2024

Position Paper: Economics of Data Sharing

STUDY ON THE EMERGENCE AND CREATION OF VALUE WITHIN DATA-SHARING ECOSYSTEMS

ERIC BROUSSEAU

LUCAS EUSTACHE

JOELLE TOLEDANO

gaia-x INSTITUTE

Dauphine | PSL

CHAIRE GOUVERNANCE ET RÉGULATION

INTRODUCTION

The aim of this paper is to set out the analytical position adopted for a study of data sharing ecosystem as part of a collaboration between Paris Dauphine University and the Gaia-X Institute. It provides a first draft of ecosystem model, following on the study carried out in 2023[1], in the context of new European regulations (Data Act, Data Governance Act). This position paper thus provides a framework for a documented analysis of data ecosystems, and this analysis will be used to test the robustness of this framework, as well as to refine it. This research proposal will then have to be fed with empirical elements (data, interviews), in order to allow to test and calibrate the proposed theoretical framework.

The aim of this study is to gain a better understanding of data sharing ecosystems on two levels: that of participants, and that of data sharing ecosystem orchestrator

This dyadic approach provides a holistic view of data sharing ecosystems. We will analyze;

- The value that participants extract from the ecosystem. What interests and benefits they can derive from their participation in the ecosystem.
- The role of data orchestrators both as drivers of emergence of a data sharing ecosystem, and as service providers to stakeholders once the ecosystem is operating.

As a result, we will analyze, on the one hand, the participants in the ecosystem (which are usually both users and providers of data and related services), and, on the other hand, the orchestrators, who are organizing the ecosystem.

Data ecosystems bring together a range of stakeholders wishing to exchange data and complementary services. Theses ecosystems can be considered as "clubs", since to be viable they must enable participants to extract benefits from their contribution to the common pool of resources (i.e. shared data and derived services).

We will also investigate the diversity of orchestrators' business models: from pure technical facilitators to strategically integrated agents, depending on the organization of the value chains they address. Indeed, the ecosystem orchestrator might simply provide technical services (standards, provision of a platform, user/supplier catalog) or a set of commercial services (value-added services, sale of enriched data, etc.).

[1] Presentation available here: https://chairgovreg.fondation-dauphine.fr/en/ressources/1603/replay-data-sharing-europe-dga-and-da-legal-consensus-achievement-implementation

I. Data Sharing Ecosystem Organization

A data sharing ecosystem consists of two main types of actors: orchestrators and participants. Participants engage in bilateral or multilateral exchanges within the ecosystem. They are typically firms that provide or use data or value-added services within the ecosystem. The orchestrator is responsible for coordinating the data sharing ecosystem. This orchestrator can be a dominant firm within the value chain from which the ecosystem emerges, or it can be an intermediary appointed by the ecosystem's participants to manage coordination. This orchestrator can have different roles: it can be ‘strategic’ or ‘technical’, which will be developed later.

There can be multiple levels of orchestration within a data sharing ecosystem: ecosystem orchestration and use case orchestration. These two functions can be performed by the same entity or different ones. When there are different orchestrators, there is a division of competences. There will be a strategic orchestrator at the ecosystem level and a technical orchestrator designated for each use case.

The overall goal of the ecosystem is to generate use cases. A use case is the realization of value from data and service sharing among participants for the benefit of all stakeholders. This use case can be a finished product, like an application, the training of an AI, or the improvement of an existing product or process, such as enhanced maintenance services or increased efficiency in supply chain stock management.

There are two levels of economic analysis within the ecosystem. The first level pertains to the ecosystem orchestrator. The orchestrator must have a viable business model to sustain the ecosystem. One major challenge for the orchestrator is subsidizing the ecosystem's creation until it reaches critical mass. The orchestrator can have various revenue sources, such as a subscription model, commissions on transactions within the ecosystem, or subsidies (from the government or other entities).

The second level of economic analysis concerns the use case. A fundamental aspect is the distribution of value created by the use case (ensuring that no party captures all the benefits without reciprocation) and the sharing of the use case costs (particularly when a use case is financed exclusively by one of the stakeholders). The orchestrator partially addresses this issue by providing either a technical or strategic solution. The technical solution aims to reduce the cost of implementing the use case, for example, through standardization or offering a catalog of potential suppliers and users. The strategic solution, in addition to the technical one, might include cross-subsidization to facilitate participant collaboration in the use case (by offering a value-added service for free to the least incentivized parties).

II. Benefits for ecosystem participants: A gradual approach

One way of organizing data sharing, especially in the B2B market, is through data sharing ecosystems. These data ecosystems constitute a spectrum of "clubs". A club allows to provides goods that are non-rivalrous (i.e. goods which consumption by one agent does not prevent use by other agents) but which are excludable (i.e. goods which access to can be technically and economically efficiently prevented). Indeed, when access cannot be prevented, no one has incentives to contribute to the provision of the good. The creation of a club, which purpose is to discriminate between members and non-members, allows to avoid the “free rider” problem by restricting access to benefit derived from the good to closed group of participants who contribute to production (Sandler & Tschirhart 1997). This is one of the differences between data-sharing ecosystems and open data platforms, where data sets are freely accessible to all. The services provided by the orchestrator together with the shared data and the service derived from them and benefitting to the ecosystem’s members constitute the “club good” provided to the stakeholders of the data-sharing community. The club owes its existence to the willingness of its members to participate. Each participant must therefore contribute (Through the supply of data, and of value-added services or financially) to benefit from the service. Within the Skywise Ecosystem, for example, participants must commit to providing data to Airbus to benefit from the platform provided by the latter free of charge. The governance and the relationship among stakeholders in the club establish its ability to emerge and its sustainability, given the economic characteristics of the value chain in which it is implemented.

Participants will be encouraged to join an ecosystem, and thus contribute to it, if they derive benefits from it. There are several levels of benefits, from the most direct and tangible for participants to the most prospective. It is important to note that these benefits can be approached sequentially, at distinct stages of an ecosystem's maturity. A recent data ecosystem should focus more on the first direct levels of gain, to reach the critical mass of participants. Later, when both the ecosystem and the participants have developed a sufficient level of data-sharing maturity, they may discover less direct, while potentially more significant benefit. However, these benefits do not come on their own; they are associated with costs. Benefits are acquired from a greater integration of the stakeholders, which requires transformation of production processes and of firms’ organization. For examples on one hand the use cases that require weaker integration of stakeholders, could be compliance requiring information sharing (like traceability). On the other hands, certain use cases aimed at creating a new product, enabled by data sharing, require a high level of maturity and integration for stakeholders, as in the case of federated artificial intelligence training, for example.

It is important to specify that this costs and benefits typology does not apply uniformly to all participants and ecosystems. Depending on the type of use case, the level of maturity of stakeholders in their data governance, and their previous ability to collaborate, initiatives may start at an advanced level of the sequentially presented below. For example, this could be the case for collaboration on a pre-identified service enabled solely by data sharing, such as federated artificial intelligence training. Stakeholders can directly achieve a highly integrated and advanced level of benefit, depending on the nature of the use case and their own maturity related to data sharing and processes. The benefits and costs are presented in the following figure:

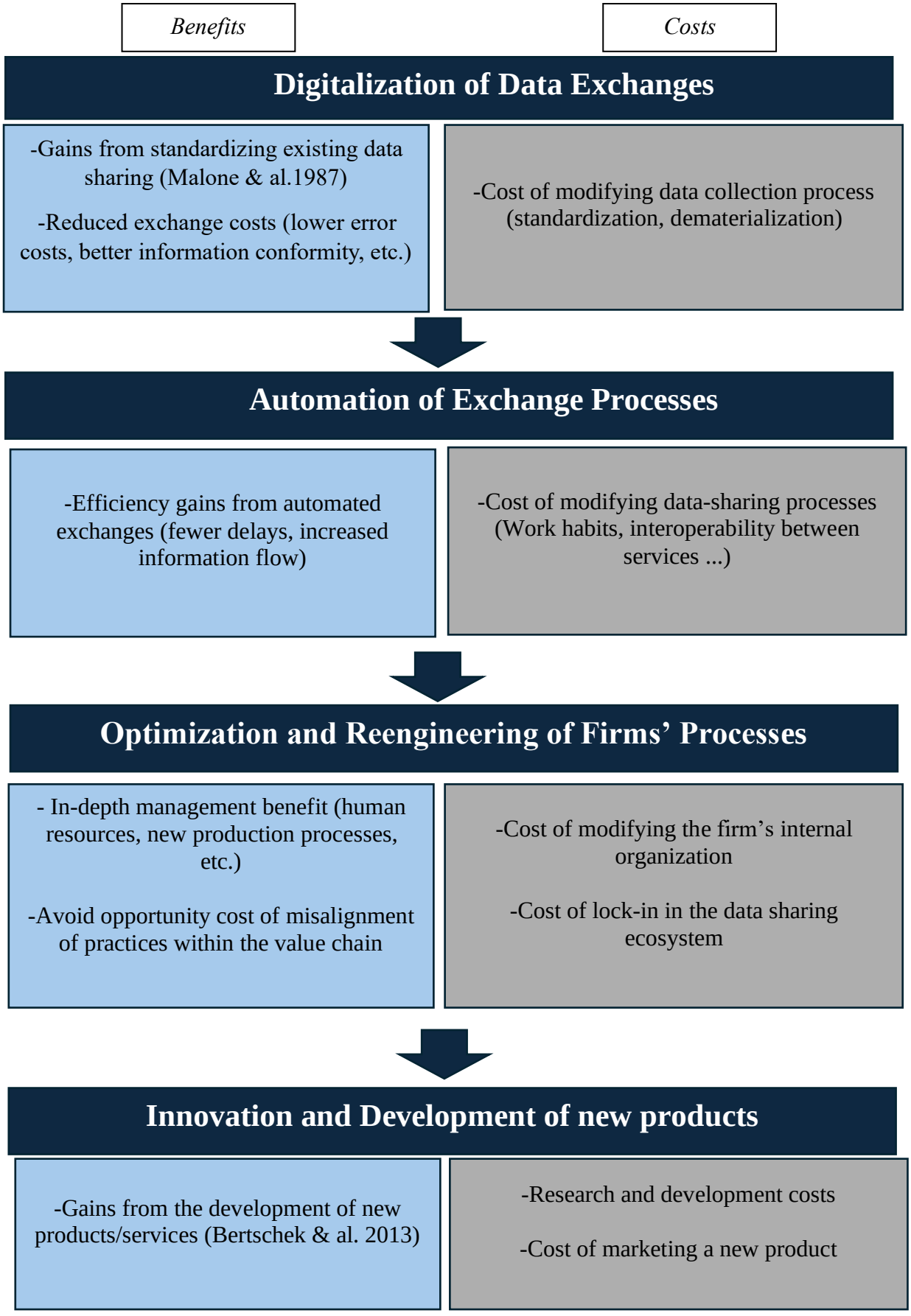


Figure 1 Representation of ecosystem participants' benefits

This perspective highlights two important points for the dynamic of data sharing ecosystem:

- The first level of benefits relates to cost savings and constitutes a direct and identifiable gain for all parties. Gains in innovation and new product development are unlikely to drive the emergence of a data-sharing ecosystem.
- The highest levels of value added request the costly integration and the reorganization of the value chain.

This framework to analyze categories of cost and benefit will enable us to classify ecosystem use cases to better understand how participants create value, which will also help us to refine the typology.

This ability to extract benefits depends in part on the structure of the ecosystem. In other words, the interaction between the structure of the value chains in which the participants are involved, and the characteristics of the agent in charge of orchestrating the ecosystem.

III. Governance and Business Model of Data Sharing Ecosystem Orchestrator : A Dynamic Perspective

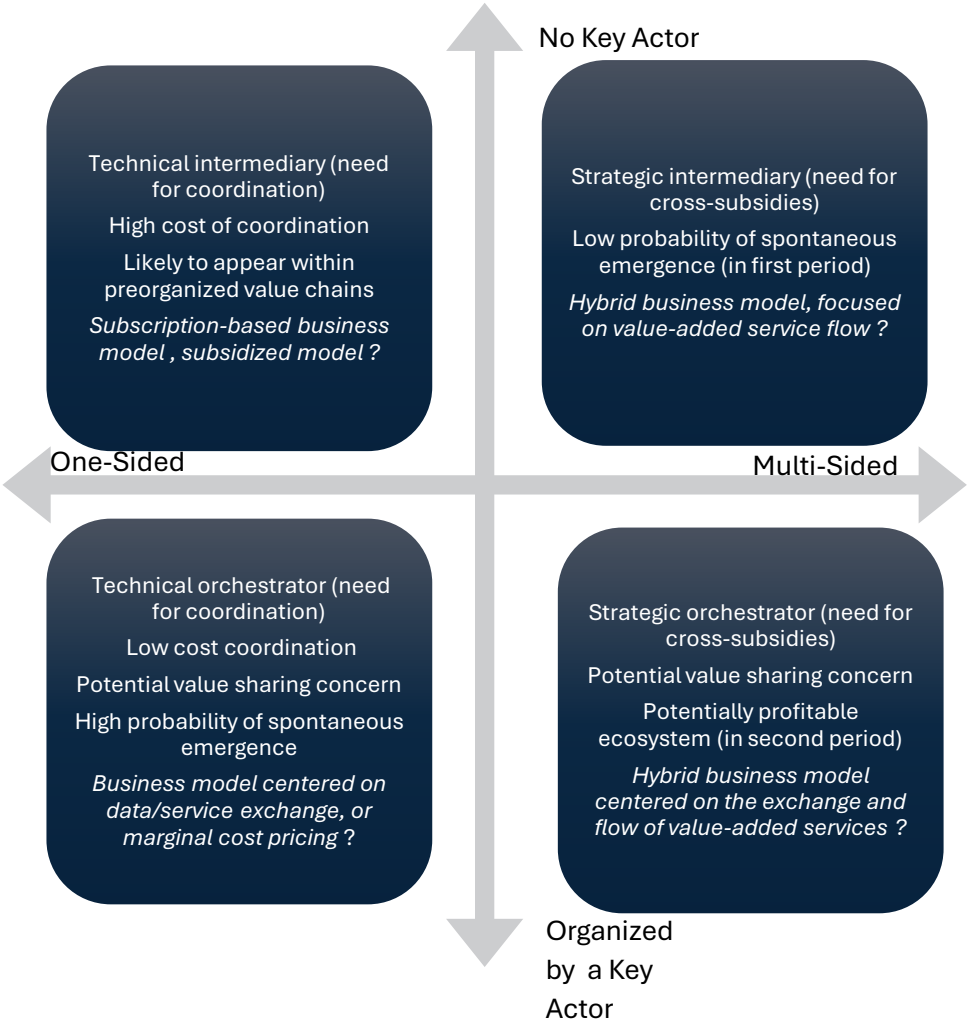
Data ecosystems are not uniform and may rely on various business and governance models, and result from alternative emergence processes. On the one hand Ecosystems can be built on well-established value chain in which some key players would act as ecosystem orchestrator. On the other hand Ecosystems can emerge in less organized value chains without a key player able to build and coordinate the stakeholders. It results a need for an intermediary to support the emergence and performance of the ecosystem. This heterogeneity stems from the characteristics of the various values chains. The aim of this section is to understand the relationship between value chain structure and ecosystem orchestrator (Belavina & Girotra 2012 and Brousseau & Glachant 2023), and which business model is best suited to a given type of organization. Ecosystems orchestrators can be represented in two dimensions that reflect the characteristics of the value chain:

- One sided or multi-sided orchestrator: The ecosystem is one sided when the orchestrator facilitate communication between the participant who form one distinctive group which exhibit same-side network effects (within their value chain) and have interchangeable roles (Staykova & Damsgaard 2014). If the need is the same and is identified by all the players, then the orchestrator will only be a technical intermediary, dealing one by one with each player in the value chain, with the latter not needing any strategic intervention from the orchestrator to conduct the data exchanges. For example, this could be the case of a value chain that encounters a common problem of product traceability. Once the need has been identified by all participants, orchestration provides only vector to carry out the data shared. However, if needs are diverse and the interests less aligned, then the orchestrator will need to be more a strategic intermediary. Parties with the least incentive to share data should be subsidized by those that benefit the most. This may be the case for a problem encountered by one part of the value chain, which needs data from another part, the latter having little or no incentive to resolve the former's need. It is important to consider the dynamic aspect of this dimension, where an ecosystem may initially be single sided before evolving, through the integration of new participants and new value chains, and diversification of the uses cases towards a multi-sided ecosystem.

- Atomicity of the ecosystem: This dimension highlights the presence of a key players who can orchestrate the ecosystem. A key player is a firm in a central position and or market power in at least one segment of the value chain, such that it is essential for the other players in the chain. The presence of a key player influences the ability of an ecosystem to emerge. However, while making the emergence of an ecosystem easier, a key player can also pose a problem, particularly when it comes to sharing value. The key actor's dominant position enables it to benefit from its market power to capture the value created by the ecosystem to mainly its advantage, which can disincentivize participants to join the ecosystem. In some cases, the ecosystem may be founded in

opposition to this key actor, to enable other stakeholders in the value chain to reduce its abuse of market power. In its absence or if participants are unwilling to take the risk inherent in its dominant position, they will have difficulties to coordinate at the emergence state. These two continuous dimensions form a matrix that will help us consider several perspective. In particular, the issue of spontaneity of emergence of data spaces and the issue of the value added by the orchestrator within the ecosystem (i.e., just technical facilitators serving to reduce transaction costs, or stakeholders offering value-added services, etc.).

The resulting matrix enables us to study the characteristics of value chains and their participants. Certain elements are obvious from the dimensions used, for example, the need



of coordination. Others correspond to hypotheses that need to be confronted with existing ecosystems, notably the question of the orchestrators' business model.

IV. Life cycle of Data Sharing Ecosystems

We can identify 3 major stages in the life cycle of an ecosystem. Its emergence, the reaching of critical mass and the diversification of its activity. It is important to take into account the dynamic dimension of data-sharing ecosystems. By the very definition of ecosystems, data sharing is a long-term process. In fact, this temporal dimension has effects on the organization of data ecosystems

This phase is relatively broad: from the gathering of a group of participants who decide to share their data in a structured way through an ecosystem, all the way to the establishment of an operational structure (not necessarily a viable one). This phase is crucial to the development of the initiative, and can be subject to opposing tensions. From the need for a closed group to create a competitive advantage for its members and foster mutual trust, to the need to make it as open as possible to remove barriers to entry for potential to reach a critical mass to be viable. The difficulties inherent at this stage, particularly in terms of overcoming initial fixed costs, call into question the need for external funding (public or not).

The second stage corresponds to reaching a sufficient number of use cases by increasing the integration and participation of existing members). Critical mass that determines the minimum level of activity to guarantee viability, i.e. reaching this mass is made possible by the network structure of data ecosystems. Increasing the number of participants increases at the same time the cost of coordination for the Ecosystem, as the potential value created. The viability afforded by reaching critical mass enables the ecosystem to reach a more advanced stage of maturity in its development, and to diversify.

Enlargement can be understood in several ways. It corresponds both to the diversification of use cases put forward by the ecosystem and to the enlargement of the set of participants (and their interests) (one-sided / multisided). This enlargement implies a change in the ecosystem's business model and organization and enables stakeholders to potentially benefit from higher value-added use cases. However, there are limits to this diversification, as the opening of the group is accompanied by an increase in coordination costs, as well as an increase in the need for cross-subsidies. The need for cross-subsidization is directly linked to the question of value sharing. Poor equalization of both costs and benefits limits firms' participation in the data ecosystem. This increase in costs may outweigh the benefits of diversification.

V. Need for “Trust”

This analysis by stage of maturity is necessary both from a descriptive point of view to understand the challenges of data ecosystems and their participants. In addition, from a normative point of view, this analysis enables us to understand the measures to be taken depending on the stage where the ecosystem and its participants are.

Trust is a central element in data-sharing ecosystems as it governs the interactions between participants. It can be defined as the willingness of participants to mitigate the risks inherent within the ecosystem. This trust is essential for addressing two major types of risks: technological risk and the risk of orchestrator opportunism.

The technological risk involves the potential for unauthorized access to or use of the data or services exchanged within the ecosystem. This risk is particularly heightened when the data in question hold strategic importance for the participants. Even with contractual safeguards in place, these protections may be deemed insufficient to compensate for potential losses in the event of a data security breach. Trust thus becomes a critical factor in enabling deeper integration of participants within the ecosystem.

The second risk pertains to the potential opportunism of the ecosystem orchestrator. This risk manifests in two distinct ways. First, the orchestrator makes investments in the form of specific assets, meaning technological and infrastructural resources designed exclusively for the ecosystem. These assets, by their nature, cannot be repurposed in other contexts without incurring significant costs. The orchestrator may engage in opportunistic behavior by leveraging these assets to create participant dependency on the ecosystem. As participants become more integrated, their ability to withdraw or diversify their relationships diminishes, thereby providing the orchestrator with the opportunity to exploit this increased dependency for profit. Second, during the ecosystem’s enlargement phase, the orchestrator may guide decisions to prioritize its own interests at the expense of other actors.

The neutrality of the orchestrator thus emerges as a crucial mechanism for ensuring trust among participants and preventing opportunistic behavior. A neutral orchestrator ensures impartial governance, which strengthens trust and encourages the integration of new actors within the ecosystem. This neutrality is manifested through the absence of incentives for the orchestrator to engage in such behaviors.

However, the neutrality of the orchestrator presents a challenge for economic viability. A neutral intermediary cannot generate significant profits without undermining participants' trust. Therefore, there exists a tension between the need for neutrality to maintain trust and the economic sustainability of the orchestrator's business model. This tension argues for either subsidies from ecosystem participants or public funding to ensure that the orchestrator has the necessary resources to support the ecosystem.

References

Belavina, E., & Girotra, K. (2012). The Relational Advantages of Intermediation. *Management Science*, *58*(9), 1614–1631.

Bertschek, I., Cerquera, D., & Klein, G. J. (2013). More bits – more bucks? Measuring the impact of broadband internet on firm performance. *Information Economics and Policy*, *25*(3), 190–203.

Brousseau, E., & Glachant, J.-M. (n.d.). Postscript: Institutions of International Economic Governance: Dynamics and Challenges. In E. Brousseau, J.-M. Glachant, & J. Sgard (Eds.), *The Oxford Handbook of Institutions of International Economic Governance and Market Regulation* (p. 0). Oxford University Press.

Malone, T. W., Yates, J., & Benjamin, R. I. (1987). Electronic markets and electronic hierarchies. *Communications of the ACM*, *30*(6), 484–497.

Sandler, T., & Tschirhart, J. (1997). Club theory: Thirty years later. *Public Choice*, *93*(3), 335–355.

Staykova, K., & Damsgaard, J. (2014). A Model of Digital Payment Infrastructure Formation and Development: The EU Regulator´s Perspective. *2014 International Conference on Mobile Business*.