Data Governance Decisions for Platform Ecosystems

Sung Une Lee
[1]Data61, CSIRO
[2]University of New South Wales
Sungune.Lee@data61.csiro.au

Liming Zhu
[1]Data61, CSIRO
[2]University of New South Wales
Liming.Zhu@data61.csiro.au

Ross Jeffery
[1]Data61, CSIRO
[2]University of New South Wales
Ross.Jeffery@data61.csiro.au

Abstract

Platform ecosystem has become an information system research subject after many years of industry success. The concept of platform ecosystem facilitates fast and self-growing of a platform by encouraging data contribution/consumption of multiple networks, and thus the importance and value of data in platforms is accentuated. It is essential to understand how data should be managed in platform ecosystems where there is complicated relationships between multiple participating groups. However, this topic has been rarely addressed in industry and academia. Industry governance frameworks focus on organizational data, and prior research on platform ecosystem is still in early-stage. To response to the limitation, we propose critical data governance decisions for platform ecosystems, and discuss how they have to be implemented in practice. This study supports right decision making about data, and facilitates a secure platform ecosystem. We perform a case study to illustrate the practical implications of this study.

1. Introduction

The Facebook-Cambridge Analytica data scandal today is one of the hottest topics in the IT press. A number of news articles report that this scandal affects the share prices and reputation of Facebook. It raises public awareness of the business risks caused by data abuse or misuse. This concern has been highlighted for some time in both academia and industry.

A platform ecosystem (PE) can reach critical mass by data contribution from multiple external parties [1]. The collected data is analyzed or shared to add value to the PE, and used by the platform owner, partners or family companies and users. Such complicated interactions between multiple parties providing, using or sharing data may arise data abuse or misuse. PEs need to impose certain regulations to mitigate risks resulting from the use of data by multiple parties [2].

Data governance refers to comprehensive control, including processes, policies and structures about data assets. Data governance for PEs has to orchestrate the complicated processes and relationships affected by multiple parties' participation [3]. Lack of or poor implementation of data governance can lead to unclear ownership and access rights of data contributors and invisible use of data [4]. Existing governance frameworks deal with general concerns for an enterprise where there is simpler and clearer data ownership and limited use of data. Those concerns have been articulated by a number of studies [5-8]. However, prior studies have been less focused on data and data governance in PEs [9], and there is a lack of an understanding how data governance should be managed such as what are the impact area of data governance decisions for a viable and sustainable PE. In the previous study [5], data governance factors for PEs are identified. We here focus on what decisions should be made and how they should be implemented for practical data governance based on the factors. The decisions and practices can be used by practitioners when they improve existing data governance or design new one. For researchers, this paper delivers broad information and knowledge of PE and data governance. Through a case study, we validate the theoretical concepts discussed in this paper. We identify how the theoretically important governance decisions are addressed in the real world, and illustrate the practical implications of this study.

Next section provides broad information to support understanding of PE and data governance. Section 3 describes the methodology of this study. We then discuss data governance decisions and management practices. The result of a case study is presented in section 5. We conclude this study in section 6.

2. Background

There are multiple types of governance such as IT/information/data governance. IT governance supports right decision making about IT assets to ensure IT investments support business objectives, but data governance focuses on data assets [10]. The term information governance is often used in the same sense

URI: https://hdl.handle.net/10125/60072
ISBN: 978-0-9981331-2-6

as data governance by some authors [11], but it addresses information issues rather than individual data pieces [12]. IT/information governance, however, often includes data governance [11]. Thus, data governance should align with the goals and concepts of higher-level governance [10]. A goal cascading mechanism in industry governance frameworks shows that stakeholder's needs, enterprise goals, IT-related goals and information/data level goals must be aligned [13].

A PE is defined a platform which constitutes two or more sided networks transacting with each other [3]. It allows interactions between multiple groups by providing a meeting place [14]. It is regarded as the building blocks that act as a foundation upon which an array of firms can develop complementary products, technologies or services [15]. For example, YouTube has a group which provides videos. The other group watches the videos. The groups facilitate various benefits and grow by providing data by themselves [1]. Every PE collect data from the participating groups which contribute data such as content or non-content like logs, and uses/shares the collected data. The main purpose of the use of data can be different according to the platform type (e.g. content portal/social network), business purpose (e.g. commercial/non-commercial) or platform strategy. Facebook uses the collected data for the business and reap the benefits of ecosystem growth such as high revenue, but Apple does not use user data for commercial purpose. Nonetheless, both (all PEs) use user data for service/product improvement, service use analysis and communication with users. While traditional organizations easily control participants (employees) and the relationship between them, platform owners have limited power to fully control platforms as there are multiple parties contributing, deriving and using data [3]. It can result in losing control of the use of data (data abuse/misuse), lawsuit by disgruntled users and low quality of data [16]. There are data breach cases of Facebook and AOL.

Facebook-Cambridge Analytica scandal [17], was publically uncovered this year. It is reported that 50 million user profiles are shared (sold) and used without permission. A similar case has been found in 2008 [18]. One research project team collected 1,700 user profiles from Facebook and then publically opened the data. The source of data could be quickly identified. An AOL case occurred in 2006. AOL published the search log data of users to the public, and the data was identified as Personal Identifiable Information (PII) data soon after the revelation [19, 20]. AOL didn't open any PII data. However, the log data was easily turned to PII data since it was categorized by user and the data provided lots of information of individuals. The three incidents remain some data governance issues such as unauthorized use of data and high ambiguity of control mechanisms in the use of data.

The current state of data governance of industry PEs is still immature [4]. There is a lack of consideration of various sources of data. PEs generally focus on user content, and thus there is a lack of clear definition of who owns or uses non-user content (e.g. logs or keywords). Data usage in the supply chain is also invisible to users. The policies of platforms are imprecise, and thus how, when, and who uses the data are not clear. This issue is claimed by researchers as one of the critical challenges [5, 8], which should be resolved for trust between platform owners and the users and business success [9, 11].

The findings and concerns are supported by prior studies. A number of studies address unclear data ownership [5-7], the importance of user contribution model [2, 21, 22] and invisible data usage [8] as challenges. However, how such concerns should be managed in data governance of PEs has received little attention in both industry and academia [9].

The results of analysis on 19 existing industry governance frameworks and academic works [10, 13, 14, 16, 23-33] shows that most of them address general roles and responsibility of stakeholders within an enterprise. It can lead to difficulties in newly applying or improving data governance in practice when there are multiple networks. Yet, prior studies pay more attention to the concept of PE and control mechanisms as they are still at a relatively embryonic stage. How to manage data is largely neglected, and the importance of visibility of a data supply chain is overlooked.

3. Methodology

This study used various data sources to identify scientifically important aspects and grounds, and the practical implications of data governance for PEs. We conducted a literature review, survey of existing governance frameworks, industry PEs and data breach incidents, and a case study on one industry PE.

3.1. Literature review and survey

For the literature review, we conducted keyword search using specific query and exchangeable keywords [31]. As the keywords, "platform ecosystem", "multi-sided platform" or "two-sided platform" and "data governance" or "management" were used. We included literature which addresses platform governance, the characteristics of PE, or role of data in PE, to get broad information and knowledge. We then drilled down to specific interests based on the result of the first step of a literature review. We used

"ownership", "access", "privacy", control", "conformance", "data breach", "monitor" and "provenance" for the detailed search.

Using the result of a literature review, we surveyed five main industry governance frameworks: COBIT 5.0, ISO/IEC 38505-1, DGI framework, Informatica framework and IBM information governance. We also surveyed PEs to identify how governance practices are implemented and what practices are overlooked in the real world. Four commercial PEs (Facebook, YouTube, EBay and Uber) and two non-commercial PEs (RIBIT: Australian platform and SW bank: Korean platform) are included. We conducted the survey by analyzing the policies and websites, and reviewing academic papers or news articles. In our previous studies, we surveyed most the mentioned governance frameworks and PEs. In this study, we replaced ISO/IEC 38500 with 38505-1 (as the data governance standard has recently been released), added new platforms (the two non-commercial PEs), and used different lens to identify specific data governance decisions for PEs.

Three data breach cases (two Facebook cases and one AOL case) were analyzed by reviewing academic papers and news articles. We reviewed the cases from the point of view of data governance, and identified significant lessons learned which should be considered in data governance for PEs.

All the collected data were distinguished and categorized in the form of a table. The data was examined and crosschecked among the different data sources. Based on the refined data, we first identified fundamental principles which should be commonly considered in every data governance decision area. We then identified important governance decisions and practices which should be made and implemented for successful management of data in PEs.

3.2. Case study

A case study was conducted to validate the theoretical concepts we discuss in the next section [20], and illustrate the practical implementation and possible implications of this study. We selected Platform A which is currently running and managed by the government agency. We chose the platform as one of the authors of this paper used to work at the platform. We surveyed the platform to understand how the PE is addressing theoretically important governance decisions in reality: i.e. how and if the proposed decisions and practices are implemented in practice.

We used five sources of evidence to collect data from the case following Yin's principles [44]: documentation, archival records, interviews, direct observations, physical artifacts. We first analyzed the policies and websites with other documents. We then reviewed the collected data and validated them through interviews with the former and current managers of the platform. We got detailed information and opinions. To do so, we prepared ten open-ended questions based on the governance decision questions identified in this study (the section 4). The interviews were carried out through online channel (phone calls) because the interviewees are overseas.

We analyzed the collected data using the identified governance decisions and practices (four decisions domains and 13 practices). We classified and summarized the results of how the platform implements the data governance decisions. We used a simple metric (sufficiency) to test if the platform implements the proposed data governance decisions and practices. We used "not implemented/partially implemented/implemented" as follows.

Not implemented: no document and observed activity.

Partially implemented: found either document or activity, but implementation is not fully satisfied. E.g. there is defined use cases of data in policies, but what types of data are used for each purpose is not clear.

Implemented: either document or activity, and implementation is fully satisfied.

In the last step, we discussed the results and draw conclusions. We first presented how the platform implements the data governance decisions. We then identified the gaps between our discussion (theoretical considerations) and the practical implementation. We identified potential risks and opportunities based on the gaps. What effects different implementation causes was analyzed to understand the context of the case.

4. Data governance decisions for PEs

There is a broad consensus among researchers that data governance must find answer to the questions of what decisions need to be made and which roles and how the roles should be involved in decision-making process [10, 29]. In this study we concentrate on the first question to identify critical decisions.

4.1. Key principles for decision making

IT/data governance frameworks are generally built on fundamental principles which present sets of guidelines and considerations for all decisions [10, 13, 25, 26]. In traditional governance, the principles focus on generic goals and a universal approach to manage the data of an enterprise [29]. We pinpoint specific principles for a PE based on the characteristics of a PE. They serve as a starting point for designing new data governance or evolving legacy one. The first principle (4.1.1) supports to identify significant governance

decisions, and the other principles provide key considerations to implement the decisions.

4.1.1. Align with platform governance concepts and business goals. Data governance goals should align the business goals and higher-level governance goals/ concepts to maximize the value of a PE [10, 24]. The business goals influence the direction and design of data governance. If a PE aims to increase user satisfaction, it needs formal and strict control mechanisms to increase the quality of data [34]. Likewise, higher governance concepts affect data governance decisions. Roles, revenue sharing, trust and control are the key concepts of platform governance [9, 20, 32]. Roles in data governance refer to a form of data ownership with clear responsibility. It allows a PE to protect data and the rights of a data owner/subject. Revenue sharing concept gives the idea that a platform owner should consider a reward for data contributors. Trust is regarded as a prerequisite factor to success [9, 20, 35]. To improve trust, high transparency of the use of data is essential in data governance. Trust can be increased by sharing decision rights with platform users. Otherwise, rigorous control mechanisms have to be implemented by a platform owner, and the result or process of decision making must be open to all participating groups. Control has been addressed in literature as a vital factor for the successful use of data [1, 30-33]. It is related to the concerns of how to monitor and preserve the use of data and how to conform to data governance rules.

4.1.2. Consider all participating groups. In traditional data governance, there are simple and clear roles for data management such as create store, update, archive and delete [25]. Data governance of a PE needs to address complicated relationships and interactions between multiple parties. The participating groups of a PE consist of platform owner (including the roles of platform sponsor, orchestrator and provider) and platform user groups (supply side and demand side users). All the groups play critical roles in data governance of a PE. Governance policies thus should be equally applied to all parties to be fairly applicable rules for everyone [33]. Thus, every participant should be given the same opportunity and accessibility as it results in more participation and ideas. It ultimately leads to new innovation [36]. This principle enables a PE to develop realistic data governance which can be realized by starting with a good understanding of the needs of all participating groups. It allows a PE to share a data management strategy which should be delivered to all participants. If a PE needs more participation and trust, a platform owner can give users more chance to join the decision-making processes in certain ways. It helps a PE to design and implement data governance from all the perspectives of parties.

4.1.3. Cover all types of data. Platform data is collected from various source like human or systems. Industry PEs generally focus on user content [4]. The other types of data are often ignored in the decision making process of data governance. It can lead to ambiguous and incomplete governance decisions. PEs generally have a focus on privacy laws to protect Personal Identifiable Information (PII) data. However, PII and non-PII are not immutable [37]. Non-PII data can be PII data by combination of extra information (as shown in the AOL data breach case). The importance of non-user content thus must be highlighted for a secure platform. In addition to this, the value of non-user content increases because of advertising, the main source of the revenue of majority PEs. Non-user content like service use information (e.g. logs) is used for a targeted advertising by PEs. A targeted advertising mechanism shows how such data is used through invisible and hidden markets [38]. It grows worries of data abuse and privacy violation with ethical issues [8, 38]. To reduce the risks, data governance of PEs should take into account how to make a visible supply chain for all types of data in a PE.

4.1.4. Consider different platform context; one size does not fit all. Platforms have to consider different business strategies, goals and market regulation. Such different contingencies affect data governance [29]. This principle gives the idea that data governance decisions can be flexibly made based on the context of a platform and tailored for efficient implementation. For instance, Apple (app store) and Facebook show explicitly different governance decisions on the control mechanisms [20]. Apple aims at providing good quality services, and therefore it adopts tight control through manual reviews. In contrast, Facebook has loose control by allowing any input with no restrictions.

Governance decisions often result in serious consequences as shown in the Facebook-Analytica scandal. Since Facebook allows the apps to collect user data (even the friends' data) for higher market share and revenue, the risks of data misuse/abuse and privacy violation increased a lot. In contrary to this, Apple's policies do not allow the apps to collect user data, and restrict the use of user data for an advertising [39].

4.2. Decision domains

4.2.1. The architecture overview. Decision domains refer to data governance areas which should be controlled to achieve the business goals of a PE. In the

previous study [4], seven data governance factors are identified for PEs (Table 1). We transform them to decision domains by categorizing based on the similar characteristics and aspects (Figure 1). The first four factors in Table 1 are identified as the main decision domains as they are regarded as core to set governance policies and strategies. The rest of the factors are considered as subdomains since they generally support other decisions [10, 13, 27, 28]. The decisions domains are identified to specifically manage the complicated situation and relationship of a PE. Therefore we do not discuss here all the domains which can appear in a universal data governance framework.

Every decision needs to be made by harmonizing all the considerations and information of the decision domains [10]. As shown in Figure 1, the decision domains are tightly interrelated to support right decision making in alignment with the principles.

There is a common consensus in both industry and academia that the conceptual difference of governance and management should be considered [10, 13, 24, 25]. While governance means decisions which should be made to ensure effective management and the use of data, management means a set of practices for the implementation of the decisions. Based on this concept, we introduce core governance decisions for PEs and the separated management practices.

Table 1. Data governance factors for PEs [4]

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

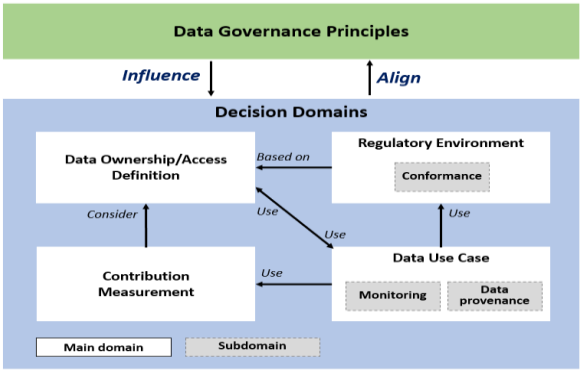

Figure 1. The data governance decision domains

4.2.2. Governance decisions. 1) Regulatory environment. The potential decisions of this domain are "*what regulations, specific policies, standards and guidelines should be considered?"* and *"how does the regulatory environment influence the uses of data?"*.

For the first decision, identifying external legal requirements and internal policies, and contractual agreements must be implemented. For example, when a PE deals with personal information such as name or address in Australia, "Privacy Act 1988" should be considered to identify the legal requirements. In addition, the decision model of data ownership/access rights should be established based on legal aspects. For example, creativity, originality, investment and source of data can be considered. The aspects are derived from the review of regulatory environment such as Berne Convention and its derivatives [40, 41], European Court of Justice (ECJ) in 2004 (William Hill case [40]) and the policies of platforms (Table 2).

Table 2. Regulations for data ownership

Category	Description	Regulation
Creativity	Creative data (video/photo)	Berne Convention and its derivatives
	Non-creative data (profile/log)	
Originality	Original data (new, raw data)	
	Derived data (modified, transformed data)	Depends on context
Investment	Non-creative and managed data by a platform owner	Court cases (e.g. European Court of Justice (ECJ)
	Non-investment data	
Source	Internal (data created in a PE)	General policies of PEs
	External (data by users)	

A certain mechanism to track and notify the compliance of the regulations should be taken into account. Identifying external/internal compliance requirements, setting conformance targets and auditing them must be carried out. The concept of due process is regarded as a pivotal control mechanism to cope with the risks of data abuse/misuse. It forces desirable behavior of participants [8], and supports successful implementation of data governance. Platform data is often used by external users such as partners or researchers. The use of data should be confirmed if it is legally permissioned. In particular, if the data is taken out and possibly disseminated for secondary use, the openness of the data and platform policies must be checked. All those processes have to be audited by third parties to avoid bias or conflict of interest, and keep transparency of a PE.

2) Data ownership and access definition. This domain refers to the decisions of "*who owns and uses the data in a PE?"*. It has been focused as a central concept of a platform design [9, 33, 42]. The decisions enable a PE to clarify the roles, responsibilities, and

comprehensive rights to data of all the corresponding participants including the data owners and subjects.

Defining data ownership and access rights of all types of data is identified as the practices of the decisions. To support implementation and keep the integrity and consistency of the outcomes of the practice, it is necessary to collaborate with other domains (Figure 2). The data classifications of all types of data which are defined in data use case domain should be used. The clarity of data ownership and access definition is improved since there might be rarely missing data in the definition. Relevant regulations identified in the regulatory environment domain must be used to develop a decision model for data ownership/access rights. As stated, the decision should be made based on the relevant regulations, laws or court cases [10]. To help practitioners understanding, we present a potential decision model which can be considered in the real world (Figure 3). The model is established based on the identified regulations introduced in Table 3. It supports a primary decision of who is the owner of (specific) data between a platform owner and the users (data contributor) of a PE. The decision should be carefully made because it is related to revenue sharing. It often leads to lawsuit like the Huffington Post case in 2011 [21].

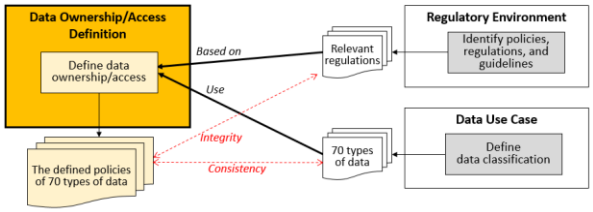

Figure 2. Collaboration with other domains

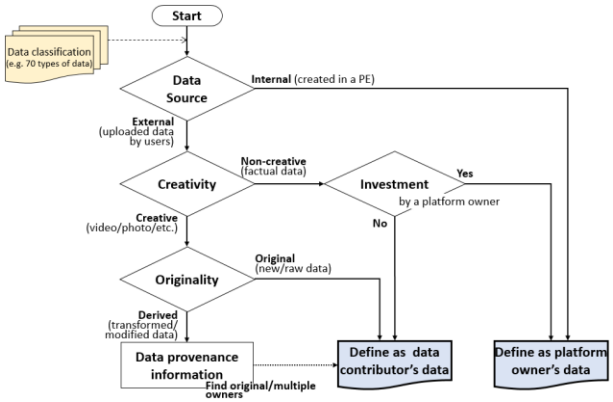

Figure 3. A data ownership decision model

Defining clear access rights facilitates platform transparency. A certain method should be available to stakeholders for giving appropriate information and security. Yet, the accessibility of data contributor to the data can be restricted by the policies or context of a PE; a platform prohibits users' access to the last password for a security reason. The governance decision makers need to consider such particular context for every single type of data in a PE. We suggest a Contribute, Own and Access (COA) matrix to support and simplify such complicated circumstance (Table 3). It allows users to clearly understand the definition of what data can own/access (or not), and to use the legitimate rights to data properly.

Table 3. An example of the use of a COA matrix

Data type	Contribute(C)	Own(O)	Access(A)
Video/photo	√	√	√
Location	√	-	-
Service use	√	-	√
Last p/w	-	√	-

Table 4. Facebook data classification

Level 1 (2)	Level 2 (8)	Level 3 (> 70)
User profile	User content	Video, photo
	Extra information of user content	Created time of photo
	User information	Name, Email
	Information about a user from other users	Post by others
	Information about a user from Facebook companies	User id, Name
	Information about others	Post to others
Service use information	Service use information	logins, logouts
	Service use information from third-party	log

3) Data use case. For PEs, how to use data is critical concern to win markets. Therefore, a series of questions, "*what types of data are collected and what are the uses of data for the business?"* and *"how should data be used without losing control?",* should be addressed in this domain.

To support the decisions, defining a data classification gives good understanding of different types of data [10] as a PE collects data from various sources. Majority data is from users as they upload content such as video, image or user information (human-sourced data) [43]. While a user uses platform services, the platform systems leave data like logs, search keywords or location (machine-generated data). This type of data is generally referred to service use information. Data is also collected through system processes through transactions, reference tables or interactions (process-mediated data). All the types of data should be considered and included in a data classification. To show an example, we identify three levels of data classification of Facebook by analyzing the policies (Table 4). The first level consists of user profile (from human) and service use information (from machine and process). The second level is divided into eight categories (six and two categories respectively). The last level of data classification comprises more than 70 types of data.

In addition, the governing body needs to decide appropriate data use cases of the collected data in alignment with the business goals. According to the result of our survey on the policies of PEs, 11 use cases have been commonly found: e.g. provide, improve and develop (test) services, communicate with users, and show and measure ads and services. The use cases must include the information of what types of data can be used for each case. It helps a platform to detect and prevent the unauthorized use of data in a data supply chain [25]. For this, the data classification identified in the previous step should be used and confirmed if every type (level 3) of data is belong to at least one of the use case and vice versa.

Monitoring and data provenance can be used as mechanisms for detecting and notifying all activities in the use of data, and tracking the derivation history of the data [8, 10]. Monitoring of the use of data should be implemented based on the defined use cases information for visible and reliable data use. Data provenance allows a platform to reserve all activities about data, identify all the associated stakeholders and prevent denial of data manipulation. It can be used to explicitly measure the contribution of data providers when there is a multiple ownership issue.

4) Contribution measurement. The success of a PE depends on the contributions made by the participating groups. Therefore revenue sharing is one of the critical governance concepts of a PE [9, 20, 33]. A number of studies note that a PE should consider the concerns such as "*what is the business value of data?"* and *"what rewards are needed for contribution of data owner?"* to encourage the contributions of the users and share the benefit from the growth of a PE.

Every participant of a PE always expects immediate rewards or future benefits [21]. The first step is to identify the specific parameters of a contribution measurement model which are related to the business success [22]. The parameters can be various depending on the business type of a PE. Some platforms like Facebook or YouTube generally rely on advertising, and grow by user content. The uploaded content plays a major role to attract the other side users and reach critical mass. User preference, likes and affiliated groups are also valuable because of targeted advertising. Meanwhile, the number or amount of service supply/purchase is the most important for the different types of business platforms such as Uber since the platform charges service fees from the users.

In the next step, proper types of rewards for users should be identified. There are three main types of rewards which can be generally considered to motivate contributors: exposure, subsidy (e.g. direct cash transfer in the form of advertising-revenue sharing, credit, physical goods, free information or technical support) and reputation [2, 21]. The different types of rewards can be used singly or in combination depending on the capability and context of a PE. Subsidy is regarded as one of important launch strategies of a PE [2]. Yet, for dominant PEs like Google, exposure can be a good choice as it has zero marginal cost but provides a big advantage to the beneficiaries. Figure 4 shows the concept of contribution measurement management.

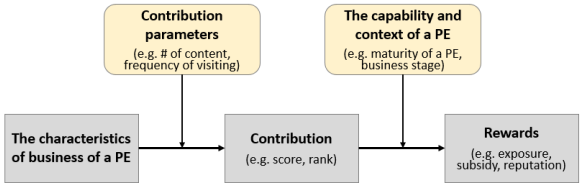

Figure 4. Contribution measurement management

Identifying the beneficiaries of rewards can be simple or complicated. If there is a single contributor, the contribution measurement will be 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 (data use case domain) helps 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.

5. Case study

The implementation of a case study presents the summary table (Table 5) to identify how and if a PE implements data governance decisions and practices in reality. We populate the decisions (with the practices) suggested in section 4 for Platform A. As noted, sufficiency of implementation of data governance is used as a metric. We illustrate the practical implementation and possible implications of governance decisions.

Platform A is a type of content portal launched in 2013. It collects software assets (development knowledge or documents), and provides the data to IT companies or individual developers for reuse. Around 3,300 software assets are currently registered in the platform. The platform is open to everyone. Any individual or company can join the platform. Yet, all data governance decisions are made by the platform owner. The governance configuration is formal and authorized-based. There are manual review processes to strictly control the quality of input data.

Table 5. The results of the case analysis

Decision	Practice	Description (implementation of Platform A)	Result (sufficiency)		Potential risk/opportunity
Regulatory environment	Identify relevant regulations	Identified the Privacy Act of the government.	▲	Partially implemented	. Limited boundary of data governance decisions (data types and regulations considered in the policies) . Lack of a comprehensive audit process and implementation
	Identify legal aspects for data use	Focus on personal information.	▲	Partially implemented	
	Identify compliance requirements	Focus on personal information based on the Privacy Act.	▲	Partially implemented	
	Audit all the use of data	Found regular audit activities to confirm legal use/and maintenance of personal information.	▲	Partially implemented	
Data ownership /access	Establish a decision model	Not found any evidence.	X	Not implemented	. Limited boundary of data governance decisions (limited types of data) . Absence of systematical decision making
	Define data ownership	Defined ownership of user content/profile.	▲	Partially implemented	
	Define access rights	Defined access rights of participants.	▲	Partially implemented	
Data use case	Define a data classification	Not found a data classification form (found technical documents only).	▲	Partially implemented	. Limited boundary of data governance decisions (limited types of data) . Absence of a control mechanism of unauthorized data use
	Identify data use cases	Documented the use of data by the platform or third-parties in the policies.	▲	Partially implemented	
	Monitor/record all activities in the use of data	Not found any evidence (how to monitor/trace the use of data).	X	Not implemented	
Contribution measurement	Identify specific parameters	Identified # of views/likes/votes as parameters for measurement the quality of user content.	O	Implemented	. Needs for test the effectiveness of current strategy for innovation
	Identify proper types of rewards	Identified exposure as a type of reward used in the platform.	O	Implemented	

5.1. Results of the case analysis

Regulatory environment— We confirmed that all the practices investigated in this domain are partially implemented since the platform focuses on personal information and the relevant regulation only. The Privacy Act of the government is identified as the critical and only regulation. A set of compliance requirements is identified and implemented based on the regulation by external auditors. Yet, the audit is confined to the personal information management.

Data ownership and access—There have not been found any idea of how to define data ownership and access rights. The policies shows personal information and uploaded content are owned/accessed by the provider. Yet, there is no clear understand of how data ownership and access rights should be defined and what legal aspects should be considered for the decision. It leads to difficulties to include all types of data in the decision making system. System data such as logs or service use information is not only currently defined as to who owns the data, but also there is no prepared decision model for future.

Data use case— We have not found any monitoring/data provenance mechanisms to control use of data by internal/external users. How to control (monitor) unauthorized data use is not clearly defined (except user reporting). When it comes to data classification, different types of data are defined in the technical documents. The definition, however, is not a form of a data classification which is generally well categorized and organized. The use of the collected data by the platform or third-parties is identified and briefly documented in the policies.

Contribution measurement— Platform A aims at high reputation and satisfaction of users by providing good quality of data. In this sense, the value of data is measured based on the number of views, likes or votes taken from the platform users. As a form of reward, the platform is using exposure (ranking) for the contributors. The contribution measurement of the platform is simple but effective enough as it is based on the users' participation (# of views, likes and votes). It reduces administrative work for the platform owner.

5.2. Discussion

This case study shows the fact that the theoretically important governance decisions are addressed in the real world. The implementation of the four decisions (regulatory environments, data ownership/access rights, data use case and contribution measurement) have been found in data governance of the platform. Yet, there are some findings which should be discussed to improve data governance capability as follows.

First of all, there is limited boundary of governance decisions. As noted, data governance decisions should be made including all types of data collected, used and shared by a PE. In particular, clear ownership/access rights of all types of data is crucial to manage data in a PE without losing control as there are complicated relationships and interactions by multiple parties. However, Platform A has the focus on personal information and user content, and other types of data like system data have less attention. It results in a lack of implementation of governance practices in all decisions such as limited definition of data ownership, access rights, data classification, and data use cases. It also affects the service agility or reputation of a PE. In this platform case, some users recently inquired if they can have access to the information of who viewed/downloaded

their content. The platform couldn't respond to the inquiry because the access rights of users to such system data (like Table 3) are not clearly defined.

Secondly, lack of control mechanisms can be the main cause of invisible use of data which can lead to data misuse/abuse. There are insufficient documentation and activities for monitoring and audit the use of data. It causes limited (or no) implementation of data use control. In addition to this, the data used for each use case is not precisely defined, and thus there are difficulties to identify unauthorized use of data. According to our investigation, even though this issue can affect negatively on the platform (e.g. less secure), it has not been recognized before by the governing body. It has been now accepted as a potential hazard which should be seriously addressed.

We identified several factors which cause the issues discussed above. The first reason stems from a lack of awareness of the needs and importance of data governance in the context of PE. It is derived from absence of adequate information and experience about those concerns. Platform A is a non-commercial platform. The business context allows the quite limited use of data. Unlike commercial PEs such as Facebook, there is not any family companies or third-party partners to sell or share the data. Any data in the platform is not used for productization or advertising purpose. Such context causes less attention to the identified issues by the platform.

This case study provides understanding of how and if an industry PE addresses the data governance decisions in reality. We identified some issues to discuss potential risks and opportunities and help correct decision making. The results of the study allow PEs to see what decisions they need to consider when setting up or improving data governance.

6. Conclusion

A PE needs to orchestrate complicated context, processes and relationships occurred among multiple parties contributing and using data. Lack of data governance of a PE can cause destructive consequences such as data abuse/misuse, and lead to market failure. Traditional data governance focuses on in-house control of data, and prior research on platform governance is still in its infancy. There is a need for a reference model for PEs to support correct decision making, but it has not been found.

In this paper, we proposed data governance decisions for PEs which should be made to ensure effective management and use of data. We also broadly discussed what practices need to be implemented for the decisions. For this, we surveyed industry platforms and reviewed governance frameworks and literature. This study delivers lots of ideas and considerations to practitioners by presenting how the identified decisions can be implemented. We also provided potential models and examples based on the survey on industry PEs and literature review which can be applicable in practice. We carried out a case study to illustrate the practical implications. Through the case study, we showed that this study is practically applicable and can be a leverage to increase the capability of data governance of PEs. Yet, there are several limitations that remain in this study. The case study has possible validity issues as it was carried out and assessed by one author with her working experience and limited number of interviewees. In addition, there is an external validity issue as this study uses a single case. Future work is planned to conduct a multiple-case study. To do so, we will select multiple platforms underlying different context and business models to generalize the findings and compare what might be the reasons for a different implementation.

7. References

[1] D.S. Evans, "Platform Economics: Essays on Multi-Sided Businesses," Competition Policy International, 2011.

[2] G. Parker, and M.W. Van Alstyne, "Platform Strategy," in the Palgrave Encyclopedia of Strategic Management, 2014.

[3] A. Smedlund, and H. Faghankhani, "Platform Orchestration for Efficiency, Development, and Innovation," in 2015 48th HICSS, 2015, pp. 1380–1388.

[4] S.U. Lee, L. Zhu, and R. Jeffery, "Data Governance for Platform Ecosystems: Critical Factors and State of the Practice," in the 21th PACIS, 2017.

[5] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big Data: Issues and Challenges Moving Forward," in 2013 46th HICSS, 2013, pp. 995–1004.

[6] S. Kaisler, W. H. Money, and S. J. Cohen, "A Decision Framework for Cloud Computing," in 2012 45th HICSS, 2012, pp. 1553–1562.

[7] H.V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J.M. Patel, R. Ramakrish-nan and C. Shahabi, "Big Data and Its Technical Challenges," Communications of the ACM (57:7), 2014, pp. 86–94.

[8] K. E. Martin, "Ethical issues in the Big Data industry," MIS Quarterly Executive (14), 2015, p. 2.

[9] 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.

[10] V. Khatri, and C. V. Brown, "Designing data governance," Communications of the ACM (53:1), 2010, pp. 148–152.

[11] F. De Abreu Faria, A.C. Gastaud Macada, and K. Kumar, "Information Governance in the Banking Industry," in 2013 46th HICSS, 2013, pp. 4436-4445.
[12] C. Dimick, "Governance Apples and Oranges," Journal of AHIMA; Chicago, (84:11), 2013, pp. 60–62.
[13] ISACA, "COBIT 5: A Business Framework for the Governance and Management of Enterprise IT," 2013, (available at https://www.isaca.org/COBIT/Pages/ COBIT-5-Framework-product-page.aspx).
[14] D.S. Evans, "Governing Bad Behavior by Users of Multi-Sided Platforms," SSRN Scholarly Paper No. ID 1950474, Rochester, NY: Social Science Research Network, 2012.
[15] Gawer, A., Platforms, Markets and Innovation, Edward Elgar Publishing, 2009.
[16] A. Hagiu, "Strategic Decisions for Multisided Platforms," MIT Sloan Management Review (55:2), 2014, p. 71.
[17] N. Lomas, "Facebook data misuse scandal affects "substantially" more than 50M, claims Wylie ", (available at https://techcrunch.com/2018/03/27/facebook -data-misuse-scandal-affects-substantially-more-than-50m-claims-wylie/).
[18] M. Zimmer, "'But the Data Is Already Public': On the Ethics of Research in Facebook," Ethics and Information Technology (12:4), 2010, pp. 313–325.
[19] B. Ives, and V. Krotov, "Anything You Search Can Be Used Against You in a Court Of Law: Data Mining in Search Archives," Communications of the Association for Information Systems (18:1), 2006, p. 29.
[20] A. Hein, M. Schreieck, M. Wiesche, and H. Krcmar, "Multiple-Case Analysis on Governance Mechanisms of Multi-Sided Platforms," In Multikonferenz Wirtschaftsinformatik, 2016.
[21] 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.
[22] 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.
[23] 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.
[24] P. Weill, and J. W. Ross, "IT Governance on One Page," SSRN Scholarly Paper No. ID 664612, Rochester, NY: Social Science Research Network, 2004.
[25] ISO, (available at https://www.iso.org/standard/56639. html).
[26] G. Thomas, The DGI data governance framework, The Data Governance Institute, 2006, (available at http://www.datagovernance.com/wp-content/uploads/2014/ 11/dgi_framework.pdf).
[27] Informatica, "Holistic Data Governance: A Framework for Competitive Advantage," 2013, (available at https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/white-paper/holistic-data-governance-framework_white-paper_2297.pdf).
[28] C. Ballard, J. Baldwin, A. Baryudin, G. Brunell, C. Giardina, M. Haber, E. O'neill, and S. Shah, "IBM Information Governance Solutions," 2014, (available at http://www.redbooks.ibm.com/abstracts/sg248164.html?Open).
[29] 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.
[30] 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.
[31] J. Manner, D. Nienaber, M. Schermann, and H. Krcmar, "Governance for mobile service platforms: A literature review and research agenda," in ICMB, 2012.
[32] 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.
[33] Tiwana, A., Platform Ecosystems: Aligning Architecture, Governance, and Strategy, Newnes, 2013
[34] S.U. Lee, L. Zhu, and R. Jeffery, "Designing Data Governance in platform ecosystems," In: 2018 the 51st HICSS, Hawaii, 2018.
[35] Choudary, S. P., Platform Power: Secrets of billion-dollar internet startups, 2013.
[36] 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.
[37] P.M. Schwartz, and D.J. Solove, "The PII problem: Privacy and a new concept of personally identifiable information." NYUL rev., 2011.
[38] Dempster, C., D.S. Williams, and J. Lee, The Rise of the Platform Marketer: Performance Marketing with Google, Facebook, and Twitter, Plus the Latest High-Growth Digital Advertising Platforms, John Wiley & Sons, 2015.
[39] K. Leswing, "Tim Cook speaks out on the Cambridge Analytica scandal, says Facebook's collection of user data 'shouldn't exist'," (available at https://www.businessinsider.com.au/apple-ceo-tim-cook-on-facebook-cambridge-analytica-scandal-2018-3).
[40] E. Harison, "Who owns enterprise information? Data ownership rights in Europe and the US. Information & management," 47(2), 2010, 102-108.
[41] R. Elliott, "Who owns scientific data? The impact of intellectual property rights on the scientific publication chain," Learned publishing, 18(2), 2005, 91-94.
[42] L.D. Thomas, E. Autio, and D.M. Gann, "Architectural Leverage: Putting Platforms in Context," The Academy of Management Perspectives (28:2), 2014.
[43] 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.
[44] Yin, R. K., Case Study Research: Design and Methods (Applied Social Research Methods, Vol. 5), Sage Publications, Beverly Hills, CA. Rick Rantz Leading urban institutions of higher education in the new millennium Leadership & Organization Development Journal, 1994, 23(8), 2002.