, from 1995 to 1999, and subsequently, he was an Associate Professor with the Graduate School of Business Sciences, University of Tsukuba, from 1999 to 2005. In the field of artificial intelligence, he has created a new domain known as "Chance Discovery" to discover events with significant impact on decision making. He has delivered many keynote talks about Chance Discovery at conferences, such as the International Symposium on Knowledge and Systems Sciences, the International Conference on Rough Sets and Fuzzy Sets, the Joint Conference on Information Sciences, and Knowledge-Based Intelligent Information and Engineering Systems. Chance Discovery has been embodied as an innovators' marketplace, a methodology for innovation that is borrowed from principles of market dynamics. He has also published 100 journal articles and initiated symposia and workshops on data-based approaches to business innovation. His original concepts and technologies have been published as books and monographs by global publishers, such as Springer, Verlag, and Taylor and Francis. The two most important books among these are *Chance Discovery* (Springer, 2003, Eric von Hippel gave the opening) and *Innovators' Marketplace: Using Games to Activate and Train Innovators* (Springer, 2012, Larry Leifer gave the opening). He has edited special issues as a Guest Editor for journals mainly related to chance discovery such as *Intelligent Decision Technologies* (2016), *Information Sciences* (2009), *New Generation Computing* (2003), and the *Journal of Contingencies and Crisis Management* (2002).

• • •