

Designing Data Governance in Platform Ecosystems

Sung Une Lee

¹Data61, CSIRO

²University of New South Wales

Sungune.Lee@data61.csiro.au

Liming Zhu

¹Data61, CSIRO

²University of New South Wales

Liming.Zhu@data61.csiro.au

Ross Jeffery

¹Data61, CSIRO

²University of New South Wales

Ross.Jeffery@data61.csiro.au

Abstract

As platform ecosystems such as Facebook or Twitter are rapidly growing through platform users' data contribution, the importance of data governance has been highlighted. Platform ecosystems, however, face increasing complexity derived from the business context such as multiple parties' participation. How to share control and decision rights about data assets with platform users is regarded as a significant governance design issue. However, there is a lack of studies on this issue. Existing design models focus on the characteristics of enterprises. Therefore, there is limited support for platform ecosystems where there are different types of context and complicated relationships. To deal with the issue, this paper proposes a novel design approach for data governance in platform ecosystems including design principles, contingency factors and an architecture model. Case studies are performed to illustrate the practical implications of our suggestion.

1. Introduction

Platform is defined as the building blocks that act as a foundation upon which an array of firms [1]. It is sometimes called a business ecosystem or platform ecosystem (PE). PE has recently become a key business trend in industry. Sustainable growth of PEs is enabled through network effects which are based on the interaction of multiple participating groups [2-4]. As the popularity of PEs and the value of data in PEs are increasing, the concern about data abuse or misuse is also growing. Therefore, the role of data governance in PEs has received attention since it encourages desirable behavior of all participating groups [5].

Data governance refers to who holds the decision rights and is held accountable for decision-making about data assets [9]. Traditional data governance focuses on data assets within an enterprise, and therefore there are simple and clear relationships to be addressed. Data governance for PEs, however, should

consider the different business context and concepts as there are the multi-sided networks of participating groups and the interactions between them [5]. How to partition the decision rights and power of a PE between a platform owner and platform users, therefore, becomes a big challenge [6, 10]. However, there is limited research on this issue. According to the analysis on 19 existing industry and academic governance frameworks [5], most of them focus on in-house control for an enterprise. There lacks consideration of external participating groups which contribute and/or use data in PEs. An elaborate analysis of the interaction of roles and responsibilities and the design of decision-making structures in the context of PE is missing. Prior studies on platform governance largely neglect the importance and role of data [5, 10]. Those claimed issues are derived from a lack of understanding and consideration of data in PEs, and the underlying complicated relationship caused by multiple participating groups [10]. Designing data governance helps deal with the complexity of PEs, and focus on identifying fundamental considerations [9]. Architecture is a tool for precisely describing data governance elements and the interconnections between them, and simplifying the complexity when designing data governance in PEs [6].

In this paper, we aim to provide an overall design approach for data governance in PEs that can be used by practitioners to focus on important data governance issues and to develop an effective data governance strategy and design for business success.

The next section provides a literature review. We explain the methodology of this study in section 3. Section 4 introduces our data governance architecture. It presents design principles which should be followed, contingency factors to help platform owners' decision-making for design choices (centralized or decentralized type), and a data governance architecture model. We focus on decentralized data governance which needs to tackle more complexity. Section 5 shows case studies to illustrate the practical implications of our suggestion. We then conclude and discuss the limitations of the study.

2. Literature review

Data governance includes processes, policies and structures for comprehensive control. Traditional organizations can easily handle participants (employees) and the relationship between them. Yet, platform owners have limited power and ability to fully control platforms as there are multiple parties contributing, deriving and using data [5, 7]. In this different context, how to design data governance effectively aligning the business goals and platform concepts have not been discussed in detail.

Design principles addressed in prior research show the fundamental considerations in designing platform governance. Transparent, fair, simple, realistic, shared value, and participation are identified as the main principles [4, 6]. The literature discusses not only the principles but also the business influence like costs, benefit, or platform growth. However, there is an absence of explanation about how the principles can be affected by different types of governance design (e.g. centralized or decentralized). A data governance type is determined by a platform owners' design choice. While centralized governance means that a platform owner takes all control and responsibility, decentralized governance shares them with platform users. A number of studies have addressed design choices (how to balance ownership/power of all sides in the ecosystems) [6, 10]. Hein et al. [11] addressed governance structure, which refers to centralized or decentralized governance as a critical mechanism. The authors claimed that it involves how the authority and responsibility for decision-making is divided among participating groups. Schrieck et al. [10] also described the distribution of power in PEs. Those prior studies, however, do not focus on data, and thus how data governance should be designed is not discussed.

According to "contingency theory", the relationship between some characteristic of an organization and the organization's effectiveness is determined by contingencies [40]. The theory respects the fact that each PE needs a specific data governance configuration by looking at current and future context or strategy. It helps platform owners to systematically analyze their contingencies and to select a better choice. However, the use of the theory in data governance has received little attention. Weber et al. [8] applied the theory to data governance, and it is reported as the first study. The authors proposed a contingency model for designing data governance, emphasizing the model contributes to the business goals of a company. The study has a focus on the general context for an enterprise. It lacks consideration of how it can be applied in a different business context such as PEs. We reviewed literature on PE to find useful ideas, and thus

to overcome the limitation of the model. A network effect for platform's growth and profit [2, 13-17], platform governance mechanisms and strategies, single/multi-homing, platform maturity and open strategy [16, 17], control mechanisms [6, 10, 11, 18], and different types of market structures [3] are addressed as key aspects of platform governance. These aspects, however, have not been explained in terms of how they are related to data governance.

Based on the contingencies of PEs, platform owners can decide their governance type. When a platform owner chooses decentralized data governance, he needs to handle a more complex architecture. It is necessary to consider the decision domains that need to be shared, and who should make the decisions. Decision domains of data governance for PEs are proposed by Lee et al. [5]. The authors identified seven data governance factors for PEs: data ownership/access definition, regulatory environment, contribution measurement, data use case, conformance, monitoring and data provenance. The factors are used to place decision rights between platform owners and users at design time. A decentralized data governance architecture describes all those considerations in a structural design to reduce the complexity [6].

To help successful implementation of a decentralized architecture, technical considerations are also important. Blockchain is identified as one of the possible technologies. It has emerged to support a digital currency based on a peer-to-peer network without trusted authorities [37]. It has been recently highlighted as decentralized governance for data provenance or data ownership [37, 39]. Choudary and Van Alstyne [4] noted that blockchain is one of the most innovative and revolutionary governance forms, and supports trustworthy interactions based on decentralized architecture. It enables transparent and trustworthy governance as any transaction in the blockchain is verified by every node, and permanently recorded. Based on the mechanisms, Zyskind et al. [39] proposed a decentralized personal data management system to achieve a transparent data supply chain, data ownership and privacy. The data provenance issue is addressed in Liang et al. [36] through a provenance database and a blockchain network to provide tamper-proof records for transparency of data accountability. These studies, however, overlook the complicated relationships and context which can occur in PEs through multiple participating groups, various types of data and a complex data supply chain.

In conclusion, all available existing data governance sources from scholars and practitioners only focus on organizational concern. Moreover, platform governance studies fall short of attention to data governance.

3. Methodology

This study is based on two main methodologies: a literature review and case study (Figure 1).

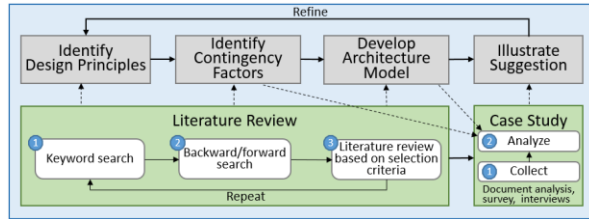


Figure 1. Overview of the research process

We conducted three steps for the literature review: keyword search, backward and forward search, and literature review based on selection criteria. Firstly we broadly searched to find the literature which addresses PE strategy, governance concepts and mechanisms. The main keywords were used “((platform ecosystem OR multi-sided platform OR two-sided platform) AND (governance OR management))”. The selection criteria was “Include concepts, strategy, or governance mechanisms in the context of PE”. We excluded the papers that are not related to platform governance, too high level topics (e.g. overview), specific domains and technologies or not academic. We then performed a quick review (abstract/introduction/conclusion) followed by backward and forward search and a full content review. We repeated the three steps to narrow down the search scope or to find the literature which has specific topic with different keywords and selection criteria. For instance, “(blockchain AND governance)” were used for our decentralized governance architecture model. We chose the papers which address data governance domains of PEs such as ownership, data provenance or contribution. We also used the literature searched in previous studies [5].

Case studies were carried out to illustrate our suggestion: the contingency factors and decentralized governance model. We used currently running or developed platforms for the studies. We chose two (Platform A and B) undergoing different business stages, which provide platforms and support participation of multiple user groups. Data collection was performed by document analysis (Platform A and B), survey of the web sites (Platform A), and interviews (Platform B). The context of Platform A was described by one of the authors of this paper in the beginning as she worked with the platform in the past. The interviews for Platform B were conducted with two members of the platform, using eight closed/fixed-response questions and two standardized/open-ended questions. We analyzed the collected data using the

contingency factors and decentralized model. We analyzed the influence of the contingency factors on the data governance design choices of the platforms, and illustrated the possibility of the concept and technical requirements of our data governance model.

4. Data governance architecture

4.1. The principles of design

We begin by introducing the principles which should be considered for designing good data governance in the context of PEs. We identify six fundamental principles: transparent, fair, simple, realistic, shared value and participation [4, 6].

Transparent means that platform governance should give a clear view to every participant. It includes internal transparency for platform companies themselves and external transparency for platform users. **Fair** is relevant to applicable rules for everyone. There should not be inconsistent rules, and thus governance policies should be equally applied to all participating groups. In addition, every participant should be given the same opportunity and accessibility. Fairness results in more participation and ideas, and after all it leads to new innovation [4]. Moreover, it makes more wise and productive use of data in PEs [34]. **Simple** presents that data governance in PEs should be designed and implemented as efficiently and effectively as possible by making its structure simple. It is strongly related to costs and benefit of platforms. **Realistic data governance** can be realized by starting with a good understanding of what governance practices are applicable and how they work based on design choices and trend (toward centralized or decentralized). **Shared value** is a management strategy which should be delivered to all participants. Data governance design, therefore, should contain the strategy, and it should be implemented through the realized systems. **Participation** presents that platform owners should give all participating groups of PEs a chance to join decision-making processes in certain ways. In decentralized data governance, it can be highly encouraged and implemented.

The principles should be considered for both design types (centralized/decentralized). Yet, the degree of achievement of each principle can be different. Figure 2 shows decentralized data governance supports a higher expectation of transparency, shared value and participation. If a platform focuses on trust-based governance or less/no restriction to platform users, the platform is more visible and easy to participate and have a common consensus between participants. It is made possible by sharing power and control of

decision-making with platform users. In decentralized data governance, platform owners should understand, consider, and reflect platform users' needs and characteristics, and it might be more difficult than in centralized data governance. Thus, it can lead to less realistic data governance. In contrast, centralized data governance enables simpler and more realistic data governance since platform owners do not need to consider platform users' issues as much. "Fair" might not be affected by any design choice.

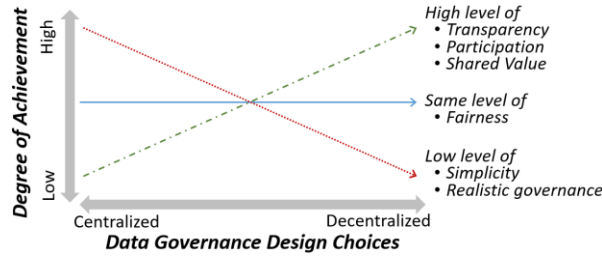


Figure 2. Design choices and principles

4.2. Contingency factors

Depending on platform owners' design choice, platforms can expect different advantages (Table 1). It is necessary to think about what platform contingencies affect design choices to find the best choice and thus to get business success of each PE (Figure 3).

Table 1. Centralized vs. decentralized governance

Perspective		Platform owners	Platform users
Design	Centralized data governance	Pros: Easy to control user behaviors and to align business goals/strategies	High quality of data or services (including strong security)
		Cons: Slow growing and lots of resources	Invisible platforms, lack of trust and accessibility
Decentralized data governance	Pros	Reduce effort, increase growing and user satisfaction/participation	Enhance trust and increase motivation and benefit expectation
	Cons	Hard to control platforms and user behaviors and to change goals/strategies	Complicated processes, slow decision-making and less secure

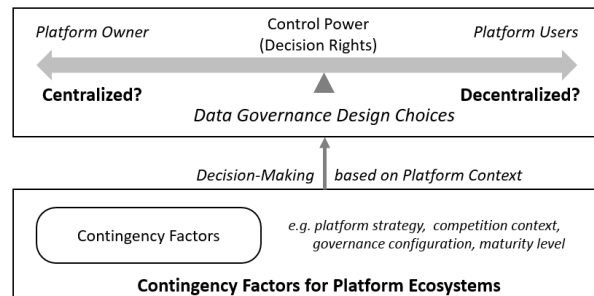


Figure 3. Design choices and contingency factors

To support right decision-making of platform owners, we suggest the contingency factors which are

adopted from a previous contingency model for data governance (Weber et al. model) [8] and platform governance studies [2-4, 6, 8-11, 13, 16-21, 27, 29, 31-33]. Weber et al.'s model provides generic contingency factors for an enterprise which are derived from prior research and workshops with practitioners. The factors are identified for data governance, and thus useful as a basis of this study. We differentiate them by focusing on platform governance characteristics (Figure 4).

Weber et al.'s contingency factors		Contingency factors for PEs (our suggestion)	
Performance strategy		Platform strategy	Modified
Diversification breath		Multi-homing strategy	Modified
Degree of market regulation		Degree of market regulation	Selected
Organization structure		Governance configuration	Modified
Competitive strategy		Platform market structure	Modified
Process harmonization		Process harmonization	Deleted
		Open strategy (new)	Added
		Platform maturity (new)	Added

Figure 4. Contingency factors

First of all, we examined Weber et al.'s model to select applicable factors for PEs. We selected "degree of market regulation" as regulatory environment is also an important factor in using data in PEs [5]. Secondly, we removed "process harmonization" because it refers to business processes of an enterprise. Next, we modified the remaining factors to adjust to PEs' situations through a literature review. Four factors were modified: "platform strategy", "multi-homing strategy", "governance configuration" and "platform market structure". The factors align with the basic concepts of Weber et al.'s model, but they differ in the definition and range of each factor. Lastly, we found new factors which are not addressed in the referred model. We identified two factors ("open strategy" and "platform maturity") as new factors. The factors are commonly discussed when addressing platform governance in academic literature. In total, we identify seven contingency factors which influence the design choices of data governance in PEs (Table 2).

Platform strategy— Weill and Ross [13] addressed profit and growth of organizations to measure IT governance performance. Weber et al. [8] adopted this concept to data governance contingency. The authors noted that if an organization focuses on profit, the decision-making authority in governance will be toward centralization. In contrast, if growth is emphasized, the governance will be decentralized. In the context of PEs, however, profit and growth can be achieved at the same time thanks to network effects of PEs. As PEs are increasing revenue based on data from platform users, low quality of data issues are also rising [19]. Some platforms have addressed the issues by adopting formal control or monitoring to drive out low quality data [8]. However, strict control for high quality in centralized governance can make

Table 2. Contingency factors for data governance in PEs

Contingency factor		Definition (trend: centralized <-> decentralized)	Relevant data governance factor	Reference
Platform strategy		Performance objective of a PE (quality <-> profit or growth)	Data ownership/access definition, data use case, conformance, monitoring	[3, 8, 13, 19]
Multi-homing strategy		Degree of affiliation in a PE (single-home <-> multi-home)	Data ownership/access definition and use case	[8, 16, 17]
Degree of market regulation		Degree of regulation regarding the use of data in a PEs (highly regulated <-> no regulation)	regulatory environment, conformance	[8, 9, 27, 29, 30]
Governance Configuration	Governance mechanisms	Type of governance of a PE (authority-based <- contract-based -> trust-based)	Contribution measurement, conformance, monitoring, data provenance	[6, 18, 20, 21]
	Control mechanisms	Type of control of a PE (formal <-> informal control)	Data ownership/access definition, conformance, monitoring, data provenance	
Platform market structures		Type of platform market structures based on competition (monopoly platforms <- intersecting -> coincident)	Data ownership/access definition, data use case, monitoring, data provenance	[3, 17, 31]
Open strategy		Level of openness of a PE (close <-> open)	Data ownership/access definition, data use case, monitoring, data provenance	[4, 10, 11, 17, 32]
Platform maturity		Level of maturity of a PE (mature platforms <-> immature)	Contribution measurement, monitoring, data provenance	[2, 4, 10, 11, 17, 33]

platform growth slow. **Multi-homing strategy**— Some platform owners require their partners to affiliate exclusively with them to offer novel content (single-home). Meanwhile, some platforms allow affiliation with competing platforms to encourage participation (multi-home): e.g. multiple payment cards in EBay [3, 17]. Multi-homing is related to openness. If a platform chooses “open”, less permission rule to the users is followed [16, 17]. It means governance should move toward decentralization to share decision rights [6]. **Degree of market regulation**— This factor is adopted from the existing model [8] as it is consistent with concern of regulatory environment of data governance in PEs [5]. A highly regulated environment requires more centralized data governance for compliance. On the contrary, less (or no) regulation might enable platform owners to divide decision rights through a decentralized approach. **Governance configuration**— Governance mechanisms can be categorized into three: authority-based, contract-based and trust-based [18, 20]. Authority-based mechanisms can be realized within centralized governance by employing the platform owner’s power to enforce desirable behavior based on policies. Trust-based mechanisms are used for gaining a certain amount of participants, and then attracting other side users. The mechanisms are generally combined with strong incentives to reach the desired goal [2, 18]. In general, authority-based mechanisms are related to formal control like input, output or behavior control, but trust-based mechanisms are enabled by informal control such as clan/social control [18, 21] in decentralized governance. A contract-based mechanism needs less formality, but still prohibits the unauthorized behavior in using data in platforms. **Platform market structures**— Platform market structures (coincident, intersecting, and monopoly) have been addressed under platform competition concerns [3]. Coincident platforms are recognized when there is too much competition in the market; supply sides (n) > demand

sides (m). Monopoly refers to having no competition: n = 1. Coincident platforms (in particular, entrants) might be necessary to be more attractive than incumbent platforms. They tend to be open and share the platforms to encourage users’ participation and to win market [17]. Monopoly leans to centralized governance to strictly control the platforms. **Open strategy**— Open strategy is an important part of the design of a PE [10, 11]. A platform owner opens the platform when there is lack of resources, need for adoption or innovation [17]. If the degree of openness increases, platform users can easily access the platforms, and platform owners need to share control with users [10]. In consequence, it is toward decentralized. Yet, limited openness leads to high process control, quality and user satisfaction: e.g. Apple’s app store [4, 11]. **Platform maturity**— A newcomer platform has more permissive governance to enhance participation and to reach critical mass when market share is small [17]. To attain enough growth, a platform owner needs less restriction but more trust to attract and lock the users in the platform [4, 10]. In particular, trust is regarded as a prerequisite for a PE to survive among competitors [10, 11, 33]. Thus, less mature platforms tend to prefer multi-homing, open and less negotiation (share control) to capture the benefits of growth. In contrast, when market share is large enough (at maturity), governance has a tendency to be toward tighter control [17].

As shown in Table 2, each contingency has a relationship with particular data governance factors. The data governance factors should be focused and enforced based on the linked contingency. For instance, if a platform aims for high quality performance, the platform needs to strengthen data ownership and access control based on clear data use cases [9, 13, 22, 23] for strict responsibility and accountability of data quality [19]. Conformance and monitoring should also be followed to improve quality. Meanwhile, a PE which adopts a trust-based governance needs to think about a certain reward for

data contributors to encourage participants [18, 24, 25]. Shared conformance, monitoring and data provenance should be considered to achieve visibility of a data supply chain for trust between a platform owner and users [26-28]. In the same vein, degree of market regulation is strongly related to regulatory environment and conformance to avoid serious court issues [9, 27, 29, 30].

4.3. Data governance architecture model

A centralized governance architecture is simpler as there are only internal issues to consider. In contrast, a decentralized architecture needs to deal with complicated relationships and interactions among multi-sided networks of participants. To cope with this complexity, we present how to partition data governance control power with platform users. We then propose a decentralized architecture model and the technical considerations. Blockchain is introduced as a possible technology which can support our model.

4.3.1. The locus of data governance. To encourage participation of platform users and to achieve visibility of PEs, some governance practices need to be directly implemented by platform users. We partition the governance factors and practices introduced in the literature review section (Table 3).

Table 3. Data governance partitioning

Governance factor	Locus of data governance	
	Platform owner side	Platform user side
Data ownership/access	. Define data ownership and access rights	. View data ownership and access rights
Regulatory environment	. Identify relevant regulations, laws or cases . Develop/use decision models for data ownership/access definition	-
Contribution measurement	. Develop contribution measurement models . Measure/notify the contribution to the owner	. Measure the contribution of data
Data use case	. Define data categories and use cases	. View the defined categories and use cases
Conformance	. Build processes/rules . Audit processes/rules	. Audit processes/rules
Monitoring	. Monitor/view all data activities(events) . Notify specific events to the data owner . Review user report and response to the reporter	. Monitor/view activities (or events) regarding uploaded /generated user data . Report misuse/abuse of data
Data Provenance	. Identify or authenticate data owner	. Trace the history of the used of data

Platform owners should carefully consider the regulatory environment to correctly define data ownership and access rights based on the identified legal aspects [9]. Platform owners also need to define

how to use the data in PEs for business goals. Understanding and defining different types/sources of data are thus very important to platform owners [9, 35]. In these cases, platform users' participation is little. However, the mechanisms and consequence of decision-making by platform owners should be disclosed to all participant groups [5]. In addition, platform users also need to clearly know and understand who, when and how uses their data [26]. However, prior research and case studies claim a lack of user contribution model and invisible data usage as critical challenges [5, 10, 24-26] even though user contribution measurement has been considered as a key mechanism to support revenue sharing and business success [5, 24, 25]. To encourage high visibility and fair and clear contribution measurement, decentralized data governance should play a critical role by enabling platform users to monitor and validate the whole process and history of the use of data in a supply chain [6, 9, 26].

4.3.2. Architecture overview. As illustrated in Figure 5, we suggest our decentralized model for PEs based on the design principles and data governance partitioning mentioned in the previous sections.

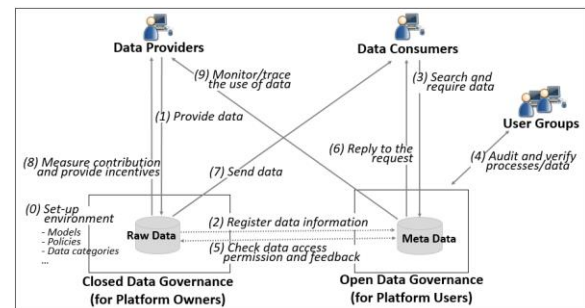


Figure 5. Decentralized data governance

There are three main entities, which interact with each other: platform users (data providers, consumers and user groups), closed and open data governance. The closed governance is for platform owners. It supports internal practices like data ownership/access definition and contribution models. The open data governance enables every platform user to fairly access platform data and transparently see or audit all activities and events of platforms. The following describes the interactions in detail.

(0) Platform owners should set up environment prior to occurrence of the interactions. It includes developing models for data ownership/access and contribution measurement, and defining data categories/use cases, aligning the business objectives and strategies. (1) Data providers can upload or generate during their use of PEs. The data is stored in

the closed data governance systems as raw data. (2) The information of raw data is registered in the open data governance as a form of metadata according to the predefined data categories and the purpose of the use of data. The registered data is exposed to any platform users who want to use them. (3) Data consumers can search and require the registered data. (4) Verifying data integrity, checking access permission, and auditing the processes are implemented by platform user groups through the open data governance systems to avoid data abuse or misuse (To enable user groups' participation for these activities, some technical considerations and mechanisms should be considered. We will discuss this in the next section in detail). (5) In order to check data access permission, the closed data governance should provide correct information of raw data based on the data ownership and access definition. (6)(7) According to the feedback of the open data governance, data consumers can receive the data via the raw data repository in the closed data governance. (8) Platform owners measure the contribution and reward the contributors as they create or add value in the use of data. (9) The owners of the data can see all the events and trace the history of the use of the data during the transaction and/or after. Furthermore, the open data governance can automatically measure the contribution of data based on the usage of data.

This model highly enhances transparency and participation of PEs by providing an open data governance platform. All participating groups can share control, value and culture, and clearly see the use of data. Platform owners can reduce costs and efforts by partially handing over some parts of their control power to platform users.

4.3.3. Technical considerations. To discuss how the suggested model is implemented, we identify the technical requirements. Firstly, the open data governance should be completely open to every platform user. It also should be decentralized without any intervention of platform owners. Meanwhile, the closed data governance needs to be invisible and protected. The two separated governances, however, have to cooperate/collaborate with each other to keep consistency and integrity of data, and seamless services. Secondly, security issues should be considered as the governance model allows any participants (platform users) to join the open data governance system. To support this, all activities/events which take place in PEs have to be safely preserved/recorded. It can also improve clear data ownership and provenance. The recording can be used to prevent or detect denial of data usage or data manipulation by someone including platform

owners. This is important because it is necessary to improve transparency of PEs, where there are multiple participating groups and thus there are worries of data abuse/misuse and privacy violation [5]. Lastly, decision making by user groups by auditing the processes/data should be clear and stable, and implemented by reasonable mechanisms.

Blockchain technology meets the requirements mentioned above. It provides a shared and distributed architecture [36-39]. Based on the concept, the hashed data of raw data (uploaded data by platform users) can be generated and transferred to the open data governance. This enables the two governance (closed and open) to be independent, but still keeps data security, consistency and integrity by verifying the hash code against the raw (original) data. Once the hashed data embedded in the transaction by the blockchain network (the open data governance in our model), the data becomes immutable. This is possible since every participant of the blockchain network stores the data as a form of a cryptographically-signed chain. Thus, it is nearly impossible to change the data [37] unless someone can own enough computational power to invalidate the chain. This mechanism satisfies our needs regarding security and data provenance. In addition, auditing every transaction for the use of data and making a decision can be conducted by every participant's voting. It is implemented with clear rules without platform owners' intervention. Therefore it is regarded as a trustworthy and transparent mechanism.

5. Case study

We conduct case studies on two platforms (Platform A and B). First of all, we compare the platforms based on our contingency factors (Table 4). We illustrate how the governance design types are characterized depending on the different context. Secondly, we analyze Platform B in detail to show the practical implication of the decentralized architecture model suggested here.

Platform A is a content portal which was launched four years ago. Through the platform, software assets (development knowledge or documents) are provided and reused by IT companies or individual developers. To encourage user participation, it provides several types of benefits to the participants: e.g. subsidy and seeding like UI/UX guide. The government supports the platform by legislating the rule that every government SW R&D project should register the outcome to the platform. According to the result of the case study, Platform A tends toward centralized. The market regulation and platform maturity of the platform lead

Table 4. The results of the case studies on platform A and platform B

Platform A -> Toward “Centralized Data Governance”		Platform B -> Toward “Decentralized Data Governance”	
Contingency factors and the context	Trend	Contingency factors and the context	Trend
<p>* Platform strategy: Quality The business goal is reputation and satisfaction of users. Thus, it aims at high quality of data rather than growth of the platform. Profit is not considered as the platform is a non-profit platform.</p>	Centralized	<p>* Platform strategy: Growth The purpose of the platform is to share as much data as possible, and the quality of data is not a big concern of the platform. Thus, it aims at growth of the platform.</p>	Decentralized
<p>* Multi-homing strategy: Single-home It tends to prohibit multi-homing for high reputation/differentiation.</p>	Centralized	<p>* Multi-homing strategy: Multi-home In order to open the platform to everyone and to gain more data from providers, the platform allows multi-homing.</p>	Decentralized
<p>* Degree of market regulation: No (less) regulation Every uploaded data is public data based on the prerequisite of the platform policies (except user information). There is less amount of sensitive data; e.g. personal identifiable information.</p>	Decentralized	<p>* Degree of market regulation: No (less) regulation The data stored in the platform is basically public data to open to everyone as there is no raw data and personal information. The platform has only metadata provided by data owners.</p>	Decentralized
<p>* Governance configuration: Authority-based/Formal control There are strict/formal control processes and clear ownership mechanisms by the platform owner.</p>	Centralized	<p>* Governance configuration: Trust-based/Informal control Reputation mechanisms will be used by platform users. Users’ reputation score can be used for input control by the government.</p>	Decentralized
<p>* Platform market structures: Monopoly As the platform is supported by the government, the platform dominates the market as a monopoly platform.</p>	Centralized	<p>* Platform market structures: Intersecting There are some identified competitors (e.g. data.gov.au and the other government platforms). However, there is no serious competition.</p>	Blended
<p>* Open strategy: Close It tends to open to demand side users, but still requires login first (only authorized users can access to the platform for the use of data in the platform). For supply side, the platform is relatively closed as it requires high quality and reputation.</p>	Centralized	<p>* Open strategy: Open The strategy of the platform is completely open. There is no any restriction to join the platform, and thus everyone can equally use the data in the platform. Platform users only need to login using blockchain accounts.</p>	Decentralized
<p>* Platform maturity: Immature The platform launched four years ago, but it is still considered as an immature platform due to the growth rate of data and users.</p>	Decentralized	<p>* Platform maturity: Immature The platform has not started its business yet.</p>	Decentralized

to decentralized governance, but the other factors show centralized governance trend of the platform.

Meanwhile, Platform B has not officially started the service yet. The project team recently finished the prototype system, and thus it is in the very initial phase. The platform is an open data registry platform. It allows data owners to upload data, and data analysts can use the data for analytics jobs. The data owners can monetize their data since it provides payment systems. According to the interviews and document review, the governance trend is decentralized. The vision of the platform is open, and thus it aims to share data with every participant. Since the governance trend of Platform B is identified as decentralized, we carry out an in-depth analysis to illustrate how our decentralized model (Figure 5) can be implemented in the real world. We survey the platform in three dimensions: architecture, process and function.

Architecture— Platform B is developed based on blockchain. There are two entities named “on-chain and off-chain” based on the existence of a blockchain network. In the off-chain (without a blockchain network), there are hosted raw data, and usage policy specification and a policy checker. Meanwhile, dataset and analytics job registries which interact with data consumer (data analyst) and data provider are in the on-chain. The on-chain also includes modules for tamper-proof and payment to support high security and monetization. **Process**— Data providers can register datasets on blockchain. Data consumers can post analytics jobs with the requirements and measurement criteria of datasets. They can select the datasets which

meets the requirements among the registered datasets. The platform conducts policy compliance checking to validate if the access to the datasets is valid. Lastly, the data analysts run the analytics jobs with the datasets, and pay for the datasets. **Function**— The main functions of the platform are dataset and analytics job registration, policy checking and payment. The additional functions are dashboards for data consumers and data owners. It supports presentation views which show the results of data searching and policy checking, the contribution of each dataset selected by data consumers, and the profit of data owners.

The overall concept is similar to our architecture model. Raw data is stored in the off-chain and only metadata is provided to the on-chain network. This mechanism helps more privacy of datasets as it stays far from direct exposure to public. The platform utilizes advantages of blockchain such as transparency, secure and stable services, or fairness. This case can be a good reference of a practical governance model which can be used in industry. However, there are some limitations which should be dealt with for PEs. First of all, there is a lack of consideration of a scalability issue. Majority PEs may produce a large amount of data every day, and have various types of data. For instance, Facebook has more than 70 categories of data [5]. All types of data do not need to be opened to or shared with platform users. It is necessary to categorize and filter data before transferring them to an open platform. It is able to reduce the burden of data processing for better performance. Secondly, while the open data registry

focuses on simple and clear data ownership and contribution, data governance in PEs should deal with multiple parties contributing, deriving and using data. This concern results in complicated relationships, ownership and contribution measurement [5]. Platform B, in particular, cannot answer the question of how to measure the contribution of shared or derived data which may involve multiple data owners through transformation or analysis. Lastly, the policy compliance checker is only applicable for tabular data focusing on ETL (extract, transform and load) phase. As PEs generally deal with various types of data including video, audio, text and geographic data. Different mechanisms should be adopted.

To sum up our case studies, we showed that our contingency factors are applicable to systematically analyze and characterize the data governance design of PEs. For currently running platforms (e.g. Platform A), the contingency factors can be used to review current governance direction. For a newcomer (e.g. Platform B), it enables to examine if the vision and strategies of the platform align with the data governance design choice. The interviewees gave a positive feedback that our suggestion is effective and useful to analyze their context and to help decision-making for governance design choices. One interviewee noted that different types of users (providers and consumers) need to be considered when analyzing the governance configuration. We also illustrated the concept and technical approach of our architecture model by in-depth analysis of Platform B. The interviewee's comment and the identified limitations from the case studies will be addressed in the future research.

6. Conclusion and limitations

In this paper, we proposed a data governance architecture for PEs. We discussed data governance design choices (centralized or decentralized), and contingency factors to support decision-making. We also proposed an architecture model for decentralized data governance based on the design principles suggested in this paper. We introduced blockchain technologies which support our model. To illustrate the practical implications of our suggestion, we carried out case studies on two platform ecosystems.

In order to provide in-depth understanding and knowledge of data governance and PE, we conducted a literature review following three steps: keyword search, backward/forward search and literature review. The suggested contingency factors which influence data governance design of PEs, were differentiated from prior studies based on the characteristics, strategy, and governance concepts and mechanisms of PEs. The case studies were performed by applying our

contingency factors with survey and interviews. We showed how the factors can be used in the real world. The decentralized governance model was also illustrated through a case study on a blockchain-based system which implements decentralized data governance. We also identified significant challenges which should be addressed in the future research.

This study has several limitations. The case studies conducted are not sufficient to validate our models. One of the case studies was carried out using a qualitative description of the author's working experiences and information from the document analysis. We conducted only two interviews for the other case. In addition, our architecture model should be implemented to clearly show if the model is applicable in the real world. At the same time, an extensive and a systematic literature review should be considered to refine our study.

7. References

- [1] Gawer, A., *Platforms, Markets and Innovation*, Edward Elgar Publishing, 2009.
- [2] M. Cusumano, "Technology strategy and management. The evolution of platform thinking," *Communications of the ACM*, 2010, (53:1), p. 32.
- [3] D. S. Evans, "Platform Economics: Essays on Multi-Sided Businesses," *Competition Policy International*, 2011.
- [4] Parker, G. G., M. W. Van Alstyne, and S. P. Choudary, *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. WW Norton & Company, 2016.
- [5] S. U. Lee, L. Zhu, and R. Jeffery, "Data Governance for Platform Ecosystems: Critical Factors and State of the Practice," in the 21th Pacific Asia Conference on Information Systems (PACIS), 2017.
- [6] Tiwana, A., *Platform Ecosystems: Aligning Architecture, Governance, and Strategy*, Newnes, 2013.
- [7] A. Smedlund, and H. Faghankhani, "Platform Orchestration for Efficiency, Development, and Innovation," in 2015 48th Hawaii International Conference on System Sciences, 2015, pp. 1380–1388.
- [8] K. Weber, B. Otto, and H. Österle, "One Size Does Not Fit All---A Contingency Approach to Data Governance," *Journal of Data and Information Quality* (1:1), 2009, pp. 1–27.
- [9] V. Khatri, and C. V. Brown, "Designing data governance," *Communications of the ACM* (53:1), 2010, pp. 148–152.
- [10] M. Schreieck, M. Wiesche, and H. Krcmar, "Design and Governance of Platform Ecosystems—Key Concepts and Issues for Future Research," in Twenty-Fourth European Conference on Information Systems (ECIS), 2016.
- [11] A. Hein, M. Schreieck, M. Wiesche, and H. Krcmar, "Multiple-Case Analysis on Governance Mechanisms of Multi-Sided Platforms," In *Multikonferenz Wirtschaftsinformatik*, 2016.

- [12] A. Tiwana, B. Konsynski, and A. A. Bush, "Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics," *Information Systems Research* (21:4), 2010, pp. 675–687.
- [13] P. Weill, and J. Ross, "A Matrixed Approach to Designing IT Governance," *MIT Sloan Management Review* (46:2), 2005, p. 26.
- [14] T. Eisenmann, G. Parker, and M. W. Van Alstyne, "Strategies for two-sided markets," *Harvard business review*, 84(10), 2006, 92.
- [15] T. Eisenmann, G. Parker, and M. Van Alstyne, "Platform envelopment," *Strategic Management Journal*, (32:12), 2011, pp. 1270–1285.
- [16] G. Parker, and M. Van Alstyne, "Innovation, openness & platform control," In *Proceedings of the 11th ACM conference on Electronic commerce*, 2013, pp. 95-96.
- [17] G. Parker, and M. W. Van Alstyne, "Platform Strategy," in the *Palgrave Encyclopedia of Strategic Management*, 2014.
- [18] J. Manner, D. Nienaber, M. Schermann, and H. Krcmar, "Six Principles for Governing Mobile Platforms," in *Wirtschaftsinformatik*, 2013, p. 86.
- [19] A. Hagiui, "Strategic Decisions for Multisided Platforms," *MIT Sloan Management Review* (55:2), 2014, p. 71.
- [20] M. De Reuver, and H. Bouwman, "Governance mechanisms for mobile service innovation in value networks," *Journal of Business Research*, (65:3), 2011, pp. 347–354.
- [21] W. G. Ouchi, "A Conceptual Framework for the Design of Organizational Control Mechanisms," *Management Science*, (25:9), 1979, pp. 833–848.
- [22] P. Weill, and R. Woodham, "Don't Just Lead, Govern: Implementing Effective IT Governance," SSRN Scholarly Paper No. ID 317319, Rochester, NY: Social Science Research Network, 2002.
- [23] P. Weill, and J. W. Ross, "IT Governance on One Page," SSRN Scholarly Paper No. ID 664612, Rochester, NY: Social Science Research Network, 2004.
- [24] Q. Tang, B. Gu, and A. B. Whinston, "Content Contribution for Revenue Sharing and Reputation in Social Media: A Dynamic Structural Model," *Journal of Management Information Systems* (29:2), 2012, pp. 41–76.
- [25] K. Chai, V. Potdar, and E. Chang, "User contribution measurement model for web-based discussion forums," in *2009 3rd IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, 2009, pp. 347–352.
- [26] K. E. Martin, "Ethical issues in the Big Data industry," *MIS Quarterly Executive* (14), 2015, p. 2.
- [27] A. Ghazawneh, and O. Henfridsson, "Governing Third-Party Development through Platform Boundary Resources," in the *International Conference on Information Systems (ICIS)*, AIS Electronic Library (AISeL), 2010, pp. 1–18.
- [28] A. Ghazawneh, and O. Henfridsson, "Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model: Control and Contribution in Third-Party Development," *Information Systems Journal* (23:2), 2013, pp. 173–192.
- [29] S. Kaisler, W. H. Money, and S. J. Cohen, "A Decision Framework for Cloud Computing," in *2012 45th Hawaii International Conference on System Sciences*, 2012, pp. 1553–1562.
- [30] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big Data: Issues and Challenges Moving Forward," in *2013 46th Hawaii International Conference on System Sciences*, 2013, pp. 995–1004.
- [31] G. Parker, and M. W. Van Alstyne, "Six challenges in platform licensing and open innovation," *Communications & Strategies*, No.74, 2009, pp.17–36.
- [32] A. Gawer, and R. Henderson, "Platform owner entry and innovation in complementary markets: Evidence from Intel," *Journal of Economics & Management Strategy* (16:1), 2007, pp. 1–34.
- [33] Choudary, S. P., *Platform Power: Secrets of billion-dollar internet startups*, 2013.
- [34] G. Clarkson, and M. W. Van Alstyne, "The social efficiency of fairness", 2014.
- [35] D. Firmani, M. Mecella, M. Scannapieco, and C. Batini, "On the Meaningfulness of 'Big Data Quality' (Invited Paper)," *Data Science and Engineering* (1:1), 2016, pp. 6–20.
- [36] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability," in *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, IEEE Press, 2017, pp. 468-477.
- [37] X. Xu, et al. "A Taxonomy of Blockchain-Based Systems for Architecture Design." *Software Architecture (ICSA)*, 2017 IEEE International Conference on. IEEE, 2017.
- [38] M. Atzori, "Blockchain technology and decentralized governance: Is the state still necessary?," 2015.
- [39] G. Zyskind, and O. Nathan, "Decentralizing privacy: Using blockchain to protect personal data," *Security and Privacy Workshops (SPW)*, 2015 IEEE. IEEE, 2015.
- [40] Donaldson, L., *The Contingency Theory of Organizations*. Sage Publications, Thousand Oaks, CA., 2001.