MODES OF GOVERNANCE IN INTER-ORGANISATIONAL DATA COLLABORATIONS

Complete Research

Broek, Tijs van den, TNO, Van Mourik Broekmanweg 6, 2628 XE Delft, the Netherlands, tijs.vandenbroek@tno.nl

Veenstra, Anne Fleur van, TNO, Van Mourik Broekmanweg 6, 2628 XE Delft, the Netherlands, annefleur.vanveenstra@tno.nl

Abstract

Big data and data-driven innovation are drivers for economic growth. To capture this growth, data often need to be shared among organisations. However, many challenges to sharing data among organisations exist. This paper investigates how governance is organised in inter-organisational data collaborations. First, based on literature, four archetypical modes of governance are identified: Market, Hierarchy, Bazaar and Network. Subsequently, these theoretical modes are investigated empirically by exploring governance modes in four use cases. Based on a cross-case comparison, we find that major challenges to data sharing are the commercially sensitive nature of data and privacy risks. Due to legal implications, sharing of personal data always takes place hierarchically. Therefore, coordination and control over data need to be firmly in place before organisations engage in data sharing. Further research should look into how these aspects can be organised in inter-organisational data collaborations to foster innovation.

Keywords: Big Data, Governance, Inter-organisational Collaborations, Data Sharing

1 Introduction

Organisations increasingly collect, store and process data. This ‘data deluge’ requires unconventional data infrastructures, such as processing power and storage capacity, to keep up with the variety, volume, velocity, variability, complexity and value of big data (Katal, Wazid and Goudar, 2013). A core principle of big data is data maximisation: more data (combinations) mean more opportunities to extract value (IWGDPT, 2014). Analytics and visualisations assist organisations in exploring big data for valuable insights. The advent of big data promises organisations valuable business analytics to improve their operational efficiency, the effectiveness of products and services, and the development of new products, services and business models (Gopalkrishnan et al., 2012). Public organisations have great expectations of big data to inform policy-makers and develop solutions to societal challenges, such as resource efficiency, sustainability and healthy ageing (Borgman, 2012; Bertot and Choi, 2013). Organisations are, thus, keen to invest in these infrastructures as data are seen as valuable and intangible assets to the organisation (Applegate, Austin and McFarlan, 2003; Kathrin and Brown, 2010; Gopalkrishnan et al., 2012; Van Veenstra and Van den Broek, 2013).

As big data require large investments in infrastructure and skills, and datasets are often scattered among organisations, data collaborations are formed (Bertot and Choi, 2013). Data collaborations are arrangements between three or more organisations. Consequently, organisations jointly establish data protocols, data exchange and reporting mechanisms and analyse data (Bertot and Choi, 2013). To mitigate these risks while maximising the value of data collaborations, organisations design and implement governance structures. IT governance refers to the arrangements that enable organisations to formulate, communicate and assess policies and procedures that arrange formal control of IT activities (Sambamurthy and Zmud, 1999). Research on governance related to big data often focuses on single organisations (Kathri and Brown, 2010; Gopalkrishnan et al., 2012). However, inter-organisational collaborations are notoriously difficult to manage (Ireland, Hitt and Vaidyanath, 2002), for instance because of privacy concerns or to retain a competitive advantage (Markus and Bui, 2012). In line with governance of inter-organisational systems (Kumar and Van Dissel, 1996), big data collaborations need well-designed and implemented inter-organisational governance to mitigate risks. Therefore, this paper explores inter-organisational governance of data sharing. Using an interpretative study, we study data governance in cases of big data collaboration.

The contributions of this paper are twofold. Firstly, the scope of research on governance in the context of big data is extended to the inter-organisational level. Secondly, we provide an in-depth analysis of four big data collaborations that vary according to their mode of governance. This paper is structured as follows. Firstly, we develop a theoretical framework based on inter-organisational governance research in organisational and IS studies. Secondly, we describe the methods of our empirical study and analyse four use cases to find out how these elements are implemented in practice. After a cross-case analysis and discussion of the findings, finally, we formulate conclusions and recommendations.

2 Theoretical background

2.1 Governance of inter-organisational collaboration

Inter-organisational collaboration refers to constellations of three or more autonomous organisations that collaborate to pursue collective rather than individual goals (Provan and Kenis, 2008). From a sociological perspective, inter-organisational collaboration is a form of collective action: a social organisation that creates more value than the sum of its individual participants (O’Toole Jr, 1997). Effective inter-organisational collaboration provides competitive advantage to its members in several ways (Provan and Kenis, 2008). Firstly, organisations can learn from other participants. Secondly, inter-organisational collaborations can pool resources, which increases efficiency. Lastly, collaboration can stimulate the development of new products and services, or the improvement of current products and services (Lowndes and Skelcher, 1998).

The effectiveness of inter-organisational collaboration depends on the governance that is in place. Inter-organisational governance consists of the arranged institutions and structures to ensure that individuals behave in line with the collective goals, conflicts between individuals are prevented or resolved, and the effective and fair use of collective resources within the inter-organisational collaboration (Provan and Kenis, 2008). Apart from legal aspects of governance, inter-organisational aspects include “command structures and authority systems, incentive systems, standard operating procedures, dispute resolution procedures and non-market pricing systems” (Dekker, 2004, p. 31). Inter-organisational collaborations often take the form of networks, which balance the strong incentives of the market and the structures of hierarchy (Gulati, 1995; Adler, 2001; Powell, 2003).

In literature, four archetypical inter-organisational modes of governance are distinguished: Market, Bazaar, Hierarchy, and Network (Provan and Kenis, 2008; Lowndes and Skelcher, 1998; Dekker, 2004; Demil and Lecocq, 2006). Table 1 summarises these four modes of governance, according to a number of characteristics: normative basis, incentives for engagement, control over these incentives, reasons for adoption, flexibility and durability of the collaboration, social contract, relations between the individual members, and type of coordination used.

	Market	**Bazaar**	**Hierarchy**	**Network**
Normative basis	Intellectual property	Open license	Formal hierarchy	Social contracts
Incentives for engagement	Competition	Reputation in the community	Career	Trust
Control over the incentives	High: contracts	Low: reputation in the community	High: administrative power	Moderate: reciprocity and social contracts
Reasons for adoption	Low coordination costs; high flexibility in participants	Innovation and low coordination costs	Negotiation position; strategic differentiation	Low-cost access to resources; joint solutions
Flexibility of the collaboration	High	High	Low	Moderate
Duration of the collaboration	Short term	Unlimited	Unlimited	Long term
Social contract	Formal, distrust	Informal, focus on joint production of products	Formal, bureaucratic	Informal, focused on common goals
Relation between network members	Independent	Partially dependent	Dependent	Interdependent

Table 1. Characteristics of four modes of inter-organisational governance.

The Market governance mode affords high level of autonomy to network members. Dyadic contractual agreements between buyers and suppliers diminish the need of trust between members. For example, conflicts resolution is regulated by contract law. The degree of control over the collaboration is high. The prime motivation of collaboration is competition: organisations work together to advance their competitive position. When better opportunities (e.g. lower prices) emerge in the market, organisations swiftly change their collaborations. Market governance has relatively high transaction costs due to these short-term relationships. Consequently, the identity of members is not important. In a Market governance mode, organisations can decide to pool the data in a central or pooled marketplace. Pooled resources require little coordination: members share data in the central repository and have contractual transactions when needed.

Central to the Bazaar mode of governance is a community of actors that chaotically cooperate on a common goal. A bazaar does not require formal contracts or high levels of trust to coordinate network behavior (Demil and Lecocq, 2006). Unlike the Market type of governance, users' reputation in and contribution to the community (e.g. kudos) are prime motivators to contribute to the common goal. This reputation mechanism makes the identity of community members of moderate interest to the cooperation. Unlike in markets, intellectual property is of minor importance: community members waive ownership by means of an open license, to ensure that the developed products or services are distributed widely. Members of the bazaar are fairly autonomous in their decision making, and social control is regulated by transparency and reputation in the community.

The Hierarchy governance mode emphasises formal relations between the individual members. Higher ranked members have formal power over lower ranked members in the network. Members are motivated to climb the rankings in the collaboration ('career opportunities'), and behaviour is regulated by sanctions and rewards. Consequently, the identity of members and the resulting trust is not necessary to form the collaboration. Hierarchical collaborations often include a dominant organisation or a network-specific umbrella organisation that coordinates and administrates joint efforts. Sequential inter-organisational collaboration often occurs in Hierarchical governance modes. The dominant organisation orchestrates and monitors the data exchange along the supply chain.

Networks are considered as a hybrid, yet distinctive, organisational form in between markets and hierarchies (Powell, 2003; Ring and Van de Ven, 1994). The Network governance mode relies on social contracts between members. These social contracts imply reciprocity between members: members need to trust each other, and hence knowing the identity of members and previous experience in collaboration is needed to build trust. In contrast to a market or hierarchy, coordination of network activities is a joint effort between network members and decisions are made based on consensus between all members. The reciprocal coordination of the Network mode of governance is a complex web consisting of data exchanges between individual members. As this reciprocal coordination becomes more complex and collaboration become more uncertain, the need for hierarchical coordination mechanisms increases (Gulati and Singh, 1998; Dekker, 2004).

2.2 Governance of data collaborations

IT governance defines the decision rights and accountabilities to encourage desirable behaviour in the use of IT within an organisation (Weill, 2004). Recently, scholars have drawn attention to IT governance in inter-organisational networks (Markus and Bui, 2012; Zaric, Stolze, Boehm and Thomas, 2012; Stolze, Zaric and Thomas, 2011; Pardo, Gil-Garcia and Burke, 2008; Spil, Van den Broek and Salmela, 2010). Based on a survey among IT professionals and academics, Stolze et al. (2011) argue that IS scholars should study IT governance in inter-organisational relationships. Due to an increase in sharing data, data governance becomes an important aspect of IT governance (Bertot and Choi, 2011). Data governance is defined as “who holds the decision rights and is held accountable for an organisation’s decision-making about its data assets” (Kathri and Brown, 2010, p. 149).

The four archetypical inter-organisational collaborations described in the previous section concern general governance rather than specific governance of (big) data collaborations. Therefore, the next step is to extend Table 1 and apply the four theoretical modes of governance to data collaborations. This extension is shown in Table 2. It includes the characteristics of data sharing in each of the four types of inter-organisational collaboration, the main coordination mechanisms used in the constellations, and the control individual member(s) have over the data. Furthermore, an example of each type of inter-organisational data collaboration is provided.

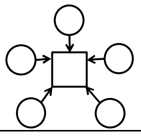
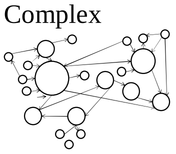
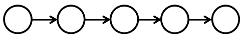
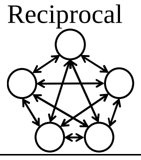
	Market	**Bazaar**	**Hierarchy**	**Network**
Type of data sharing	Pooled 	Complex 	Sequential 	Reciprocal 
Characteristics of data sharing	Buy and sell data based on (dyadic) transactions	Open up and reuse of data	Data exchange orchestrated by dominant member(s)	Lateral data exchange between individual members
Coordination mechanisms	Contracts	Data quality	Power exerted by the dominant member(s) over the others	Trust
Control over data	Remains at individual organisations	Open licence means that everyone has access to the data	Determined by the dominant member(s)	Remains at individual organisations
Example of data collaboration	Central marketplace for big data	Open data community	Supply chain network	Networked exchange of data

Table 2. Characteristics of four modes of inter-organisational data governance.

In a Market mode of governance, data supply and demand are met via a marketplace and via contracts between individual organisations. The control over data remains at individual organisations, until the data are sold. In that case, control over the data is in the hands of the buyer. In a Bazaar governance mode, data is open and supply and demand are determined by the quality of the data. This also means that everyone has access to data. In a Hierarchy, data are exchanged based on the needs of the dominant member(s) that is able to exert power over the other member(s). The Network mode of governance is a hybrid, with member(s) laterally exchanging data while retaining control over this exchange. Trust relations are formed, which form the basis of the data exchange.

To our knowledge, no study has yet connected inter-organisational governance mode and data governance. Based on the literature review in the previous section, we expect that the governance of inter-organisational data collaboration depends on the type of data sharing, the characteristics of the data sharing, the coordination mechanisms and control individual members have over data within the collaboration. In the next section, we study these relations in four cases of data collaborations.

3 Case studies

3.1 Methodology

After identification of the aspects influencing the governance of inter-organisational data sharing, such as the type of inter-organisational data sharing, characteristics of data sharing, coordination mechanisms and control over the data, the next step of this research is to explore these aspects in practice. For this, we use an interpretivist methodology that allows in-depth investigation, fitting the complexity of the matter (Klein and Myers, 1999). We use a multi case-study approach to explore and compare big data collaborations, which allows for cross-case comparison and reflection on differences and similarities between cases (Yin, 2001). To investigate the characteristics of data collaborations, we identified four domains that we expected to match the four inter-organisational governance archetypes.

A use case of personal marketing via a loyalty program by a large retailer is used to investigate the Market model of data sharing. An open data portal of a large municipality presents a Bazaar set-up. The Hierarchical collaboration is investigated by looking at data sharing around the database of a healthcare insurance company. This database is an epidemiological dataset that allows re-use of transaction data on the use of healthcare collected for administrative purposes. The fourth use case is an energy data sharing platform that aims to establish a Network type of collaboration. All use cases are located in the Netherlands. Although we sampled the case studies on the governance archetype we expected, the outcomes of the collaborations may differ in practice. For example, actors may use a hybrid form of the archetypes rather than an archetype. In line with our interpretive approach, the case studies assist us in exploring the four archetypes rather than testing them.

For the data collection we used semi-structured interviews. The interviews focused on the four aspects of data governance: the type of inter-organisational data sharing, characteristics of data sharing, coordination mechanisms and control over the data. Therefore, the interview comprised questions about the collaboration, data sharing, type of data, technical infrastructure, privacy risks and other challenges. The retailer use case was based on two interviews, with the manager of Personal Marketing and with a data consultant. The open data use case was based on six interviews, two with the director of the open data portal, one with a provider of datasets via the open data portal, two with initiators of the open data portal and two with users of the open data from the portal. The health database case was based on four interviews, with the manager of the databases, with a strategic advisor and with two users of the data. And finally the energy platform use case was based on two interviews with the project leader and a data provider at an energy company. All interviews lasted between 30 minutes and one hour and took place between November 2013 and June 2014. We complemented the interviews with desk research that included presentations, project plans, and reports about the cases.

3.2 Personal marketing

The retailer from the Netherlands is best known for owning a chain of supermarkets. The organisation uses transaction data for profiling and marketing purposes. It is expected that customers buy more when they receive offers that are tailored to their needs. The big data collaboration is established between the individual retailers, headquarters and a consultancy that performs the data analyses. The transaction data are collected using a loyalty card scheme. The users of the loyalty card scheme can register their cards online, through which it becomes possible to identify purchases of individuals. Based on the analyses of the transaction data, the company profiles customers and offers to their customers a list of products tailored to their needs. Step by step also other data sources, such as demographical data and market research, are linked to gain more insight into the customers and their needs.

The transaction data of those loyalty cards that are registered online contain personal data. These data have a unique identifier. The data collected from the individual retailers are collected in the stores, stored centrally and sent to the consultancy for performing analyses, after anonymisation. Re-identification happens after the analyses are carried out, just before personalised e-mails are sent. Only certified employees have access to these data and the company performs regular internal and external audits. The data that are collected are owned by the retailer. Upon registration, the client gives explicit consent for data processing and this consent can be revoked, after which data will be removed. Data are not sold to other organisations.

3.3 Open data portal

The open data portal of a large municipality was set up by an institute of applied scientific research within the municipality. The institute intended to re-use datasets of the municipality within its research projects and started collaborating with the department dealing with city maintenance, waste management and the public sphere. This department has a lot of geographical data as well as information on objects in the city and opened up many of its datasets. Therefore, a next step was to publish these datasets in an open data store. This open data store was also set up by the institute, but the municipality took it over after the city council embraced the notion of open data. Before setting up the open data portal, the municipality sold these data, but in the open data portal data are provided for free. The goals for the municipality in relation to open data are to increase efficiency of the organisation, stimulate innovation within the municipality and allow for re-use and innovation within other organisations, such as app developers, and to increase transparency and accountability.

While most of the data is merely provided in the portal, sometimes the municipality collaborates with the users of the datasets. In some cases, the users of open data request specific datasets to be opened up, such as datasets with geographical and real-time data. Data are published in a format that fits the type of data, such as SQL for database information and csv for geographical data. Also metadata is added to make the published datasets more easy to find. The municipality mainly collaborates with start-ups and app developers to create value from open data. An example is the ‘tree spotter’ app, showing information on all 180.000 trees in care of the municipality. Other organisations do not publish data via the portal. A next step in the development could be adding social media or co-creation of services, such as interactively show social media data related to the objects in the public sphere. The municipality currently also explores the option of the portal being run commercially. This might mean that other (semi-)public organisations, such as schools and hospitals, will be able to publish their data in the portal too.

Since the data that are published contain no personal data, no primary privacy concerns occur. However, when data are published and combined with other data sources, privacy risks may occur if data can lead to re-identification. A special mentioning should be made of geographical data, such as addresses, which are not personal data in itself, but could easily lead to re-identification. This is also an issue within the municipality. Some departments would like to use the citizens’ registry to do better analyses for maintaining the public sphere, for example, but because of data protection legislation this is not allowed. Still, privacy issues may occur with the open data portal, as it is impossible to determine all possible combinations with open data, which may lead to re-identification.

3.4 Healthcare database

The expertise centre of the largest health insurance company in the Netherlands is responsible for improving the quality of health processes. In order to do so, the centre collects data on the use of healthcare services of approximately 4.8 million citizens in a large-scale healthcare database. The database contains detailed information about medical care and costs incurred over a period of twelve years. Additionally, the database includes detailed information about patients and healthcare providers. Healthcare providers automatically send these data to the insurance company, and the company checks the data quality. While the data are generated for administrative purposes, such as the administration of health reimbursements and quality audits, the expertise centre aims to innovate and improve the effectiveness of healthcare by supporting research based on these data. Therefore, the database is occasionally accessible to external researchers (e.g. from research institutes or pharmaceutical companies). About twenty requests for access are accepted every year.

Access to these data is thus hierarchically organized and strongly controlled by the insurance company. A review board, with staff from the insurance company and research institutes, evaluates the ethical, theoretical, methodological and societal quality of the requests for access to the data. Compliance to the Dutch data protection act is an important requirement for acceptance. In order to prevent the insurance company from any legal and reputational damage that can result from poorly executed or commercial research, publications that result from research based on these data also require approval from this review board. Furthermore, guidelines and procedures for data management, such as anonymisation of the data are internally monitored and externally audited. The insurance company does not directly provide anonymised data to external parties. A Trusted Third Party (TTP), a non-profit organisation, pseudonimises the data to minimise the risk of re-identification. Access and recombination of the data is provided through this service.

To minimise privacy risks, the insurance company emphasises the importance of transparency and widely communicates its data policies to its clients. When signing the health insurance contract, clients automatically accept that their data could be shared for scientific purposes. This procedure, however, is not an informed and explicit consent as formulated in regulation, as that would be too time-consuming. While the right on privacy of clients is important, it is carefully considered and compared with scientific and societal goals. Ownership of healthcare data is not strictly determined, however. It is unclear if data are owned by the patients, the healthcare organisation or by the insurance company. Patients are increasingly seen as owners of their data, which makes it unclear how to implement governance for sharing healthcare data. Therefore, it is currently written down in a data security policy to be signed by all users, but there is currently no audit to check compliance to the security policy. Similarly, it is not clear who is liable in the case of data misuse. The strict procedures by the review board aim to prevent any misuses, but after sharing the data, the only control the review board can exert is to block a publication. Liability is thus determined case by case, as strict guidelines regarding ownership could increase the threshold to share data for scientific purposes and increase the administrative costs, for example to audit compliance.

3.5 Energy data platform

Energy data is an asset in the energy market as it may facilitate matching energy demand and supply, offer services to motivate consumers to save energy, or inform municipalities where illegal energy consumption takes place (e.g. drugs labs). However, energy suppliers and energy grid operators in the Netherlands are reluctant to share energy data. The main reasons are the costs of sharing data, lack of IT skills and knowledge to extract value out of the data, uncertainty about the benefits of sharing data, privacy risks, and stakeholder complexity. In 2013, a grid operator, a national research institute and a telecommunications provider started a project on setting up an open energy data platform in the Netherlands. The goal of this platform is to share data to stimulate energy-efficiency, innovative energy services and a transition towards sustainable energy. Platform members form an organisational network together with the telecommunications provider as platform provider. The open energy data project is currently in its inception phase.

Digitisation of administrative processes, smart grids, smart meters, smart thermostats, mobile applications and social media rapidly increase the amount of energy data in the Netherlands. For example, an energy company aims to increase the approximately 30,000 smart thermostats to 2.2 million thermostats within five years. The open energy data platform will include data from grid operators (e.g. energy peaks or leakages and invoices), smart thermostat data (e.g. energy consumption on an aggregated level), relevant telecommunications data (e.g. drops in modems to detect energy fall-outs) and relevant open data from government agencies (e.g. geographical information). Not all data in the project are freely available to other parties: tariffs vary from free to commercial tariff. The project strives for data maximisation on the long term: the network is open to new participants, such as other grid operators, energy suppliers and energy service developers (e.g. energy mobile applications), and the platform also aims to link data from other domains, e.g. logistics or house construction data, to stimulate cross-sectoral innovation. However, this has not happened yet.

The open energy data platform data vary in aggregation level. Regional energy consumption data pose no privacy risks, as combination with other datasets is unlikely to lead to re-identification of individuals. On the other hand, data from smart meters and thermostats present more risk. For example, thermostat data may be available per six households. The collection and analysis of smart meter and thermostat data is monitored by the Dutch data protection authority and the Dutch consumer and market authority. Next to compliance to data protection legislation, members argue that the public perception about privacy is an important issue for accepting energy data collaboration. Privacy risks may increase in the future, as network members will link datasets across sectors. The network members jointly own the open energy data infrastructure, and consider individual members to be owner of the data they publish on the platform. Custom-made data agreements between network members regulate the licensing of data.

On the short term, network members prefer to solve conflicts regarding data agreements on a one-to-one basis, as they do not want to endanger their relations or inflict reputational damage to the network. Accountability is based on social contract between individual members. In the long term, project members may want to arrange accountability through foundation of an umbrella organisation consisting of representatives from all network members, including the platform provider. This network umbrella organisation will govern the open energy data, including the monitoring of data management and regulation. An external organisation will audit the platform. Project members state that an open energy data platform requires transparency towards energy consumers, when it processes data on the individual. However, questions on how transparency towards consumers can be arranged and who is responsible in the open energy data project to provide this transparency remain unanswered as of yet.

4 Findings

The findings from the four use cases in the previous section are presented in Table 3. Four aspects of the data collaborations are looked into more closely: the type of inter-organisational data sharing, characteristics of data sharing, coordination mechanisms and control over the data.

	Personal marketing	**Open data**	**Healthcare database**	**Open energy data**
Type of data sharing; mode of governance	Hierarchical; all data are owned by the retailer	Bazaar mode	Hierarchical; the insurance company determines data sharing	Network; different types of data sharing co-exist
Characteristics of data sharing	Commercial relation between the retailer and the consultancy	Free supply of open data by the municipality	Determined by the review board	Mixed; different types via one platform
Coordination mechanisms	Standardisation between individual stores and the main branch; contract between the retailer and the consultancy	Quality and usefulness of the data to users	Careful deliberation of the usefulness of the research against the potential privacy infringement	Mixed for the individual data sharing activities, but overall standardisation via the platform
Control over data	Only the retailer; individuals can opt in and opt out	Everyone has access to the data	Strictly controlled by the insurance company	Individual organisations remain in control over their own data

Table 3. Findings on inter-organisational governments from the use cases.

Regarding the *type of data sharing* within the big data collaboration, none of the cases were found to represent the Market governance mode, which means that in none of the cases an example could be found in which data was shared openly for a commercial purpose. All cases indicated that they could not yet establish a business case for sharing data in this manner. Research and innovation were the most often found reasons for sharing data in inter-organisational collaborations. The cases varied from one-to-one data sharing (personal marketing), to one-to-many (open data portal and healthcare database), to many-to-many (energy data platform). The one-to-one model offers organisations most control; the many-to-many model is most complex. There appears to be a relation between complexity and openness of collaboration. The personal marketing and healthcare database cases represent closed models, the open data portal and the energy data platform represent an open form of collaboration.

Regarding the *characteristics of data sharing*, sharing and combining data does not take place on a large scale, which also means that few collaborations take place at the moment. Furthermore, few cases show sign of data maximisation, which means that the potential of big data to come up with unpredictable applications is not yet realised. However, all cases expect that the use of data will increase in the future.

The *coordination mechanisms* differed substantially among the cases. In the hierarchically organised use cases, the retailer and the healthcare database, the dominant organisation mainly set the standard for data sharing, either via a contract or a protocol. In the open data portal, no coordination except for (technical) maintenance takes place. The quality and usefulness of the data determine re-use. In the Network mode of governance, all types of coordination co-exist, based on the specific data and the organisations involved. The only coordination taking place is that all data are shared via the platform.

Control over data also varied among the cases. In the Hierarchy government mode it is determined by the dominant organisation, which tightly controls what happens with the data. In the open data portal, the Bazaar mode of governance, no control over the data is exerted. In the Network governance mode, individual organisations retain control over their data, leading to different outcomes of data sharing.

5 Discussion

The use cases show that the governance mode of data sharing in inter-organisational collaborations is influenced by the characteristics of the data sharing, the coordination mechanism and the control organisations retain over their data. Similar to the findings from Markus and Bui (2012), we found that two important reasons for wanting to keep tight control over data were the commercial sensitivity of data and the privacy risks involved. The clearest case in which commercial sensitivity is involved, is in the case of personal marketing. All data remain clearly within control of the retailer that does not want their competitors to get the same insight into their customers' behaviour. But also in the case of the energy data platform this was mentioned as a barrier to establishing inter-organisational data collaboration. While this is a barrier to data collaborations, it may be overcome by installing appropriate governance mechanisms.

The other factor that had a strong influence on the coordination mechanism and the control over the data was the risk of privacy infringement. When personal data were not involved, such as in the case of the open data portal, free and open data sharing was observed, but as soon as personal data were involved (in the other three cases), the coordination mechanism called for was strict control of a hierarchical nature. This was even the case when no apparent privacy infringement could be observed (yet), but the threat of re-identification was sufficient to call for a hierarchical governance mode. The reason for this is the existence of data protection legislation, which requires organisations to retain control over any personal data they process. As data minimisation is an important principle of personal data legislation, this means that organisations need to have a clear ground for processing or sharing personal data.

This ground for data processing can be a specific purpose alone (but this means that data cannot be shared), based on a strong generic purpose (such as a scientific purpose), or based on (informed) consent by the data subject. The cases do not show that specific purpose binding is considered a problem by organisations. In all use cases organisations are very careful to process data, which means that they are also careful in determining the purposes for data processing before asking consent. In case data are shared for a generic purpose, for instance for a societal goal, this needs to be controlled tightly. This was the reason for installing a review board in the case of the healthcare database. The healthcare database case also explained that that the costs of obtaining proper (informed) consent from the data subjects are expected to be higher than the revenues. Therefore, consent is usually obtained by having people accept general terms, which is not very elegant, nor does it have a strong legal basis. All cases hold that there are still many uncertainties involved in sharing data within a network of organisations.

An important challenge for data collaborations is the apparent incompatibility of data maximisation (the premise of big data) and data protection legislation. Combining datasets from different data holders may result in re-identification of individuals (Gopalkrishnan et al., 2012; Roosendaal, 2013). For example, Californian researchers were able to re-identify patients based on multiple open data sets (El Emam et al., 2012). Furthermore, European data protection legislation requires organisations to define clear and urgent goal to collect, store and apply data. When this pre-defined goal is achieved, the same legislation requires organisations to delete their data. The explorative nature of big data, however, implies a lack of pre-defined goals or applications and stimulates organisations to expand rather than delete datasets. Whereas European data protection legislation requires data minimisation, big data is based on the notion of data maximisation. While within a single organisation, control over data can be more easily exerted (Kathri and Brown, 2010), this is especially challenging in inter-organisational data collaborations. Further research should thus look into how coordination and control over data can be organised in inter-organisational data collaborations to allow for data sharing and foster innovation.

6 Conclusion

Organisations aiming to share data need to determine how data collaborations take place, which data they will share, and how data governance can ensure proper data sharing, which takes into account the sharing of commercially sensitive data that is compliant to data protection legislation. Based on literature, four modes of governance were identified: Market, Bazaar, Hierarchy, and Network. Subsequently, we explored four use cases that we expected to match to the four inter-organisational governance archetypes. We did not find an example of the Market governance mode in the use cases we examined. The organisations involved seemed to have difficulties in setting up a purely commercially viable model for cross-organisational data sharing as they want to retain control over their commercially sensitive data. Furthermore, we found that any data collaborations involving personal data need to put a hierarchical Governance mode in place for that specific purpose in order to retain control over the data. As the data maximisation notion behind big data may, thus, not be compatible with the data minimisation notion of data protection legislation, in order to spur innovation, further research should look into how control over data can be organised in inter-organisational data collaborations in order to allow for data sharing in a responsible manner.

References

Adler, P. S. (2001). Market, hierarchy, and trust: The knowledge economy and the future of capitalism. *Organisation Science*, 12(2), 215-234.

Applegate, L. M., Austin, R. D., and McFarlan, F. W., (2003). *Corporate Information Strategy and Management*, 6th ed., McGraw Hill, New York.

Bertot, J. C., and Choi, H. (2013). Big data and e-government: issues, policies, and recommendations. In: *Proceedings of the 14th Annual International Conference on Digital Government Research*, 1-10.

Borgman, C. L. (2012). The conundrum of sharing research data. Journal of the American Society for Information Science and Technology, 63(6), 1059-1078.

Dekker, H. C. (2004). Control of inter-organisational relationships: evidence on appropriation concerns and coordination requirements. *Accounting, Organisations and Society*, 29(1), 27-49.

Demil, B., and Lecocq, X. (2006). Neither market nor hierarchy nor network: The emergence of bazaar governance. *Organisation Studies*, 27(10), 1447-1466.

El Emam, K., Arbuckle, L., Koru, G., Eze, B., Gaudette, L., Neri, E., and Gluck, J. (2012). De-identification methods for open health data: the case of the Heritage Health Prize claims dataset. *Journal of Medical Internet Research*, 14(1).

Gopalkrishnan, V., Steier, D., Lewis, H., and Guszcza, J. (2012). Big data, big business: bridging the gap. *In Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, 7-11.

Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of Management Journal*, 38(1), 85-112.

Gulati, R., and Singh, H. (1998). The architecture of cooperation: Managing coordination costs and appropriation concerns in strategic alliances. *Administrative Science Quarterly*, 43(4), 781-814.

Ireland, R. D., Hitt, M. A., and Vaidyanath, D. (2002). Alliance management as a source of competitive advantage. *Journal of Management*, 28(3), 413-446.

Katal, A., Wazid, M., and Goudar, R. H. (2013). Big data: Issues, challenges, tools and Good practices. In *2013 Sixth International Conference on Contemporary Computing (IC3)*, 404-409.

Khatri, V., and Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148-152.

Klein, H.K. and Myers, M.D. (1999). A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *MIS Quarterly*, 23(1), 67-93.

Kumar, K., and Van Dissel, H. G. (1996). Sustainable collaboration: managing conflict and cooperation in interorganisational systems. *MIS Quarterly*, 279-300.

Lowndes, V., and Skelcher, C. (1998). The dynamics of multi-organisational partnerships: an analysis of changing modes of governance. *Public administration*, 76(2), 313-333.

Markus, M. L., and Bui, Q. N. (2012). Going concerns: the governance of interorganisational coordination hubs. *Journal of Management Information Systems*, 28(4), 163-198.

O'Toole Jr, L. J. (1997). Treating networks seriously: Practical and research-based agendas in public administration. *Public Administration Review*, 45-52.

Pardo, T. A., Gil-Garcia, J. R., and Burke, G. B. (2008). Governance structures in cross-boundary information sharing: Lessons from state and local criminal justice initiatives. *In proceedings of the 41st annual Hawaii International Conference on System Sciences*.

Powell, W. (2003). Neither market nor hierarchy. *The sociology of organisations: classic, contemporary, and critical readings*, 315, 104-117

Provan, K. G., and Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of public administration research and theory,* 18(2), 229-252.

Ring, P. S., and Van de Ven, A. H. (1994). Developmental processes of cooperative interorganisational relationships. *Academy of Management Review*, 19(1), 90-118.

Roosendaal, A. (2013). *De informatiefuik*. Business Contact, Amsterdam.

Sambamurthy, V., and Zmud, R. W. (1999). Arrangements for information technology governance: A theory of multiple contingencies. *MIS Quarterly*, 23(2), 261-290.

Spil, T. A., Broek, T. A, van den and Salmela, H. T. (2010). It Takes Two to Tango: The Fit Between Network Context and Inter-organisational Strategic Information Systems Planning. *International Journal of Strategic Information Technology and Applications*, 1(1), 23-41.

Stolze, C., Zarvić, N., and Thomas, O. (2011). Working in an inter-organisational context: The relevance of IT Governance and Business-IT Alignment. *International Journal of Computer Science and Information Security*, 9(8), 1-4.

Veenstra, A. F. van, and Broek, T. A. van den (2013). Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organisation. In: Wimmer, M.A., Janssen, M. & Scholl, H.J. eds. Electronic Government 2013, Koblenz, Germany. Lecture Notes in Computer Science (LNCS), 50–61.

Weill, P. (2004). Don’t Just Lead Govern: How Top-Performing Firms Govern IT, *MIS Quarterly Executive* (3)1, 1-17.

Zarvić, N., Stolze, C., Boehm, M., and Thomas, O. (2012). Dependency-based IT governance practices in inter-organisational collaborations: A graph-driven elaboration. *International Journal of Information Management*, 32(6), 541-549.