future internet

MDPI

Article

Secure Internal Data Markets

Peter Kieseberg [1,*,†], Sebastian Schrittwieser [2,†] and Edgar Weippl [3]

[1] Institute of IT Security Research, St. Pölten University of Applied Sciences, 3100 St. Pölten, Austria
[2] Research Group Security and Privacy, University of Vienna, 1010 Vienna, Austria; sebastian.schrittwieser@univie.ac.at
[3] SBA Research, 1040 Vienna, Austria; eweippl@sba-research.org
* Correspondence: peter.kieseberg@fhstp.ac.at; Tel.: +43-660-3126291
† These authors contributed equally to this work.

Abstract: The data market concept has gained a lot of momentum in recent years, fuelled by initiatives to set up such markets, e.g., on the European level. Still, the typical data market concept aims at providing a centralised platform with all of its positive and negative side effects. Internal data markets, also called local or on-premise data markets, on the other hand, are set up to allow data trade inside an institution (e.g., between divisions of a large company) or between members of a small, well-defined consortium, thus allowing the remuneration of providing data inside these structures. Still, while research on securing global data markets has garnered some attention throughout recent years, the internal data markets have been treated as being more or less similar in this respect. In this paper, we outline the major differences between global and internal data markets with respect to security and why further research is required. Furthermore, we provide a fundamental model for a secure internal data market that can be used as a starting point for the generation of concrete internal data market models. Finally, we provide an overview on the research questions we deem most pressing in order to make the internal data market concept work securely, thus allowing for more widespread adoption.

Keywords: local data market; internal data market; on-premise data market; secure data market

Citation: Kieseberg, P.; Schrittwieser, S.; Weippl, E. Secure Internal Data Markets. *Future Internet* **2021**, *13*, 208. https://doi.org/10.3390/fi13080208

Academic Editor: Arcangelo Castiglione

Received: 12 July 2021
Accepted: 5 August 2021
Published: 12 August 2021

1. Introduction

The topic of legally compliant and ethically correct trade with data has increasingly moved into the focus of science and industry in recent years, especially with regard to the increased use of machine learning. "Data markets"—trading platforms where companies can offer their data (possibly also anonymised) and buy the data that they need—are seen as the infrastructure of the future in the sector of data-driven business and research (see [1] for one of the first platforms in this direction and [2] for an early modelling approach). For many organisations, especially small and medium-sized enterprises but also various research institutions, the design and implementation of their own data market is not possible, both from the point of view of the required resources and the lack of know-how, especially in the field of security. As a result, even very innovative companies are hindered in their efforts to innovate, especially if they are actually dependent on cooperation with others, e.g., external experts. For many of these companies, the currently very strongly represented concepts of large data markets also play only a very minor role, as they have several disadvantages. The most obvious of these is that the entire market does not run internally on company systems, but is designed as an external platform. This is absolutely unacceptable in many industries, especially in the field of medical products, but also in the banking sector or in the processing of personal data. In addition, many of these companies want to use a data market as a purely internal tool and do not want to have to offer the data to external parties. This is especially important between different departments of a company or in the environment of cooperative projects. Nevertheless, the currently popular approaches and initiatives to establish data markets are very much aimed at an

environment in which data are basically made available to everyone and, thus, follows a "one size fits all" approach. In the novel basic model provided in this paper, we provide a much more abstract view on the topic of internal data markets that allows specific tailoring to the requirements of the companies in question. This can range from security- and privacy-relevant aspects, such as data anonymisation stages, to issues of whether to charge money for data transfers, to resort to a karma point system, or to even not charge anything. In this paper, we provide the following contributions:

- A comparison of global and internal data markets with respect to security-relevant parameters from a general perspective.
- A novel abstract basic model for a secure internal data market, including audit and control.
- Flexible tailoring capabilities in the abstract model through the inclusion of pre- and post-processing capabilities.
- A set of open research questions that need consideration in order to make internal data markets work securely in environments with limited resources.

2. Background and Related Work

The collection and use of data have grown in popularity in recent years. While in the past, technical data were typically collected in the context of decision-making processes, e.g., in the context of predictive maintenance [3], in recent years, personal data have been increasingly used, especially in the area of marketing. A few large companies in particular control the entire market—above all, Google (or Alphabet). In recent years, there have been some initiatives to enable small- and medium-sized companies to use their data efficiently, especially to enable new products and services. In the health care sector, there are a number of approaches (e.g., dedicated research servers as in [4]) to make meaningful use of the data pools, whereby the involvement of external experts with all data protection-related problems, but also the use of distributed data that may not be combined centrally for legal or institutional reasons, repeatedly cause problems. In research, too, there are always results that could be passed on by the surveyors to other institutions, whether from research or industry, in order to, for example, be of use in other fields of application. Data markets (see [5] for a formal description) seem to be the appropriate answer to these needs. They allow the distribution of data, whether processed, anonymised or in raw form, as well as the development of a business model based on the transfer of one's own data. Therefore, there are already some initiatives aiming at the establishment of data markets; "Data Market Austria" [6] is a prominent example in which many national efforts are bundled. Similar initiatives also exist internationally, such as the IOTA Marketplace [7], the planned Food Data Market [8] or the targeted and GDPR-compliant search for patient data in the H2020 project FeatureCloud [9]. All these initiatives and concepts are strictly centralised and service oriented, i.e., a central operator operates a central market (also with distributed evaluation via block chains or distributed payment via Smart Contracts [10]). This does not necessarily mean that the data traded through these markets are stored on a central location; some concepts let the data reside with the data owners, and solely provide the brokering but, still, the service is centralised to some extent. In addition, services for translation into uniform metadata schemas [11] and the like are offered. However, the use case of these data markets always presupposes that organisations trade with other external organisations (under certain conditions) ready-made data sets and a quasi-free market is created in which services can be offered at a later stage. In [12], the authors also formally described the different roles and trust relationships within such a data market for the first time. The term "internal data market" was coined in [13] and provides a first, basic idea on remunerating the provision of data inside a large company or institution. Their definition of an internal data market is quite abstract and consists of three entities: the infrastructure, a buyer and a seller. It must be noted that their concept is not derived from global data markets, or any other form of data brokering with external partners, but from an ERP (Enterprise Resource Planning) perspective, alongside other types of

markets for services (e.g., marketing) inside companies [14]. Other authors [15] derive the definition of an internal data market from the data-sharing perspective and make a strong distinction between "external" and "internal" markets within an organisation, or, more generally, within a narrowly defined and consolidated consortium. Especially in the direction of the economic component of data markets, there are already some works on pricing [16] and fairness [17] amongst others, as well as work on establishing data markets for specific business models [18]. A somewhat dual concept to data markets, providing machine learning as a service, is also proposed in the literature (see e.g., [19]). Additionally, the topic of trading data independently from data markets has received some attention in the last decade [20].

In traditional IT security, attention on data markets has so far mainly focused on their usage for illegal deeds, particularly in the area of the Darknet and trading stolen data [21]. The trade-off between privacy and the return on investment in business models was also investigated in the past, as this is a field with huge potential for harm that also saw many changes with respect to the GDPR [22,23]. Furthermore, works on security in data markets, such as [24], typically focus on the privacy aspects of the people involved in the traded data sets, which is, of course, an important topic but does not cover the whole problem field of secure data markets. Other approaches focus more on the integrity of the market and the involved data, proposing, for example, blockchain-based solutions to the problem [10,25]. Related topics, such as data lakes [26], have been studied, for example, with regard to counterfeit protection in the event of data being changed by externally involved third parties [27]; however, these works relate more to global than internal markets with respect to their setup and especially the security issues surrounding them. Furthermore, the authors of [12] provided insight for general issues of data markets; still, the model proposed in this work is rather general and does not focus on issues specific to internal markets. The approach proposed in Section 4 differs thusly from this model in many key aspects.

3. Security Issues of Internal Data Markets

In this section, we provide an overview of the differences between internal and external data markets with respect to security issues. In order to identify the most pressing issues, we analyse publications on global data markets, as well as concepts from the implementation of such markets and transform them into an internal environment, i.e., all the parties in the selected market are portrayed to be within the same company, belonging to different divisions. Furthermore, we analyse the major issues of internal data sharing platforms that we gathered from our experience in this field and re-visit them in the context of an internal data market. While such an analysis is unlikely to be comprehensive, we do believe that we are capable of covering the most important aspects with respect to security. Table 1 provides an overview of the identified major differences, which are discussed in detail later on in this section.

Several of the outlined differences between global and internal data markets are directly based on the fact that internal data markets are typically an add-on and not the core property of an institution; thus, they do not receive the same attention as global markets by their respective system designers and administrators. In the following, we discuss some details.

3.1. Security Personal

Large infrastructures handling potentially critical material, such as global data markets that deal with valuable data and provide payment channels in order to reimburse for data proliferation, typically have a team of security experts dedicated to monitoring the platform and the surrounding ecosystem for (potential) threats and attacks, as well as constantly improving the security measures of the platform. Internal data markets, on the other hand, especially when introduced into smaller institutions, often do not have these resources. There might be a security team on hand, especially when the internal data market is set

up inside a larger company in order to handle data between different divisions; however, the data market is not the focus of this team, being just another application that needs to be handled.

Table 1. Security-related key differences between internal and global data markets.

	Global Data Markets	Internal Data Markets
Dedicated security personal	Typically available	Typically not available
System and Infrastructure	Modelled after DM requirements	Already existing infrastructure
Management capabilities	Can be patched/fixed/changed	External software
Volume	High	Diverse
Data intelligence	Little to none	Can be high
Users	Little control	High level of control possible
Market model	Generic	Can be specifically tailored
Interfaces	Control possible	Typically no control
Data locality	Typically low	Typically high
GDPR responsibility	None	Can be fully responsible
Institutional attention	DM is main focus	DM is a side issue

3.2. Integration

In the case of a global data market, the whole infrastructure and the system environment is modelled in order to accommodate the market software. Contrastingly, when small enterprises or small divisions inside a large company set up a data market, the market software is integrated into an already existing ecosystem, i.e., the design decisions of the surrounding system are not tailored for the market, but the market software will need to fit in. This is not only a deployment issue, but can also cause a lot of problems from a security perspective.

3.3. Market Management Capabilities

Global data markets typically develop their own software, which gives them great capabilities in patching and improving the system in case of security-related incidents, and lowers the response time to new threats. Internal markets are typically based on some off-the-shelf data warehousing solution, or bought from an external consultant, as the data market does not constitute the core business case of the company. Thus, like for every external software, the reaction time relies a lot on the external vendor.

3.4. Data Volume

Global data markets are built to accommodate a high volume of diverse data sets and have to be capable of managing large amounts of transactions. Internal data markets may be built to completely different specifications, thus also changing the attack surface and manageability of the system as a whole.

3.5. Knowledge on Data

Similar to the data volume, global data markets strive to be data agnostic, i.e., they either simply broker the data without local storage (see data locality), or provide a data storage engine that is as flexible as possible in order to be able to provide a wide variety of data sets and sources. This also reduces the capabilities for sanity checks and other security features on the data itself. This is especially reinforced in cases where the market only handles the data in an encrypted form. Internal markets, on the other hand, often do not need to be built as all-purposes markets, but only in order to deal with rather specific data sets. Furthermore, encryption of the data might not be an issue, depending on the design of the system and the entities involved. Thus, the platform could apply sanity checks and detection mechanisms against manipulation attacks in order to, for example, thwart adversarial machine learning [28].

3.6. User Model

Again, the global market needs to provide a rather low-level and flexible user model. Furthermore, it is outside of the control of the market, whether a user acts in different roles, for example, by signing up with several user accounts, in order to carry out attacks. In internal markets, depending on the setup and involved entities, access can be far more strictly monitored, often being combined with the standard user management.

3.7. Market Model

The market model of a global data market needs to allow for very flexible behaviour, which often includes the handling of complex policies and licenses. While this can also be an issue in an internal market, the complexity can be reduced in many internal cases of data exchange, e.g., by issuing company-wide policies outside the market system that automatically apply for any data shared between divisions. This can greatly reduce the complexity of the market and, thus, the attack surface. Furthermore, current data markets have a money-based remuneration model at their cores, while internal data markets could potentially be designed around alternative remuneration models, thereby omitting the payment and billing issue, further reducing the attack surface. Still, even in cases of monetary rewards for data sharing between divisions, this can be achieved outside the market, i.e., it is not functionality that necessarily has to be provided by the market software itself.

3.8. Interfaces

A global data market with a certain power, e.g., derived by its popularity, has some benefits with respect to interfaces: since data providers want to use it for making profits, it is likely that they will spent some efforts in order to provide the data in a form that is processable for the market software. This might not be the case for an internal market that is seen as an add-on with limited importance to the divisions that share the data—especially in cases where there is no monetary reward for the division sharing the data. Thus, the providers of an internal data market might need to develop their own interfaces to the (proprietary) data interfaces, potentially introducing the whole set of security issues surrounding data parsing [29]. On the other hand, depending on the data market use case, the data interfaces might be far simpler than the general-purpose interfaces required in global models, thus making it possible to apply such security measures as Language Security [30].

3.9. Data Locality

Some global data markets store data locally on dedicated databases; others, such as Data Market Austria (DMA), solely act as brokering platforms, i.e., they only store meta information on the data that are required to handle the brokering. However, the data transfer itself is completed outside of the market, basically leaving many issues, especially those related to security, to the users. In an internal market, especially when considering a small institution, data are typically local, regardless of whether the actual transfer function is an integral part of the market software, or is simply done through other means. This also opens up attack vectors through the database administrator responsible for the data store; thus, measures that take malicious actions from this side into account need to be put in place, including securing the Audit and Control mechanisms (see Section 4.2). Furthermore, this locality might also result in other security-related implications, especially in fields that require obeying certain regulations, such as BASEL or Sarbanes–Oxley (SOX) [31]. Contrary to a global market that acts data agnostic and only trades links, these issues have to be considered when designing a secure internal data market, as well as considering issues regarding the secure storage and transfer of the data itself.

3.10. GDPR Responsibility

In a global data market, all responsibilities regarding regulations, such as the General Data Protection Regulation [32], can be pushed to the respective users, i.e., data owners are responsible for the data that they provide to not be in violation of any of these regulations. The data market only provides the brokering and, sometimes, the transfer of (encrypted) data. Internal markets, on the other hand, have to take care of these regulations: as internal markets, they are typically owned by the same institution that holds the data, thus making it difficult to shift responsibilities to external entities. Furthermore, as internal markets are typically not the focus of data providers inside the company but rather a side issue, and thus, dedicated privacy experts are not available at the owner's side, they will expect the market to take on these issues.

3.11. Institutional Attention

Global data markets as we know them receive a lot of attention in their respective institutions, i.e., there are dedicated teams running them and resources dedicated to the tasks surrounding them. With internal data markets, this cannot be expected, as most institutions have a completely different focus, and the re-use of data through a data market is peripheral.

All these differences lead to a completely different attack surface and require different approaches for providing a secure data market. While there has been quite a lot of exploration on this topic regarding global markets, dedicated security research for internal markets is still missing. In Section 4, we provide a basic model for an internal data market that allows to mitigate most of the issues outlined above. However, some issues cannot be solved on a model basis, e.g., the problem of dedicated security personnel or a lack of focus inside the institution. Furthermore, the above results also show that internal data markets can be very different with respect to their requirements regarding security, and a dedicated security concept has to be designed for each internal market with respect to the requirements and surrounding ecosystem.

4. A Generalised Model for Secure Internal Data Markets

In this section, we provide a generalised model for a secure internal data market that overcomes the issues outlined in Section 3. This generalised model is designed as a starting point for the design of specific internal data markets, especially considering small institutions and companies. Furthermore, we discuss the issue of providing Audit and Control mechanisms throughout the whole data market in order to guarantee integrity of the data, where possible.

4.1. Model Overview

In this section, we discuss the proposed generalised model for internal market places. Figure 1 provides an overview of the model. Contrary to other models, e.g., that proposed in [12], in this basic model, we formally only distinguish between data owners and data consumers (users) and do not model Value Added Service Providers (VAS-providers) as specific identities. In our case, this does not represent any limitation in the modelling of so-called prosumers (who both provide and consume data), as well as external VAS providers, due to the following:

- Depending on the data record, prosumers act both on the owner's side and on the consumer's side, always according to the rules of the respective company. Therefore, it can also be modelled that a data owner acquires their own data, which can be useful with regard to the pre- and post-processing methods.
- VAS providers—actors who use data from others, develop it further by, for example, enriching it with other data, or perform additional calculations—also act in both roles: they purchase the data to be processed as consumers, and act as owners when passing on the data. Any kick-backs to the owners of the original data are simply handled via policies, just as they are when used by a normal consumer.

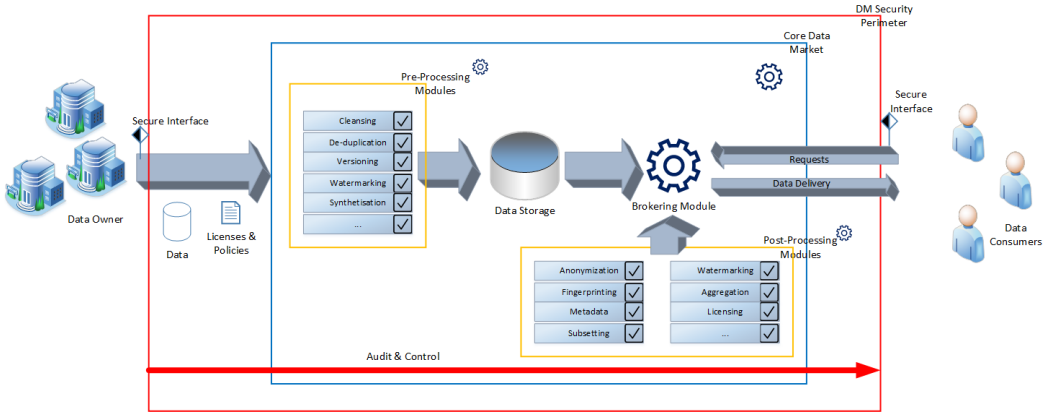

Figure 1. Generalised model overview.

The core data market (referred to as Core Data Market in the figure) is, therefore, in contrast to the basic model in [12], the sole data hub in our concept, i.e., we do not provide for any data traffic outside the brokerage platform as part of the data market model. This is especially important with regard to the security concept since this is virtually negated by a planned second communication channel outside of the brokerage (which, of course, does not mean that no further data transfer takes place, but rather that it does so outside the data market on your own responsibility). Additionally, the core data market in the basic model consists of three different parts:

- The storage entity, a secure, hardened and highly efficient database containing all data held in the data market.
- The brokering module, an engine that maps the trading of the data. This includes not only the provision of information about the data to enable a trade, but also the necessary methods for billing, payment and policy management. In addition, the brokering module has to call on the selected post-processing routines that need to process the data before actual delivery to the respective data recipient.
- The pre- and post-processing modules (marked in yellow in Figure 1) are of special importance for the pre- and post-processing of the data, as they are quasi internal VAS providers for important tools that a data market should provide. The main difference between these two types of methods is the time of application—the pre-processing methods are executed before the data is stored in the repository, whereas the post-processing methods are executed at the time of data delivery by the brokering module. While the pre-processing module do not need additional intelligence included, as they are invoked on the raw data and never run again on any data residing in the storage entity, the controller for the post-processing modules requires far more control over the whole brokering process; in many anonymisation techniques, it is of vital importance to keep track of the data that the specific recipient has already received at the time of provisioning with a new batch of data. This is especially important for fingerprinting, as receiving two differently fingerprinted copies of the same data set might enable a recipient to remove parts of the fingerprint. Especially when using fingerprinting methods based on anonymisation [33], this also requires tracking capabilities on the side of the post-processing modules.

Of course, interfaces are necessary to connect the data market to owners and consumers. While the exact definition of these interfaces is deeply rooted in the actual implementation of such a market and, therefore, cannot be provided here, the main important issue with respect to security is that the interfaces must be considered from a security perspective, as they are a major attack vector, e.g., by issuing injection attacks amongst others. Thus, in the proposed model, the interfaces are already inside the data market security perimeter.

4.2. Audit and Control

While most internal markets do not have the personal and technical resources for fully monitoring all events taking place in the market, it is advisable to introduce measures that allow for the ex-post detection of certain attacks, especially data manipulation. Thus, the model introduces an "Audit and Control" (A&C) mechanism that stores at least the following information on each data set:

- Hashes of the incoming data sets.
- Processing steps executed on the data in the pre-processing stage with the following:
  - Timestamp of the execution in order to be able to track back enrichments based on other data sets and enable reconstructing the respective versions.
  - Version number from version control, if available.
  - Hashes of input and output data sets.
  - Meta data on the execution, such as number of records in the data sets and user names in the case of manual execution.
- Access rights to the data in the data store.
- Post-processing steps executed on data delivery through the brokering module, including the same information as that for the pre-processing steps, as well as the request information from the data consumer.

From a technological perspective, there exist several methods for providing Audit and Control, e.g., blockchains, third-party controlled logging and other well-researched mechanisms for providing data and process integrity. While the integration of a transparent Audit and Control mechanism is quite straightforward for simple markets, where data are collected, stored and provided (perhaps including a fingerprint, watermark or just giving away an aggregated version), it is rather problematic to provide a transparent oversight in the case of more complex enrichment processes, especially in cases where classification based on neural networks is used, as the problem of explainability (or, rather, the lack of explainability) makes it currently impossible in most real-life use cases to determine what part of the data was actually used in the classification and is thus reflected in the delivered end result.

4.3. Comparison to Other Data Market Approaches

Much literature has been put forward in the past in order to provide practical approaches for data markets, especially considering market dynamics (e.g., [34]), but also focusing on security and privacy aspects (e.g., [35]). Several important approaches have been put forward in recent years, with a strong focus on using blockchains as a major backbone technology.

In [36], the authors propose a novel approach for a blockchain-based market called OmniLytics for matching research data with models that require training, while focusing on the issues of model privacy, data privacy and byzantine resistance. While the resulting model is highly distributed, and could be used as a local installation, it is developed with a very specific use case in mind, following a no-trust approach between the different partners. Still, it is far more specific when compared to the basic model outlined in our work, which is a positive aspect in terms of direct applicability, but also emphasises the key finding in our work that the current research is focusing on providing one specific model for data markets. As we outlined in Section 3, internal markets might have many different issues at hand, depending on the actual setup in question. OmniLytics answers a rather specific setup, which, from the basic idea, rather hints at a global market solution. This is also quite similar to [17], which is a different blockchain-based approach that focuses on issues of data availability, fairness in the payment process and user privacy, i.e., the data consumer must not be able to determine the real identity of the data provider. Again, this is a rather specific requirement, unsuitable for many applications, especially in internal data markets, where the identity of the data provider is (i) obvious based on the company structure and (ii) even required to be guaranteed, e.g., in order to guarantee the usability of

the data with respect to technical and legal questions. Again, these requirements might be useful in a global market perspective, but hardly in a local one. In [37], the authors provide a lot of insight in using blockchains for data markets, again focusing on the global case. BlockMarkchain [38] is another example of a very interesting approach for building a data market in a untrusted environment by using blockchain, especially focusing on performance issues. Still, these basic requirements are very untypical for many applications inside companies, as there typically is already a multitude of trusted and semi-trusted "third" parties involved, e.g., the IT department, management or the legal department, rendering the whole issue of no trust between participants negligible. Using blockchain (or other digital ledger technologies) especially makes sense in untrusted environments with no trusted third party available, as it can be seen in global data markets. From an internal data market perspective, i.e., from the perspective of a company that sets up an internal data market, this requirement most certainly does not hold for most application aspects (also see [15]). Still, in our basic approach, we also utilise a integrity protection measure, which can be based on a blockchain, especially in the Audit and Control (see Section 4.2). Here, the idea is not to push the whole trading mechanism to the blockchain, but merely to provide defence mechanisms against internal attackers, e.g., database administrators, as these roles are far more powerful in an internal market than an external counterpart, as they act as (semi-) trusted entities. Thus, the internal mechanisms of the market application need to defend the integrity of their data against a powerful internal adversary, in contrast to the approaches outlined above that focus on detecting market partners that try to game the market, either on the data or the payment side.

In related terms, the issue of privacy is a big topic in the area of data markets, as can also be seen in the approaches outlined above, where especially the privacy of the data provider is a typical design criterion. In [39], the authors also focus on the privacy protection measures utilised on the data itself, i.e., assuming that the traded data contain privacy-relevant information, they apply the concept of differential privacy and measure the data loss attributed to the data sets in order to (i) protect the privacy of the data subjects and (ii) find a balance between pricing and data quality. The technique is very interesting and could, of course, also be applied to certain internal markets; still, it is, again, very specific since (i) it assumes that privacy-relevant data are traded, which is not the case in many industrial environments, and (ii) that there are different partners involved in the systems, leading more to a global data market model rather than a local one. The same holds true for the slightly older PRIVATA approach [40], focusing on the balancing of data noise and pricing, which is rather specific and not relevant for many internal market concepts, where there exist a trusted authority and rules, authorizing which data could be used under what circumstances, as well as connected trading entities. Furthermore, pricing can be very different in local markets [15]; thus, we did not touch this topic in our work, as it is very use-case specific, and thus needs to be resolved on a per-market basis. In [35], the authors focus on providing a global data market solution, with much of the control being exerted by the data owner in this approach, compared to the data-sharing process.

While all the related techniques discussed in this section have great merit, they are not comparable to our work, as they (i) typically have a very specific set of requirements in mind and (ii) focus on global situations where the other partners are generally unknown (and often even protected against recognition by the approach itself). Still, the major contribution of our paper lies in outlining the many differences between local and global data markets in the area of security (also, ref. [15] provides insight on the differences between these two concepts, albeit omitting the topic of security at large); thus, while these approaches are useful in selected local scenarios, we focus on providing a very general and high-level general approach as a starting point for asking the right questions in the market design phase and providing a market specifically tailored to the needs of the specific company, consortium or even application in question (see [41] for a very specific and non-standard form of the "data market"), instead of providing a fixed market model that the partner(s) need to adapt to.

5. Open Research Questions

In this paper, we have provided an analysis of security issues in, as well as a basic model for, secure internal data markets. Still, there are several open research questions regarding this topic. The following section details the most important questions from our perspective.

5.1. Concrete Data Market Models

The basic model provided in Section 4.1 is rather abstract. In the case of implementing an internal market, this model needs to be specifically tailored to the needs of the data exchange and the side parameters of the actual ecosystem. This not only includes such issues as integration into the institution's infrastructure and process workflows, but also a specification on the billing, licensing and payment mechanisms (if applicable at all), as well as information on the data, such as the volume, complexity and time constraints. Furthermore, the concrete model also needs to specify what types of pre- and post-processing measures, e.g., fingerprinting, need to be in place, as well as how (and if) versioning of data sets is set up, especially considering that most technical companies already have some kind of versioning in place that needs to be respected in the data market. Using a concrete, specifically tailored model also provides many benefits from a security perspective, especially when the data to be shared are rather specific. The user base can be limited, thus allowing for very concrete measures with respect to access. Furthermore, the knowledge on the data itself can be used to provide additional mechanisms against data manipulation, as well as sanity checks. The main issue here is the generation of the specific models out of a general model and a set of (often unsystematic) requirements. Automating this process would greatly enhance the acceptance of such markets and reduce the setup costs; still, there is a lot of research work to be done, ranging from providing a taxonomy for the potential market owner to define and specify the requirements in a way that allows for automation, down to the automated model generation itself.

5.2. Flexible Security Model

Similar to the basic model, a general security model needs to cover a wide variety of different requirements and expectations, i.e., different markets will need different security models, depending on the entities and data sets involved, as well as side parameters derived from the surrounding ecosystem. Thus, the security model needs to provide added flexibility to cater to all these needs, which is not required in the case of a global data market, where all these parameters can be fixed beforehand.

5.3. Policies and Basic Market Model

Even though these topics are outside of the scope of our expertise, we do want to stress the issue that many internal markets might require a totally different market model, when compared to the global data market models available, which work as trading platforms between data owners and data consumers. In our opinion, the area of internal data markets allows for very interesting research in the field of alternative market models. Furthermore, the issue of policies needs further attention in internal data markets, as well as that of the effective enforcement of such policies.

5.4. Automated Deployment

Many institutions that could benefit from sharing data internally do not have the capabilities to model and implement such internal data markets. Thus, mechanisms for the automation of both the modelling phase as well as the deployment phase are required, including the issue of providing a secure environment that still allows for the required flexibility.

6. Conclusions

In this paper, we have analysed the differences between global and internal data markets with respect to security, outlining the issues that internal data markets face, but also the characteristics they have that could be used to potentially provide a higher level of security, especially considering the internal knowledge of the market owner on the traded data and the involved users. To the best of our knowledge, this is the first time global and local data markets were compared with respect to security in a structured way. Based on the results of this analysis, we see ample reasons for future research in this area, as most data markets concepts currently available typically focus on global approaches, even in cases where there is a limited number of foreseen partners. This, in our opinion, does not include all possible trust relationships between partners. Furthermore, the structure of the members and the partnerships has to be taken into account, especially considering such issues as insider attacks through highly connected partners. A special fundamental issue that needs to be taken into account as a conclusion is the status of an internal data market with respect to the rest of the institutions: while in the global market situation, the global data market is typically the core (business) unit, a local market will often only be a side attraction when compared to the core business of the company/ies in question. Thus, any issues surrounding this market will not receive the same level of attention as those in the global market scenario. On the other hand, there might be far better knowledge on the data handled through the market, providing more insight on the protection and mitigation mechanisms required.

Based on this analysis, we provided a basic model for a secure internal data market that can be used as a starting point for generating specific market models that are tailored to the needs of the institution. Here, we observe a huge need for providing mechanisms that support the semi-automated generation of instances of specific local data markets based on the outlined model. This includes supporting the analysis of partners and their (trust) relationships but also issues surrounding the technical setup. Finally, we are of the opinion that the data market concept can work for many companies on a local basis; however, a lot of additional research is required in order to make it practical for smaller institutions.

Author Contributions: Conceptualisation, S.S. and P.K.; methodology, S.S.; validation, E.W. and P.K.; investigation, P.K. and S.S.; writing—original draft preparation, P.K. and S.S.; writing—review and editing, E.W.; project administration, P.K.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Austrian Research Promotion Agency grant number 866880 (Big Data Analytics) and grant number 880592 (Secure Internal Data Markets).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

A&C	Audit and Control
GDPR	General Data Protection Regulation
SOX	Sarbanes–Oxley Act
VAS	Value Added Services

References

1. Meyer, M.H.; Zack, M.H. The Design and Development of Information Products. *Sloan Manag. Rev.* **1996**, *37*, 43–5 .
2. Wijnhoven, F. Models of information markets: Analysis of markets, identification of services, and design models. *Inf. Sci.* **2001**, *4*, 117–128. [CrossRef]
3. Daily, J.; Peterson, J. Predictive maintenance: How big data analysis can improve maintenance. In *Supply Chain Integration Challenges in Commercial Aerospace*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 267–278.

4. Popper, N.; Endel, F.; Mayer, R.; Bicher, M.; Glock, B. Planning Future Health: Developing Big Data and System Modelling Pipelines for Health System Research. *SNE Simul. Notes Eur.* **2017**, *27*, 203–208. [CrossRef]
5. Niyato, D.; Alsheikh, M.A.; Wang, P.; Kim, D.I.; Han, Z. Market model and optimal pricing scheme of big data and Internet of Things (IoT). In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6.
6. Ivanschitz, B.P.; Lampoltshammer, T.J.; Mireles, V.; Revenko, A.; Schlarb, S.; Thurnay, L. A Data Market with Decentralized Repositories. DeSemWeb@ISWC. 2018. Available online: https://openreview.net/pdf?id=rkgzBg7yeX (accessed on 11 May 2021).
7. The IOTA Marketplace. Available online: https://blog.iota.org/iota-data-marketplace-cb6be463ac7f (accessed on 11 May 2021).
8. Food Data Market. Available online: https://fundingbox.com/spaces/ledger-ledger-news-and-updates/5d8c6ed052317832f858fc59 (accessed on 11 May 2021).
9. The FeatureCloud Project. Available online: https://featurecloud.eu/ (accessed on 11 May 2021).
10. Schlarb, S.; Karl, R.; King, R.; Lampoltshammer, T.J.; Thurnay, L.; Ivanschitz, B.P.; Mireles, V. Using Blockchain Technology to Manage Membership and Legal Contracts in a Distributed Data Market. In Proceedings of the 2019 Sixth International Conference on Software Defined Systems (SDS), Rome, Italy, 10–13 June 2019; pp. 272–277.
11. Ivanschitz, B.P.; Lampoltshammer, T.J.; Mireles, V.; Revenko, A.; Schlarb, S.; Thurnay, L. *A Semantic Catalogue for the Data Market Austria*; SEMANTICS Posters & Demos: Vienna, Austri , 2018.
12. Horváth, M.; Buttyán, L. Problem domain analysis of iot-driven secure data markets. In *International ISCIS Security Workshop*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 57–67.
13. Koronios, A.; Redman, T.; Gao, J. Internal Data Markets: The Opportunity and First Steps. In Proceedings of the 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology, Beijing, China, 21–23 November 2009; pp. 127–130.
14. Lings, I.N. Internal marketing and supply chain management. *J. Serv. Mark.* **2000**, *14*, 27–4 . [CrossRef]
15. Fernandez, R.C.; Subramaniam, P.; Franklin, M.J. Data Market Platforms: Trading Data Assets to Solve Data Problems [Vision Paper]. *arXiv* **2020**, arXiv:2002.01047.
16. Liang, F.; Yu, W.; An, D.; Yang, Q.; Fu, X.; Zhao, W. A Survey on Big Data Market: Pricing, Trading and Protection. *IEEE Access* **2018**, *6*, 15132–15154. [CrossRef]
17. Zhao, Y.; Yu, Y.; Li, Y.; Han, G.; Du, X. Machine learning based privacy-preserving fair data trading in big data market. *Inf. Sci.* **2019**, *478*, 449–460. [CrossRef]
18. Agarwal, A.; Dahleh, M.; Sarkar, T. A marketplace for data: An algorithmic solution. In Proceedings of the 2019 ACM Conference on Economics and Computation, Phoenix, AZ, USA, 24–28 June 2019; pp. 701–726.
19. Joita, L.; Rana, O.F.; Freitag, F.; Chao, I.; Chacin, P.; Navarro, L.; Ardaiz, O. A catallactic market for data mining services. *Future Gener. Comput. Syst.* **2007**, *23*, 146–153. [CrossRef]
20. Lorenzo, B.; Gómez-Cuba, F.; García-Rois, J.; Gonzalez-Castano, F.J.; Burguillo, J.C. A microeconomic approach to data trading in user provided networks. In Proceedings of the 2015 IEEE Globecom Workshops (GC Wkshps), San Diego, CA, USA, 6–10 December 2015; IEEE: Piscataway, NJ, US , 2015; pp. 1–7.
21. Holt, T.J.; Smirnova, O.; Chua, Y.T. Exploring and Estimating the Revenues and Profits of Participants in Stolen Data Markets. *Deviant Behav.* **2016**, *37*, 353–367. [CrossRef]
22. Nget, R.; Cao, Y.; Yoshikawa, M. How to Balance Privacy and Money through Pricing Mechanism in Personal Data Market. *arXiv* **2017**, arXiv:1705.02982.
23. Elvy, S.A. Paying for privacy and the personal data economy. *Colum. Law Rev.* **2017**, *117*, 1369.
24. Keerthana, K.; Stefie, C.; Priyadharshini, R.; Veeralakshmi, P. Safe and Secure Data Markets using Merkle Hash Algorithm. *Int. J. Res. Eng. Sci. Manag.* **2019**, *2*, 94–96
25. Özyilmaz, K.R.; Doğan, M.; Yurdakul, A. IDMoB: IoT data marketplace on blockchain. In Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, Switzerland, 20–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 11–19.
26. Fang, H. Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem. In Proceedings of the 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), Shenyang, China, 8–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 820–824.
27. Kieseberg, P.; Schantl, J.; Frühwirt, P.; Weippl, E.R.; Holzinger, A. Witnesses for the Doctor in the Loop. In Proceedings of the International Conference on Brain Informatics and Health, London, UK, 30 August–2 September 2015; pp. 369–378.
28. Huang, L.; Joseph, A.D.; Nelson, B.; Rubinstein, B.I.; Tygar, J.D. Adversarial machine learning. In Proceedings of the 4th ACM workshop on Security and artificial intelligence, Chicago, IL, USA, 21 October 2011; pp. 43–58.
29. Poll, E. LangSec revisited: Input security flaws of the second kind. In Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 24 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 329–334.
30. Sassaman, L.; Patterson, M.L.; Bratus, S.; Locasto, M.E. Security applications of formal language theory. *IEEE Syst. J.* **2013**, *7*, 489–500. [CrossRef]
31. Sarbanes, P. Sarbanes-oxley act of 2002. In *The Public Company Accounting Reform and Investor Protection Act*; US Congress: Washington, DC, USA, 2002.

32. Regulation, G.D.P. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. *Off. J. Eur. Union OJ* **2016**, *59*, 294.
33. Kieseberg, P.; Schrittwieser, S.; Mulazzani, M.; Echizen, I.; Weippl, E. An algorithm for collusion-resistant anonymization and fingerprinting of sensitive microdata. *Electron. Mark.* **2014**, *24*, 113–124. [CrossRef]
34. Zhao, Y.; Wang, H.; Su, H.; Zhang, L.; Zhang, R.; Wang, D.; Xu, K. Understand love of variety in wireless data market under sponsored data plans. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 766–781. [CrossRef]
35. Bruschi, F.; Rana, V.; Pagani, A.; Sciuto, D. Acknowledging Value of Personal Information: A Privacy Aware Data Market for Health and Social Research. DLT@ ITASEC, 2020. Available online: http://ceur-ws.org/Vol-2580/DLT_2020_paper_6.pdf (accessed on 11 May 2021).
36. Liang, J.; Jiang, W.; Li, S. OmniLytics: A Blockchain-based Secure Data Market for Decentralized Machine Learning. *arXiv* **2021**, arXiv:2107.05252.
37. Khapre, S.P.; Dhasarathan, C.; Puviyarasi, T.; Goundar, S. Blockchain-Based Data Market (BCBDM) Framework for Security and Privacy: An Analysis. In *Applications of Big Data in Large-and Small-Scale Systems*; IGI Global: Hershey, PA, US , 2021; pp. 186–205.
38. Ehteram, H.; Toghani, M.T.; Maddah-Ali, M.A. BlockMarkchain: A Secure Decentralized Data Market with a Constant Load on the Blockchain. *arXiv* **2020**, arXiv:2003.11424.
39. Zheng, X. Data trading with differential privacy in data market. In Proceedings of the 2020 the 6th International Conference on Computing and Data Engineering, Sanya, China, 4–6 January 2020; pp. 112–115.
40. Jung, K.; Lee, J.; Park, K.; Park, S. PRIVATA: Differentially Private Data Market Framework using Negotiation-based Pricing Mechanism. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2897–2900.
41. Gadd, M.; Newman, P. The data market: Policies for decentralised visual localisation. *arXiv* **2018**, arXiv:1801.05607.