Business Process Modelling for a Data Exchange Platform

Christoph Quix[1,2], Arnab Chakrabarti[1], Sebastian Kleff[3], Jaroslav Pullmann[1]

[1] Fraunhofer Institute for Applied Information Technology FIT, Germany
[2] Databases and Information Systems, RWTH Aachen University, Germany
[3] Fraunhofer Institute for Material Flow and Logistics IML, Germany
[1]firstname.lastname@fit.fraunhofer.de
[3]firstname.lastname@iml.fraunhofer.de

Abstract. The digitization of companies and their business processes is a central component of Industry 4.0. Secure and trusted data exchange is crucial in this context, but providing data without sacrificing the control over their data is a challenge. The Industrial Data Space project seeks to define and implement such a platform supporting reliable, secure data exchange and governance enforcement among networked peers. In this paper, we report on our initial results in applying business process modeling for defining standardized processes in an open data exchange platform. We identified several participant roles and model their respective interactions. A transparent process modeling is expected to increase the trust in a data exchange platform and contribute to its acceptance and subsequent standardization.

1 Introduction

The advent of Industry 4.0 has seen an enormous growth in digitalized data. However, these digitalized data that are produced across various industries should not only be available vertically between value chains, but also distributed across horizontal organizational boundaries. The need for a platform for the secure cross-sectorial data exchange is motivated by the fact that the owners of the data want to retain control over their data. It should be transferred, distributed and processed in accordance with the explicit usage policies. In this way the data owner always determines the terms and the conditions of use for the data provided, thus maintaining the *data sovereignty* across the proposed platform.

The *Industrial Data Space*[4] [5] initiative originated in Germany to create such a platform, in which participants can exchange data securely and still keep the control over their data and maintaining their data sovereignty. Individual aspects of such a multifaceted goal (i.e., data management, semantic data description, integration, and usage policy enforcement) are addressed by dedicated Fraunhofer institutes that address the research and development (R&D) challenges and supply a prototypical implementation. In addition to the R&D project, the initiative is complemented by the Industrial Data Space Association[5] (IDSA) comprising more than 50 companies. They contribute

[4] http://www.industrialdataspace.de
[5] http://www.industrialdataspace.org

X. Franch J. Ralyté, R. Matulavičius, C. Salinesi, and R. Wieringa (Eds.):
CAiSE 2017 Forum and Doctoral Consortium Papers, pp. 153-160, 2017.

to the development of use cases, requirements, and a reference architecture, which is considered the main asset of the initial project phase. The key assumption underlying the platform architecture is an open peer-to-peer approach that enables a bilateral data exchange without any central data storage. As its main contribution, this paper presents the business architecture for the Industrial Data Space, identifying the major roles and modeling the involved processes in a formal way. The evaluation of the business architecture is based on several aspects.

1.1 Background and Related Work

A survey by IBM [1] on 25 industries from 93 countries, identified business intelligence as one of the four major technology trends in 2010. Thus the business world is going through a revolution induced by the use of data to control decision making and control analysis. One of the major reasons for this revolution in today's business processes is the rapid proliferation of the amount of data to be analyzed [3]. The changing role of data over time, from data as process to data as valued products, has lead to an increase in the need for inter as well as intra industrial exchange of data in a reliable manner.

Data as asset is also available for sale on online platforms popularly known as *Data Markets*. The emergence of this type of market place is creating enormous value for the data which is not anymore confined within the boundaries of a single organization. Both the producers and consumers are using the market place to trade their data. Even third party organizations are using the data marketplace to provide value added services like real time data analytics and predictive analytics to name a few [2]. Some of the most popular data marketplaces include Azure Data Marketplace[6] or the Qlik Data Market[7].

With the passing time there will be more data market places that are emerging in the market, but a data space, where participants can exchange data independent of the platform while maintaining the sovereignty over the data, is still missing. Data sovereignty is the important feature of our proposed data space model.

2 Architecture of the Industrial Data Space

The Industrial Data Space enables a reliable and secure exchange of data enforcing shared governance rules among participants. It has been conceptualized as an open architecture in which members interact and exchange data in a decentralized peer-to-peer manner. We briefly summarize the main features of the architecture in this section, more details can be found in the reference architecture [4]. Once registered, the participants of the Industrial Data Space may integrate and expose their data by using a *Connector*. Depending on the deployment model, the implementation of this logical component varies regarding the scale (embedded library, mobile application up to dedicated communication servers), functional coverage, and security level (e.g., as trusted platform module, with remote attestation etc.).

Regardless of the embodiment, any connector supports the Industrial Data Space protocol featuring identification, authentication and attestation as well as the secure

[6] `https://datamarket.azure.com/`
[7] `http://www.qlik.com/us/products/data-market`

transmission of data and metadata. The standardized metadata model extended by domain specific vocabularies (e.g., a taxonomy of steel grades) is a key resource allowing the brokerage and integration of enterprise data. Data sources exposed by a connector are advertised via a metadata set on a several *broker* components. This type of 'static' annotation supports the discovery and identification of relevant data sets by interested clients. Lightweight 'dynamic' metadata accompanies every data transaction indicating the provenance, target and usage restriction of the transmitted data to be enforced by the receiving connector.

The standard connector features application container management and message routing capabilities providing a runtime environment for *data apps*. Each application is executed in isolation, interacting with its environment as part of a data flow via explicitly declared and enabled interfaces. The data apps might integrate back-end systems ('system adapters') or implement a reusable data processing logic (e.g., aggregation, transformation, anonymization). The *AppStore* platform serves the distribution of properly annotated and certified applications. The metadata for the application captures, among others, its typed service interfaces and the semantic categorization linking to classes of data it operates upon. Thus, the Industrial Data Space provides a complete ecosystem for data brokerage, distributed provisioning, and processing.

A formal specification of the reference architecture is currently under development. It is planned to standardize the architecture in national and international standardization organizations (e.g., DIN and ISO).

3 Business Architecture of the Industrial Data Space

Participation in the Industrial Data Space requires the use of a software which is compliant with the Industrial Data Space reference architecture model. However, the Industrial Data Space is not limited to the software of a specific software provider as an open reference architecture model is proposed. This implies that a service in the Industrial Data Space can be provided by multiple organizations; this includes also general services in the Industrial Data Space infrastructure, such as a metadata broker or a digital distribution platform (often known as App Store). On the other hand, an organization might offer services that cover several roles.

Nevertheless, it is necessary to clarify the roles and their functions in such a network. In the following subsections, we will identify the different roles from a business perspective and describe the activities in which these roles are involved. This should contribute, on the one hand, to the business models employed by participants. However, our aim is only to model the processes within the network; business processes at the site of participators are not in our focus. On the other hand, the process models can be used the verify the technical architecture, e.g., whether all required interfaces between the components of the Industrial Data Space have been specified and whether all required information for running the business process is available. The later is the main goal for developing the business architecture.

The Fig. 1 shows the overall business architecture of the Industrial Data Space. The boxes represent the roles of participants; the arrows are (parts of) business processes which are described in more detail in section 4.

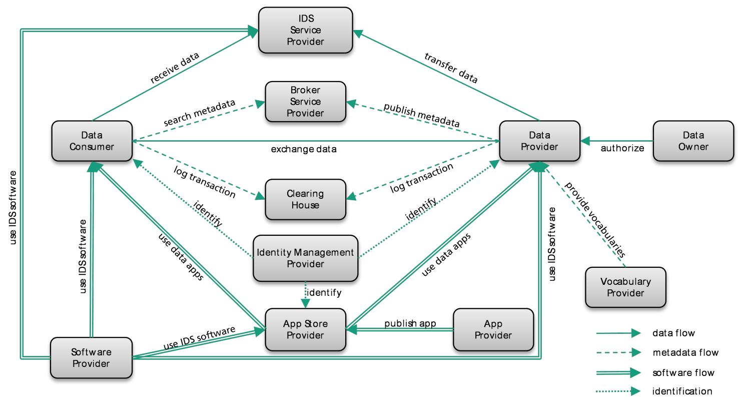

Fig. 1. Business Architecture of the Industrial Data Space

Data Owner: The data owner is legally the owner of the data. The owner might be different from the data provider in the case that the data is technically managed by a different entity. An example is a company which uses an IT service company for data management. Usually, the roles of the data owner and the data provider will be played by the same organization. The only activity of the data owner is the authorization of a data provider to publish the data owned by the data owner.

Data Provider: The data provider is an organization that manages data to be published in the Industrial Data Space. The data provider usually owns the data, but it might be also authorized by the data owner (see above). The data provider publishes metadata at a broker and exchanges data with a data consumer. Exchanging data with a data consumer is the main activity of a data provider. Usually, a broker is required to establish the connection between a consumer and a provider. However, they can also establish their connection by different means without involving a broker. The data exchange is described in more detail in section 4.

Data Consumer: The data consumer is an organization that receives data from a data provider. From a business process modeling point of view, it is the mirror entity to a provider. Thus, the activities are similar to the activities of the data provider.

Broker Service Provider: The main feature of the broker is the management of a metadata repository that provides information about the data sources available in an Industrial Data Space. There can be multiple providers of broker services (e.g., for different application domains) as the role of the broker is central, but non-exclusive. Other intermediary roles in the Industrial Data Space (e.g., clearing house, identity provider) can be played by the same organization that also offers broker services. Nevertheless, it is important to distinguish the roles from the organization, i.e., the broker service provider role deals only with metadata management, whereas the organization acting as broker service provider can also act as, for example, clearing house.

Clearing House: The clearing activities have been separated from the broker service as these activities are technically different from maintaining a metadata repository. As stated above, it might be still possible that the roles clearing house and broker service provider are played by the same organization, as they need to act as a trusted, intermediate entity between data provider and data consumer.

The clearing house should log the actions of a data exchange. After a data exchange has been completed, both data provider and data consumer need to confirm the transmission and the reception of the data by logging the transaction at the clearing house. Based on the logged data, a billing of the transactions can be performed. The log information can also be used to resolve conflicts (e.g., whether a data package has been received or not).

Identity Management Provider: For a secure operation and to avoid unauthorized access to data in the Industrial Data Space, there must be a service to verify identities. An identity needs to be described by a set of properties, e.g., that characterizes the role of the identity within an organization.

App Store Provider: The App Store provides applications that can be deployed in the Industrial Data Space to enrich the data processing workflows. The App Store Provider is responsible for managing data apps that have been provided by app developers. App developers should describe their data apps with metadata according to a metadata model describing the semantics of the services. The App Store should provide interfaces for publishing and retrieving data apps and their metadata.

Vocabulary Provider: The vocabulary provider manages and offers metadata sets (e.g., vocabularies, ontologies, reference data models, etc.) that can be used to describe data sets. In particular, the IDS Vocabulary will be provided by this role. Also other (domain-specific) vocabularies can be provided.

App Providers: App providers develop data apps to be used in Industrial Data Space. To be deployable in the Industrial Data Space, the Data Apps need to be published in an App Store. The data apps should include metadata that describe the data app (e.g., its functionality and the interfaces).

Software Providers: Software providers offer software that realizes the required functionality of the Industrial Data Space architecture. In contrast to the data apps, these software packages will not be provided by the App Store, but individual agreements between the software providers and their users (e.g., data consumers, data providers, broker service providers) have to be done. However, these agreements are outside the scope of the Industrial Data Space.

IDS Service Provider: If a participant of the Industrial Data Space does not deploy the required technical infrastructure to participate in the Industrial Data Space by itself, it can transfer the data to a service provider which hosts the required infrastructure for other organizations. This Industrial Data Space Service Provider plays then the role of a data provider, data consumer, broker, etc. and can perform the corresponding activities.

Furthermore, service providers that offer additional services to improve data in the Industrial Data Space are also covered by this role. Examples for such services are data analysis, data integration, data cleaning, or semantic enrichment. From a technical point of view, these service providers can be also seen as data providers and data consumers at the same time, e.g., they receive data as a data consumer from some data provider in the Industrial Data Space, apply their value-added service, and then offer the data in the Industrial Data Space as a data provider.

4 Modeling of Processes in the Industrial Data Space

This section describes the modeling of the processes using the BPMN language[8]. As stated before, we take a more technical perspective for modeling the processes and focus on the components and the exchange messages. Due to the limited space, we will can only show the main process which is the data exchange. The data exchange process involves four participants, namely the data provider, the data consumer, the clearing house, and the broker. The BPMN model is illustrated in Fig. 2.

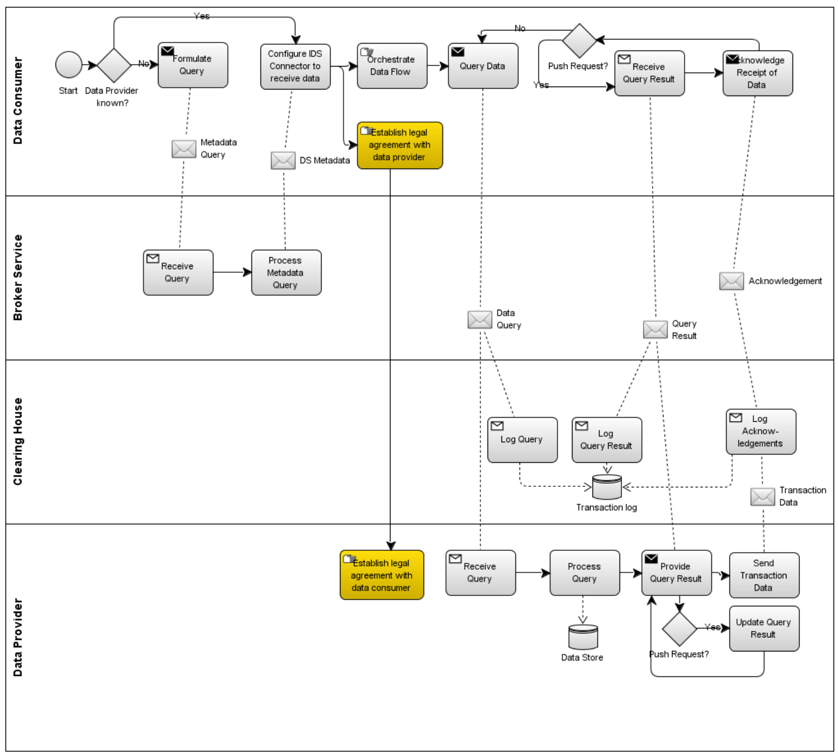

Fig. 2. Process Model for Exchanging Data

The process is initiated by the data consumer who will search for data using the broker. This is optional as the consumer might already know the provider. After selecting and receiving metadata of a data store, the consumer has to establish a legal agreement with the data provider. This might not be necessary in case of open data, or if the data exchange has already been regulated. The negotiation of this agreement is currently out of the scope of our work.

[8] `http://www.bpmn.org/` Business Process Model and Notation. We use this language as formal, yet easy to understand in its graphical representation and well supported by various tools. We used the open-source editor Yaoqiang `http://bpmn.sourceforge.net/`.

After the agreement has been established, the connector of the data consumer needs to be established. Part of this activity is the configuration of a data flow in which the received data will be integrated. After all prerequisites are fulfilled, the actual data exchange process can be initiated by the data consumer querying data from the remote connector of the data provider. The query is then processed by the connector of the provider, and the result is sent back to the data consumer. Communication between the connectors can be asynchronous; i.e., the data consumer will be notified by the data provider as soon as the result is available. Instead of a pull request, a push request can be sent, which means that the data consumer asks for updates regarding the requested data. The updated query results can be provided either after certain events (e.g., after the data has been updated by the data provider) or within certain time intervals. If a push request is made, the data consumer repeatedly receives updated query results from the provider. In case of a pull request, the data consumer can repeat the last part of the process to query data again (using the same or a different query).

The final step of the process is the logging the successful completion of the transaction. For that, both the data consumer and the data provider must send a message to the clearing house, confirming the transaction was successfully completed.

5 Evaluation of the Business Architecture

The evaluation of the business architecture and of the architecture in general is an ongoing activity within the Industrial Data Space initiative. So far, the evaluation focused on the following aspects:

Use Cases: Within the Industrial Data Space project, we are implementing several reference use cases which are based on the requirements gathered from the members of the IDSA. The business architecture can be considered as an abstract model for the use cases, i.e., the use cases should be mapped to the roles and processes of the business architecture. The use cases are from the domains supply chain management and production. Further use cases will be considered in the near future in domain-specific variants (e.g., Medical Data Space [9]). The implementation of the use cases is important feedback in the evolution of the business architecture.

This affected especially some details of the modeled processes, e.g., the distinction of push and pull activities. Also, some additional roles have been introduced to guarantee the required level of security, e.g., the identity provider.

IDSA Working Groups: Several working groups with participation of major companies have been established in the IDSA. Two working groups are relevant for the business architecture: the architecture group reviews the reference architecture model of the Industrial Data Space, which includes the presented business architecture. Another group discusses potential business models for the roles that have been identified in the business architecture.

The discussion in the working groups helped especially to clarify the roles of the business architecture.

Comparison with other Business Architectures: We have compared the business architecture of the Industrial Data Space with some other models from other domains:

[9] `http://medicaldataspace.de`

the general architecture of the Internet, the organization of an electrical energy supply network, and the deposit system for one-way bottles in Germany.

The Internet architecture is very similar to our business architecture, as it is also an open peer-to-peer architecture without centralized control. The Internet association ISOC has a similar role as the IDSA.

The electrical network has an interesting model for billing. Different providers can provide energy to the network and are reimbursed based on the amount they provided. Similar models must be also established in the Industrial Data Space.

The interesting aspects of the German deposit system for one-way bottles are the clearing house and certification of the participants to avoid misuse. This is also relevant in our case; a detailed model for certification is also under development.

6 Conclusion

In this paper we have provided a business architecture for a reliable data space platform where stakeholders can exchange data without losing its sovereignty. We described how the business processes that we presented in the paper is being developed as the Industrial Data Space platform and can contribute to the realization of the idea of Industry 4.0. The next steps are the refinement of the business architecture and detailed modeling of further processes within the Industrial Data Space. This needs to be verified with the ongoing implementations of the use cases and domain-specific variants of the Industrial Data Space, as indicated in section 5.

Acknowledgements: This work has been funded by the German Federal Ministry of Education and Research (BMBF) (project InDaSpace, grant no. 01IS15054). We thank our project partners for their comments on earlier versions of the business architecture.

References

1. J. P. Buerck, S. P. Mudigonda, S. E. Mooshegian, K. Collins, N. Grimm, K. Bonney, H. Kombrink. Predicting non-traditional student learning outcomes using data analytics-a pilot research study. *Journal of Computing Sciences in Colleges*, **28**(5):260–265, 2013.
2. J. Deichmann, K. Heineke, T. Reinbacher, D. Wee. Creating a successful Internet of Things data marketplace. *McKinsey & Company*, 2016. `http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/creating-a-successful-internet-of-things-data-marketplace`.
3. V. Gopalkrishnan, D. Steier, H. Lewis, J. Guszcza. Big data, big business: bridging the gap. In *Proc. Intl. Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, pp. 7–11. ACM, 2012.
4. B. Otto, et al. Reference Architecture Model for the Industrial Data Space. *Technical report*, Fraunhofer-Gesellschaft, 2017. `http://www.industrialdataspace.de`.
5. B. Otto, Jürjens, J. Schon, S. Auer, N. Menz, S. Wenzel, J. Cirullies. Industrial Data Space - Digital Sovereignity over Data. *Whitepaper*, Fraunhofer-Gesellschaft, 2016. `http://www.industrialdataspace.de`.