A Data Governance Framework for Platform Ecosystem Process Management

Sung Une Lee[1,2](✉), Liming Zhu[1,2], and Ross Jeffery[1,2]

[1] Architecture and Analytics Platforms Group,
Data61, CSIRO, Sydney, Australia
{Sungune.Lee,Liming.Zhu,Ross.Jeffery}@data61.csiro.au
[2] School of Computer Science and Engineering, UNSW, Sydney, Australia

Abstract. Platform ecosystem today is regarded as the key business concept of organizations to win market. Platform companies can grow fast through the data contribution of multi-sided networks. Yet, they face difficulties in managing the data resulted from complicated contribution, use and interactions between the multiple parties. The circumstance causes serious concerns about unclear data ownership and invisible use of data, and ultimately leads to data abuse/misuse or privacy violation. To alleviate to this, a particular type of data governance is required. However, there is limited research on data and data governance for platform ecosystems. We introduce a new data governance framework for platform ecosystems which consists of data, role, decisions and due processes. The framework supports organizations in understanding to show how the risks should be dealt in the processes for business success. We compare 19 existing industry governance frameworks and academic work with our framework to show current gaps and limitations.

Keywords: Data governance · Platform ecosystem · Business process

1 Introduction

All organizations use and manage data. Traditional business companies focus on the management of the "ilities" (availability or usability) of enterprise data. Since the concept of platform ecosystem (PE or business ecosystem) has been widely spread, many organizations can facilitate reaching critical mass by data contribution of two or more external sides. The collected data is analyzed or shared to add value to the companies. This generates more data and it is used by the partners or family companies and the platform users. A negative externality arises from the fact that there are complicated interactions between multiple parties providing, using or sharing data. There is a pervasive problem of data breach (data abuse, misuse or privacy issue) in business ecosystem area [1, 2]. A platform owner company should impose certain regulations on the user participation to reap the benefits of ecosystem growth [3]. Lack or poor implementation of governance causes significant destructive effects on business success.

Data governance refers to comprehensive control, including processes, policies and structures about data asset. It enables a platform owner to orchestrate the complicated processes and relationships affected by multiple parties' participation [4]. In traditional

M. Weske et al. (Eds.): BPM Forum 2018, LNBIP 329, pp. 211–227, 2018.
https://doi.org/10.1007/978-3-319-98651-7_13

data governance, data ownership is clear and simple as there is limited use of data and interactions within an enterprise. Researchers articulate a set of concerns such as unclear data ownership or invisible data supply chain in PEs [5–8]. It must be addressed in data governance to win market. In particular, the role of due process has been highlighted by researchers [8]. However, previous platform studies pay little attention to the importance and the role of data [9]. It leads to limited research in understanding how organizations should manage business processes differently for PE.

We aim to provide data governance for enterprises which run PE business. We identified data governance factors for PEs as a starting point [10]. In this study, we provide a comprehensive data governance framework which comprises three core elements, and discusses current issues and how to improve business capability. Based on that, we suggest due processes as a supporting element of the framework. The due processes encourage desirable behavior of all participating groups to mitigate business risks. This article delivers broad information and knowledge of PE and data governance through a survey on industry platforms and literature review. It helps researchers and practitioners to comprehend how data governance processes should be managed and implemented, and to plan next steps.

The next section presents background information to help an overall understanding of the general concept of data governance and PE, and the current state of academic works on platform governance including our previous studies. Section 3 describes the methodology of this study. Section 4 introduces a data governance framework for PEs. We present each element of the framework: platform data, role and decision domains, and discuss current issues and the possible solutions. We then illustrate due processes along with the data management flow of a PE. In Sect. 5, 19 industry governance frameworks and academic works are reviewed and compared with our framework. We then conclude this study in Sect. 6.

2 Background

Data governance determines who holds the decision rights and is held accountable for decision-making about data assets [11]. To support right decision-making and encourage desirable behavior, it provides comprehensive control such as processes, policies and structures. Khatri and Brown [11] noted data governance decision domains and showed how the domains align with those of IT governance. Weber et al. [12] focused on a context-based approach for data governance design by presenting how organizational contingencies influence on data governance. Those studies, however, are focused on the general business context of organizations where there are simple interactions and internal considerations in using and managing data.

PEs provide a meeting place, and facilitate interactions between two (or more) participating groups [13]. Smedlund and Faghankhani [4] noted traditional organizations easily control participants (employees) and the relationship between them, but platform owners have limited power and ability to fully control platforms as there are multiple parties contributing, deriving and using data. Governance for PEs thus should deliberate the different business context and concepts. Trust, roles, revenue sharing and control are identified as fundamental governance concepts for organizations which run a PE [9, 14, 15]. Those concepts should be implemented in data governance of PEs to encourage good practices of governance and to create value in the use of data [12]. Prior studies on platform governance largely neglect the role of data and data governance, and therefore data governance studies have been rarely found.

In our previous studies [10], we surveyed four platform companies (Facebook, YouTube, EBay and Uber) to show the state of practice of data governance. It revealed the fact that the policies of the platforms are imprecise in terms of data ownership and data usage. It can cause uncertainties and arguments between participating groups and business risks. We also reviewed 19 existing industry governance frameworks and academic works to examine if the identified issues can be addressed by them. However, there are common missing considerations of how to clarify the rights of data owner or subject and how to achieve visibility and traceability in the use of data. Through the studies, we confirmed the need for a data governance framework for PEs to support the organizations' business success.

3 Methodology

This study was conducted through three steps (Fig. 1). The first step was carried out to understand overall PE environment including who participates in a platform, how data is used in the platforms and what data characteristics are identified (①). We analyzed five PEs: Facebook (social network), YouTube (content portal), EBay and Uber (exchange platform), and Data.gov.au (public platform). The survey on the platforms were conducted by examining their policies such as data polices, privacy policies or cookies polices. We carried out a literature review to complement and confirm the result of the analysis. The second step was to identify decision domains and governance principles (②). In the previous study [10], we identified seven decision domains through reviewing literature, industry governance frameworks and the state of practice of four platforms. In this study, we refined them and identified focal principles to support the decision domains. Based on the results of the two steps, we defined due processes for the implementation of data governance. To confirm the processes, we analyzed data breach cases (AOL and Facebook) and reviewed the relevant literature. In the last step, we compared 19 existing frameworks and academic works with our framework (③). We included some IT/information frameworks for this comparison because they generally contain data governance. We used the identified decision domains, principles and due processes as the comparison factors.

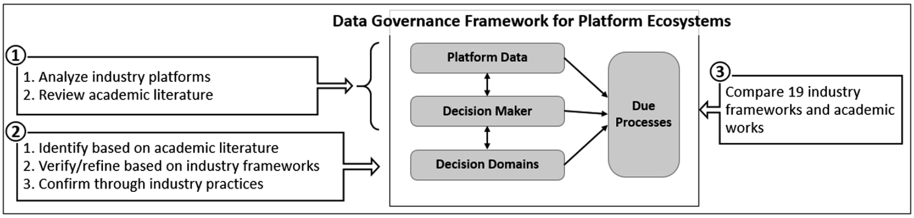

Data Governance Framework for Platform Ecosystems
①
1. Analyze industry platforms
2. Review academic literature
②
1. Identify based on academic literature
2. Verify/refine based on industry frameworks
3. Confirm through industry practices
Platform Data
Decision Maker
Decision Domains
Due Processes
③
Compare 19 industry frameworks and academic works

Fig. 1. The research methodology for a data governance framework for PEs

4 Data Governance Framework

4.1 The Conceptual Model

We start by presenting the general concept of data governance. There is a broad consensus among researchers that data governance must find answers to three questions: what decisions need to be made, which roles should be involved in the decision-making process and how the roles are involved in the process [11, 12]. Governance processes play a focal role to orchestrate and explain how the elements work together.

We develop a data governance framework for PEs based on the concept. It includes three core elements (data, decision makers, and decision domains) and due processes to support the elements. All the elements are discussed with a set of questions such as how data is collected and used in a PE, what decisions should be implemented about the data, and who can play the roles (Figure 2). The framework enables the governance body of a PE to find the answers of the questions by showing how platform data should be managed in data governance through due processes.

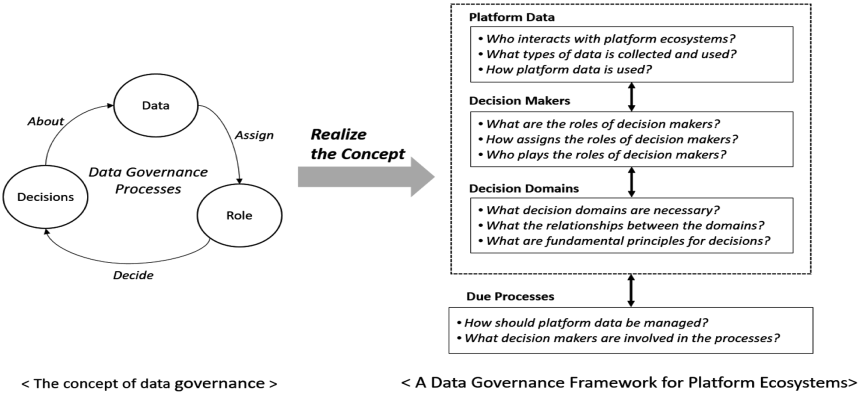

Data
About
Assign
Data Governance Processes
Decisions
Role
Decide
Realize the Concept
Platform Data
• Who interacts with platform ecosystems?
• What types of data is collected and used?
• How platform data is used?
Decision Makers
• What are the roles of decision makers?
• How assigns the roles of decision makers?
• Who plays the roles of decision makers?
Decision Domains
• What decision domains are necessary?
• What the relationships between the domains?
• What are fundamental principles for decisions?
Due Processes
• How should platform data be managed?
• What decision makers are involved in the processes?
< The concept of data governance >
< A Data Governance Framework for Platform Ecosystems>

Fig. 2. A data governance framework for PEs from general concept

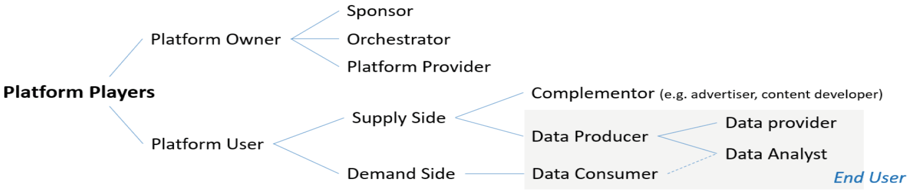


Fig. 3. Generic platform players in PEs

4.2 Platform Data

Ecosystem refers to a complex set of relationships among the elements of a given area. In a PE, data interacts with multiple participating groups. We begin with introducing platform player who interacts with a platform (Fig. 3), and then describe how platform data is collected and used, and discuss the current issues.

In general, platform player is divided into platform owner and platform user.

Platform owner consists of three roles: sponsor, orchestrator and provider. Platform sponsor owns a platform, and facilitates the co-creation of value from third-parties or establish an exchange platform he can benefit from [9]. The role of orchestrator is to organize a platform and the involved parties and processes. It is in charge of sharing standards, developing the industry vision or maintaining the integrity of a platform [4]. Platform provider is intermediary who delivers a platform. It generally includes the roles of data manager such as data collector, steward and custodian.

Platform user comprises supply and demand side. Data supply can be capable by complementor or data producer. Complementor contributes to a PE as an external party not directly related to the platform owner [9]. It offers a complementary content to the core component of the platform. Data producer consists of data provider who directly contributes data and data analyst who uses and provides data through data analytics jobs. On the demand side, data consumer refers to end user that uses platform data. Data analyst can be both data consumer and data provider if providing the outcome of the analytics jobs to the platform again. End user here is a person who accesses the platform to consume a service available on the platform [9, 15].

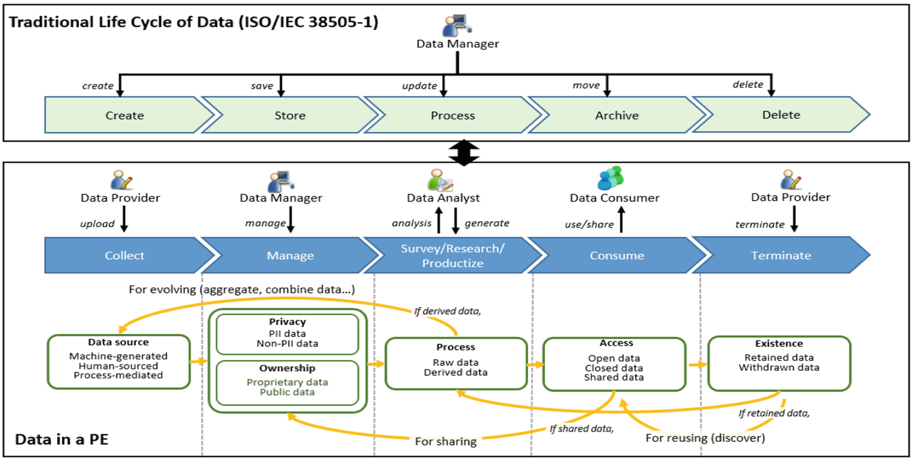


Fig. 4. Life cycle, characteristics and interactions with actors of data

The described roles can be changed over time or depending on platform strategy.

In traditional data governance, the life cycle of data is aligned with accountability for data management within an enterprise [16]. There is a simple interaction with a data manager. Meanwhile, the life cycle of platform data is based on the processes of data sharing. Platform data interacts with various platform players, and the flow map is characterized along the life cycle (Fig. 4).

Platform data is collected through providers' contribution such as uploading or generating new data. Majority data is from platform users as they upload their content such as video or image or provide user information (human-sourced data) [17]. While a user uses platform services, the platform systems may leave some data such as logs, search keywords, location (machine-generated data). This type of data is generally referred to service use information. Data is also collected through system processes such as transactions, reference tables, and relationships, as well as the metadata setting the context (process-mediated data).

The gathered data has to be examined whether it is Personal Identifiable Information (PII) data or non-PII data, and it is proprietary data or public data. Based on that, the management processes of data and the involved roles should be differed. PII is defined by Australian Acts as “Information or an opinion about an individual, or an individual who is reasonably identifiable, whether the information or opinion is true or not; and whether the information or opinion is recorded in a material form or not” [18]. However, PII and non-PII are not immutable [19]. If there are only single instances of users, it is easier to be identified by combination of the characteristics revealed in the datasets. To reduce the risk, continuous review process is necessary in the data management process. Due processes allow only expected activities of actors, and support identifying audit trails, offering interactive modeling and supporting user objections [8]. Proprietary data is claimed ownership by a specific entity or company. The owner of data should have certain decision rights and obligation about the data. In contrast, public data (e.g. crime data) is available for the public to collect or look at. As mentioned, the ownership of data is not clearly defined in platform policies. Our survey on four industry platforms [10] showed that Facebook, YouTube and Uber define data ownership of user content, but non-user content is rarely addressed. EBay documents overlook data ownership. To cope with the issue, platform owners have to consider the regulatory environment of a platform and determine an appropriate governance configuration prior to the use of data. It can reduce the risk of data misuse or abuse and protect the rights and privacy of the owner or subject of data.

The stored data in the platform systems can be internally used for their business to get useful information such as trends, statistics, significant keywords, or personal interests of users. The data can be evolved by aggregating or combining raw data, and then generates new data as derived data. 11 common use cases of platform data are identified through survey on industry platforms: provide, improve and develop (test) services, communicate with platform users, or show and measure ads and services. The cases, however, are not detailed documented in the policies of platforms. The data used for each use case is not precisely mentioned. It can result in data misuse. The issue is claimed by a number of researchers [8]. Another risk can be found when the stored data is used outside of a platform for survey or research by external partners. The exporting data should be reviewed by an appropriate policies if it can violate ownership or privacy rules or includes PII data. AOL and Facebook data breach (2006 and 2008) are reported as representative cases which the processes were ignored [1, 2].

General platform users can use platform data if data is set to “open” to everyone (open data) or a specific group or person (shared data). If a user changes the mode of his/her data into “private” (closed data), no one can use/access the data. Facebook documents that when a user post on Facebook, the user can select the audience for the post, such as a customized group of individuals, all of his or her friends, or members of a group. The platform mentions that open data is available to anyone on or off

Facebook services and can be seen or accessed through any online and offline media including search engines and TV. Open or shared data can encounter more risks of data abuse, misuse or privacy violation if there is no strong regulation or any complemental consideration. How to use, share or sell data without losing control is a critical issue of PEs [7]. Accordingly, data use cases should be explicitly defined in data governance as monitoring the use of data have to be implemented based on the use cases. A data supply chain also should be recorded to trace the derivation history of the open or shared data transparently. Such governance mechanisms should be fairly applied, and the processes and performance must be transparently shown to every participating group (in particular, to data owner and subject). Yet, the requirements are poorly implemented in industry PEs [10]. There are also claimed issues of an invisible data supply chain by researchers [8].

There is a broad consensus that data provider must have the privilege to stop sharing his data at any time. PEs provide several ways to change the mode of data sharing. A platform user can change the mode of data into private to stop sharing. The other way is to delete the data. Alternatively, the user can delete or deactivate the account. In theory, it looks as if data owners can perfectly control their data. However, in some cases, data owners lose control over their data. For instance, Facebook policies note that "information that others have shared about you is not part of your account and will not be deleted when you delete your account". That is, the shared data will be retained in the platform in the state that the owner is out of control, and continuously used/discovered by others. This issue has been discussed by researchers [20, 21]. Platform users' need for data transparency is increasing to access information which they are involved in. In addition, a certain method should be available to them for appropriate notice, consent and security.

4.3 Decision Makers

A typical data governance structure is organized within an enterprise. There is a lack of concerns about external users. In PE business context, platform users provide data, and it results in adding value to the platforms. Accordingly, they have a critical role in data governance of PEs [4]. In this sense, how to share the roles of decision making about data assets with platform users becomes an important issue [22]. We identify the key roles of decision makers for PEs including platform users: data committee/council, data manager, data owner and data subject.

Data committee/council is one of the role which is responsible for clarifying the role of data in PEs [11]. It makes decisions about the purpose of data use, desirable behaviors, and the appropriate governance mechanisms of a PE aligning the business goals. The role is generally taken by platform orchestrator.

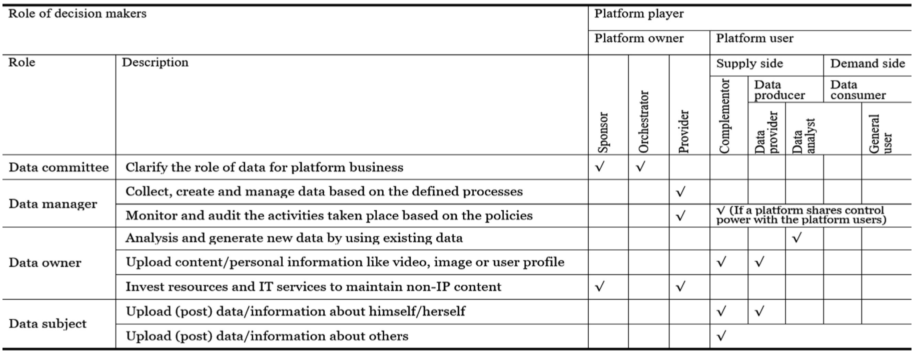
Data manager here refers to the role of internal data management in platform companies including data collector, data steward and data custodian [11, 23]. They are responsible for the implementation of data management tasks and the conformance to governance rules in platforms. Data governance design can be categorized into centralized and decentralized [12]. While centralized governance means a platform owner takes all the control, decentralized governance shares it with platform users. Therefore, some parts of the role can be implemented by the users in decentralized governance.

Data owner is an individual (or a company) who owns data by contributing it to the platforms. Data owner has ownership rights which refer to the questions of who is allowed to use data and who has decision rights [24]. Accountability of a data owner is noted as a form of verifiability in some literature. The term verifiability represents a sort of responsibility of the one who can verify data and confirm the veracity of the data before using or sharing the data [20]. Data owners should have data transparency and auditability, and access control power [21]. That is, every user has complete transparency over what data is being collected about her and how the data is used.

Data subject means a person who is the subject of personal data. If data is about a specific person, then that person can be a data subject. Data owner can be a data subject and vice versa if he/she uploads/generates data about him/her. There is an example to explain the difference between data owner and data subject in a simple way. A medical record of a patient is generated by a doctor/hospital. The owner of the record is the doctor/hospital that generated the record. The patient is a data subject because the medical record is about him, but he cannot own the record. Like a data owner, a data subject should have rights to access the data which he is involved in and a method available to him to hold data governance mechanisms accountable for appropriate notice, consent and security.

The described roles can be taken by various platform players depending on the platform strategies (Table 1). In decentralized data governance, platform users can monitor or audit the use of data or data integrity based on enabling technologies [25].

Table 1. The roles of decision makers and platform players



Role of decision makers		Platform player							
		Platform owner			Platform user				
Role	Description	Sponsor	Orchestrator	Provider	Supply side			Demand side	
					Complementor	Data producer		Data consumer	
						Data provider	Data analyst		General user
Data committee	Clarify the role of data for platform business	√	√						
Data manager	Collect, create and manage data based on the defined processes			√					
	Monitor and audit the activities taken place based on the policies			√	√ (If a platform shares control power with the platform users)				
Data owner	Analysis and generate new data by using existing data						√		
	Upload content/personal information like video, image or user profile				√	√			
	Invest resources and IT services to maintain non-IP content	√		√					
Data subject	Upload (post) data/information about himself/herself				√	√			
	Upload (post) data/information about others				√				

4.4 Decisions Domains

Decision domains refer to data governance areas which should be controlled to achieve the business goals of a PE. In our previous study [10], we identified seven data governance factors which can be used as decision domains for PEs (Table 2).

The decision domains are interconnected with each other. When platform data is used, there should be clear definition of the roles about the data such as who has accessibility or accountability, and who should be informed or consulted. A data ownership and access definition are thus regarded as a major concept when designing the business process of a PE [9, 22]. The definition should include user content and non-user content together to protect all the data and owners or subjects' rights against unauthorized use. To support this, a data ownership decision model should be developed by considering relevant regulations, laws or court cases [11]. For example, creativity (creative data: videos/non-creative data: factual data), originality (original data/derived data), investment (data managed by a platform owner or not) and source (from outside or created inside of a platform) of data can be the aspects of the model. They are derived from the review of regulatory environment such as Berne Convention and its derivatives, European Court of Justice (ECJ) in 2004 (William Hill case [26]) and the policy of platforms. Looking at the regulatory environment of a platform also supports accurately identifying and rewarding the contributors of a platform as it clarifies who adds value to the platform. The main role of data committee is to build policies for a platform based on the review on regulatory environment. The policies have to include all the considerations of how to use data, what data can be open (or not), how to share data or how to terminate data sharing.

Table 2. Data governance decision domains for PEs

Decision domain	Definition
Data ownership/access	Definition of who owns, uses and accesses platform data
Regulatory environment	Regulations, laws or court cases that could affect the use of data
Contribution measurement	Mechanisms to measure contribution against value creation to a PE
Data use case	The purpose of the collected data by a PE (how to use data)
Conformance	An audit for compliance based on strict processes and rules
Monitoring	Mechanisms to monitor the use of data (all activities related to data)
Data provenance	Means to trace the derivation history of the data transparently

When data is collected and used by a platform (platform users), if there is an only single owner, contribution measurement is simple. Meanwhile, using derived data (aggregated or transformed data) can lead to measurement issues because the data may contain a complicated ownership structure. Data provenance management can help this issue. It allows a platform to identify all the associated stakeholders and explicitly measure the contribution of each owner of the data by preserving all the record of the use of data. It also supports high visibility of the use of data [11, 27].

As stated, the purpose of data uses and the relevant data are not clearly defined in the policies of PEs. The documentation is not enough to understand how the collected data is used. All the collected data should be categorized and has a clear and limited purpose of the use of data. It enables a platform to detect and prevent unexpected use of data in a data supply chain [16].

Monitoring and conformance mechanisms facilitate visible/reliable data use. There are many data breaches caused by an invisible supply chain and unclear due processes

[1, 2]. To increase transparency of a platform and thus gain more trust from platform users, a platform owner can share control power and decision rights with the users through decentralized data governance [25].

To support right decisions, data governance should be implemented based on key principles which present sets of applicable guidelines and considerations. Through a literature review and survey on industry platforms, the following four principles are identified, which have been regarded as fundamental considerations.

Table 3. The influence of the data governance principles on the decision domains

Decision domain	Principle 1 Align with platform governance concepts	Principle 2 Meet the needs of all participating groups	Principle 3 Address all types of data	Principle 4 Consider platform context
Data ownership, access definition	Define clear roles and responsibilities	Consider all data contributors' needs and rights	Clarify ownership and access rights to all types of data	Apply different levels of governance control based on the context of a platform * Highly regulated environment, high quality of data strategy, closed platform strategy or authorized-based governance configuration - > use strict data ownership, access control, audit and monitoring by a centralized (internal) structure * In the opposite case - > share the control power with platform users and use trust-based control
Regulatory environment	Identify what regulations should conform to control	Develop a decision model for explicit data ownership	Consider extensive regulations for non-user content	
Contribution measurement	Consider a revenue sharing concept	Identify different types of contribution of participants	Measure every data contribution based on regulations	
Data use case	Build trust through a visible data supply chain	Consider how to use data without losing control	Provide a detailed data category and use cases	
Conformance	Conform governance rules through a regular audit	Involve various participating groups	Audit every data use case and its processes	
Monitoring	Control an unauthorized data use	Provide possible opportunities to all stakeholders	Make a visible supply chain for all data use activities	
Data provenance	Support efficient, effective control and clear roles	Enable data owners and subjects to trace the history of the use of data	Record all the use of data including sharing, analyzing and transforming	

Principle 1. Align with platform governance concepts and business goals. Data governance goals can be identified and determined by looking at what to maximize the value of data and a PE. The goals, thus, should align the business goals and higher-level governance like platform governance [11, 28]. The characteristics of a platform also can be considered when confirming data governance goals. If platform open strategy leans to close, the data governance should be toward the focus on strict due processes and input/output control mechanisms [25].

Principle 2. Meet the needs of all participating groups. A PE faces the complicated relationships between multiple parties. Trust between platform owners and the parties is regarded as a prerequisite factor to win business [9, 14]. It can be built by starting with a good understanding of what governance practices are applicable and how they work, and share value (management strategies of a platform). Accordingly, data governance should be designed and implemented from all the perspectives of parties.

Principle 3. Address all types of data. Data governance should be able to control all types of data in platforms. As mentioned, platform data is collected from various source. Yet, PEs are mainly focused on user content [10]. The other types of data are often ignored and thus do not addressed in data governance processes. It leads to unclear data ownership or access rights of data owner or subject.

Principle 4. Consider platform context; one size does not fit all. Platforms have to consider different business strategies and goals, and consider different levels of market regulation. According to contingency theory, such different contingencies affect data governance [12]. In the previous study, we examined the influence of specific platform contingencies on the characteristics of a platform and a data governance design [25]. This principle gives the idea that data governance can be flexible based on the context of a platform and tailored for practical implementation.

The principles affect the decision domains in a certain way. They help a platform to focus on the key considerations and ultimately enable a platform to win business. Table 3 shows how the principles are applied to the decisions domains.

4.5 Due Processes in the Use of Data

Due process is regarded as a pivotal control mechanism to mitigate a risk of data abuse or misuse as it forces desirable behavior of participants [8].

In this section, we suggest comprehensive due processes in the use of data to show how platform data should be managed in data governance. All the considerations and discussions in the previous sections are deliberated in the processes. The processes also illustrate how the roles described previously are involved and how the decision domains are implemented in data governance. This section sequentially demonstrates the processes by following the lifecycle of platform data (Fig. 4).

Data Collection Process (Fig. 5). Data collection is implemented by defining data categories and data use cases based on the principles and policies of a platform. When data is collected by a platform, it should be classified by the defined data category. The use of the data thus can be limited by the predefined and linked use cases to prevent illegal use of data (❶). The data is also characterized based on the source. It can be used

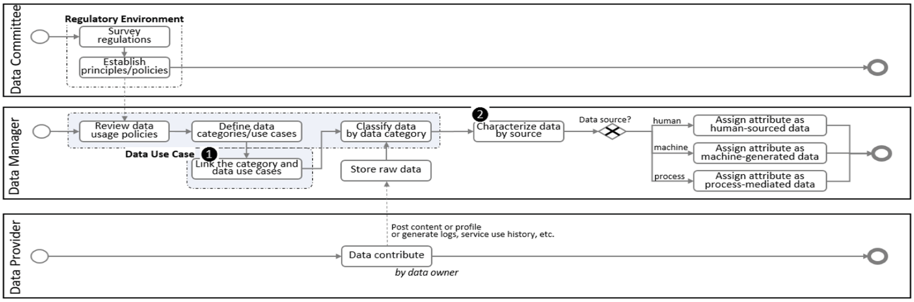

Data Committee
Regulatory Environment
Survey regulations
Establish principles/policies
Data Manager
Review data usage policies
Define data categories/use cases
Classify data by data category
Data Use Case
1
Link the category and data use cases
Store raw data
2
Characterize data by source
Data source?
human
machine
process
Assign attribute as human-sourced data
Assign attribute as machine-generated data
Assign attribute as process-mediated data
Data Provider
Post content or profile or generate logs, service use history, etc.
Data contribute
by data owner

Fig. 5. Data collection process

for data ownership recognition (❷). In general, human-sourced data is regarded as Intellectual Property (IP) content, and it belongs to the provider.

Data Management Process (Fig. 6). The main focus of this process is on the rights protection of data owner and subject. Data committee has to establish clear policies in terms of data ownership, access right and privacy based on the relevant regulations or laws. For privacy protection, every stored data should be tested (PII or non-PII data). According to the result, the data needs to be dealt by the different levels of processes and policies (❶). A model for ownership definition is developed by following the defined policies of a platform (❷). It is used when identifying the owner of data to measure contribution and assign explicit roles and responsibilities. The information should be sent to the owner and subject of the data (❸). Data provenance can be initialized for recording historical information of the data use (❹).

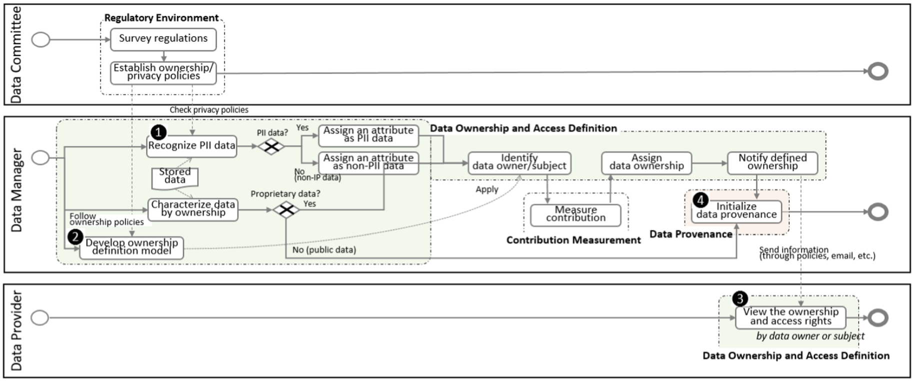

Data Committee
Regulatory Environment
Survey regulations
Establish ownership/privacy policies
Check privacy policies
Data Manager
1
Recognize PII data
PII data?
Yes
No (non-IP data)
Assign an attribute as PII data
Assign an attribute as non-PII data
Stored data
Follow ownership policies
Characterize data by ownership
Proprietary data?
Yes
No (public data)
2
Develop ownership definition model
Data Ownership and Access Definition
Identify data owner/subject
Apply
Measure contribution
Contribution Measurement
Assign data ownership
Notify defined ownership
4
Initialize data provenance
Data Provenance
Send information (through policies, email, etc.)
Data Provider
3
View the ownership and access rights
by data owner or subject
Data Ownership and Access Definition

Fig. 6. Data management process

Survey, Research and Productization Process (Fig. 7). Platform data can be used for improving the services of a platform company. In addition, it can be required for external use such as research purpose. In those cases, first of all, every access should be confirmed if it is legal and the purpose of the use meets the predefined use cases of the

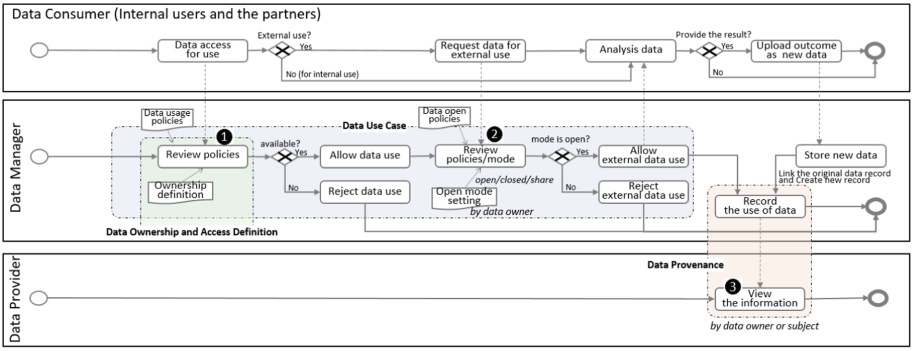


Fig. 7. Data survey, research and productization process.

data (❶). Secondly, if the data is taken out and possibly disseminated for secondary use, the openness of the data and platform policies must be checked (❷). Facebook data breach happened as the company overlooked the process [2]. A group of personal information was exported for a research project without a review process, and quickly diffused for secondary use. It resulted in revealing the data to public without consent of the data owners and subjects. Lastly, the data owners and subjects should transparently know all the information of the use of their data to support user objections (❸).

Data Consumption Process (Fig. 8). The open or shared data in a platform can be discovered and used by other users (❶). Like the previous process (Fig. 7), all the processes should follow the relevant policies and be reviewed. This process pays more attention to high participation of platform users and transparency of a platform. When a platform company shares control with platform users (in decentralized data governance), platform users can actively participate in auditing or monitoring data and data use processes (❷❸). It is made possible by enabling technologies such as blockchain which is regarded as one of the most innovative and revolutionary governance forms [29]. This process enables an organization to reduce cost and effort, and gain more trust between a platform owner and the platform users [25].

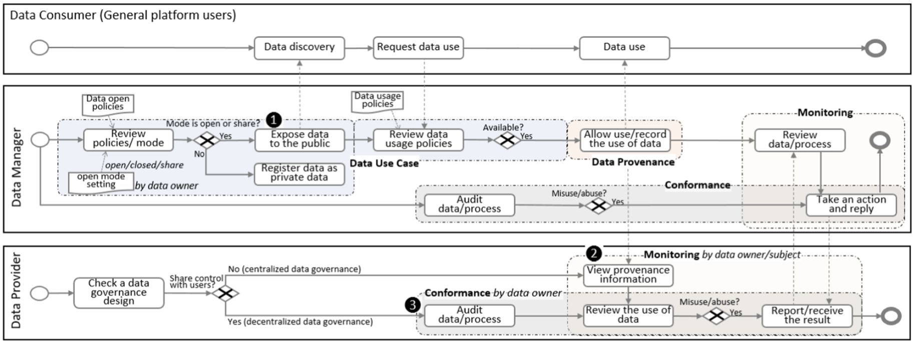


Fig. 8. Data consumption process

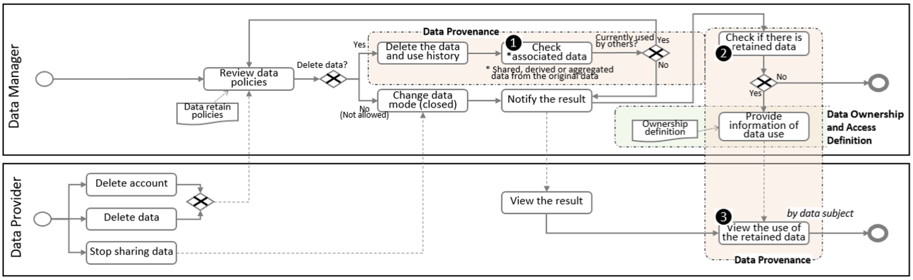

Data Manager
Data Provider
Review data policies
Data retain policies
Delete data?
Yes
No
(Not allowed)
Data Provenance
Delete the data and use history
1
Check *associated data
* Shared, derived or aggregated data from the original data
Currently used by others?
Yes
No
Change data mode (closed)
Notify the result
Check if there is retained data
2
No
Yes
Provide information of data use
Ownership definition
Data Ownership and Access Definition
Delete account
Delete data
Stop sharing data
View the result
3
by data subject
View the use of the retained data
Data Provenance

Fig. 9. Data sharing termination process

Data Termination Process (Fig. 9). When a user deletes his account or content, the content may be deleted from the platform systems. Depending on the data retention policies of a platform, the deleted data can be retained for a certain period of time, but it is ultimately deleted. However, the shared or derived data of the user can be retained and out of control of the owner (❶). Accordingly, if data is retained based on the policies of a platform, even though the owner lost the ownership, the rights of the data subject to the data should be protected and respected (❷). In this sense, the information of the use of the data must be accessible by the subject (❸).

5 Comparison

We compare 19 governance frameworks and academic works with our framework which were used in the previous study for the state of the art (Table 4). ISO/IEC 38500 is replaced with 38505-1 as it has been released as a data governance standard.

We use the principles (P), decisions (D) and due processes (DP) presented in this study as comparison factors to evaluate the comprehensive concerns and impact of the compared frameworks and studies (P and DP are added to the previous analysis). "Sufficiency" is used to examine if the factors are dealt in the frameworks [10]. A result is determined as "not covered (x)", "partially covered (○)", or covered (●)".

The result reveals three main findings. Firstly, any framework or study that covers all the considerations addressed in our framework have not been found. Most of IT/data governance frameworks focus on general roles and responsibility (D1) or role definition and control by governance bodies (D1, 5 and 6). Secondly, platform studies pay more attention to the concept of PE, and platform control mechanisms (D5 and 6). As the studies are still at a relatively embryonic stage, how to manage data is largely neglected. Lastly, while the importance of governing process is stressed by most of the frameworks, due process has not been suggested in any framework. COBIT 5.0 documents governance processes, practices and activities but how organizations implement the processes and what roles should be involved in the processes are not described. It may lead that organizations have difficulties in newly applying or improving data governance in practice.

We confirm that there are significant gaps between the compared frameworks and our suggestion which should be filled. It shows the need for our framework again.

Table 4. The result of comparison of governance frameworks

Category	IT Governance			Information/Data Governance						Governance for Platform Ecosystems							
	COBIT 5.0	Weill & Woodham (2002)	Weill & Ross (2004,2005)	ISO/IEC 38505-1	DGI Framework	Informatica Framework	IBM Information Governance	Khatri & Brown (2010)	Weber et al. / Otto & Weber (2009, 2015)	Evans (2012)	Ghazawneh & Henfridsson (2010, 2013)	Hagiu (2014)	Manner et al. (2012)	Manner et al. (2013)	Tiwana et al. (2010)	Tiwana (2013)	Our Framework
P1: Align platform governance	●	X	○	○	X	X	●	●	X	X	X	X	X	○	X	X	●
P2: Meet the needs of all participants	●	X	○	○	X	○	X	○	X	X	X	X	X	X	X	○	●
P3: Address all types of data	X	X	X	X	X	○	X	X	X	X	X	X	X	X	X	X	●
P4: Consider platform context	X	○	X	X	○	X	X	X	●	X	X	X	○	○	X	X	●
D1: Data ownership and access definition	○	○	○	○	●	●	○	○	○	X	X	X	X	X	○	○	●
D2: Regulatory environment	○	○	○	●	X	X	○	○	X	X	X	X	X	○	X	X	●
D3: Contribution measurement	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	●
D4: Data use case	X	○	○	○	○	X	X	○	●	X	X	X	X	X	X	X	●
D5: Conformance	●	X	X	●	X	●	○	○	X	○	○	X	○	○	○	○	●
D6: Monitoring	○	○	X	●	○	●	○	○	X	○	X	X	○	○	○	○	●
D7: Data provenance	X	X	X	X	X	●	○	○	X	X	X	X	X	X	X	X	●
DP: Collection, management, analysis, consumption and termination process	○	X	○	X	○	○	X	X	X	X	X	X	○	X	X	X	●

6 Conclusion and Future Work

Many organizations today adopt or consider PE for their business innovations. The concept of PE supports sustainable growth through network effects where there is multiple groups' data contribution. However, lack of organizational capability to orchestrate complicated context, processes and relationships occurred among the parties will lead to market failure. Traditional data governance focuses on in-house control of data, and prior research on platform a governance is still in its infancy.

In this study, we proposed a new data governance framework which supports an organization to mitigate business risks from the complexity of a platform and add value to the organization. We surveyed industry platforms and reviewed governance frameworks and literature. This study delivered the idea on how data should be managed when an organization adopts the concept of platform ecosystem. In particular, through the due processes, we demonstrated how organizations can implement data governance and orchestrate all the considerations of platform ecosystem. We compared the framework with 19 existing industry governance frameworks and academic works. The comparison showed that there is no existing framework or study which covers all the aspects of our suggestion in the framework.

In the next step, we will provide the use cases of the framework to assist an organization to implement data governance in practice. We will identify use case scenarios and the associated governance questions for decision-making which are critical but cannot be answered by current governance frameworks. To this end, we will perform an extensive literature review and survey industry needs.

References

1. Ives, B., Krotov, V.: Anything you search can be used against you in a court of law: data mining in search archives. Commun. Assoc. Inf. Syst. **18**(1), 29 (2006)
2. Zimmer, M.: "But the data is already public": on the ethics of research in Facebook. Ethics Inf. Technol. **12**(4), 313–325 (2010)
3. Parker, G., Van Alstyne, M.W.: Platform strategy. In: The Palgrave Encyclopedia of Strategic Management (2014)
4. Smedlund, A., Faghankhani, H.: Platform orchestration for efficiency, development, and innovation. In: 2015 48th Hawaii International Conference on System Sciences, pp. 1380–1388 (2015)
5. Kaisler, S., Armour, F., Espinosa, J.A., Money, W.: Big data: issues and challenges moving forward. In: 2013 46th Hawaii International Conference on System Sciences, pp. 995–1004 (2013)
6. Kaisler, S., Money, W.H., Cohen, S.J.: A decision framework for cloud computing. In: 2012 45th Hawaii International Conference on System Sciences, pp. 1553–1562 (2012)
7. Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., Shahabi, C.: Big data and its technical challenges. Commun. ACM **57**(7), 86–94 (2014)
8. Martin, K.E.: Ethical issues in the Big Data industry. MIS Q. Executive **14**, 2 (2015)
9. Schreieck, M., Wiesche, M., Krcmar, H.: Design and governance of PEs–Key concepts and issues for future research. In: 24th ECIS 2016 (2016)
10. Lee, S.U., Zhu, L., Jeffery, R.: Data governance for PEs: critical factors and state of the practice. In: 21st PACIS 2017, Malaysia (2017)
11. Khatri, V., Brown, C.V.: Designing data governance. Commun. ACM **53**(1), 148–152 (2010)
12. Weber, K., Otto, B., Österle, H.: One size does not fit all—a contingency approach to data governance. J. Data Inf. Qual. **1**(1), 1–27 (2009)
13. Evans, D.S.: Governing Bad Behavior by Users of Multi-Sided Platforms, SSRN Scholarly Paper No. ID 1950474. Social Science Research Network, Rochester, NY (2012)
14. Hein, A., Schreieck, M., Wiesche, M., Krcmar, H.: Multiple-case analysis on governance mechanisms of multi-sided platforms. In: Multikonferenz Wirtschaftsinformatik (2016)
15. Tiwana, A., Konsynski, B., Bush, A.A.: Platform evolution: coevolution of platform architecture, governance, and environmental dynamics. Inf. Syst. Res. **21**(4), 675–687 (2010)
16. ISO. https://www.iso.org/standard/56639.html. Accessed 27 Sept 2017
17. Firmani, D., Mecella, M., Scannapieco, M., Batini, C.: On the meaningfulness of 'Big Data Quality' (Invited Paper). Data Sci. Eng. **1**(1), 6–20 (2016)
18. Australian Government. https://www.legislation.gov.au/Details/C2018C00034. Accessed 3 Oct 2017
19. Schwartz, P.M., Solove, D.J.: The PII problem: privacy and a new concept of personally identifiable information. NYUL Rev. **86**, 1814 (2011)
20. Al-Khouri, A.M.: Data ownership: who owns "my data". Int. J. Manag. Inf. Technol. **2**(1), 1–8 (2012)
21. Zyskind, G., Nathan, O.: Decentralizing privacy: Using blockchain to protect personal data. In: Security and Privacy Workshops (SPW). IEEE (2015)
22. Tiwana, A.: PEs: Aligning Architecture, Governance, and Strategy, Newnes (2013)
23. Cheong, L. K., Chang, V.: The need for data governance: a case study. In: ACIS 2007 Proceedings, vol. 100 (2007)

24. Eckartz, S.M., Hofman, W.J., Van Veenstra, A.F.: A decision model for data sharing. In: Janssen, M., Scholl, H.J., Wimmer, M.A., Bannister, F. (eds.) EGOV 2014. LNCS, vol. 8653, pp. 253–264. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44426-9_21
25. Lee, S.U., Zhu, L., Jeffery, R.: Designing data governance in PEs. In: 2018 the 51st HICSS, Hawaii, (2018)
26. Harison, E.: Who owns enterprise information? Data ownership rights in Europe and the US. Inf. Manag. **47**(2), 102–108 (2010)
27. Informatica, https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/white-paper/holistic-data-governance-framework_white-paper_2297.pdf. Accessed 14 Sept 2016
28. Weill, P., Ross, J.W.: IT Governance on One Page, SSRN Scholarly Paper No. ID 664612, Social Science Research Network, Rochester, NY (2004)
29. Parker, G., Van Alstyne, M.W., Choudary, S.P.: Platform Revolution: How Networked Markets are Transforming the Economy and How To Make Them Work for You (2016)