Towards Solid-based B2B Data Value Chains

Andreas Both[1,2], Dustin Yeboah[1], Thorsten Kastner[1], Daniel Schraudner[3], Sebastian Schmid[3], Christoph Braun[4], Andreas Harth[3], and Tobias Käfer[4]

[1] DATEV eG, Nuremberg, Germany
[2] Leipzig University of Applied Sciences, Leipzig, Germany
[3] Friedrich-Alexander University (FAU), Nuremberg, Germany
[4] Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract. In the paper, we describe the data sharing within a data-driven B2B ecosystem using Solid technologies. Using a real-world use case, we describe our approach and implementation, as well as the (previously non-existent) referencing of a purpose ontology to fully express the intended use of shared data in a machine-readable and analyzable form. We follow the intention of establishing a safe, robust, and traceable linked-data-driven B2B ecosystem based on Solid, where high standards of data privacy and thriftiness have to be met.

Keywords: Solid · Data Value Chains · Zero Trust · Data Sovereignty

1 Introduction

In this section, we will describe the use case of data-based ecosystems (data value chain, cf. [2]), Solid technologies [4], and the challenges in such ecosystems (in particular, regarding GDPR [7] and zero-trust architectures [6,5]) that are unique regarding the strong need for a technological solution that enables data-driven value chains and could fulfill the requirements for safety, robustness, and traceability. In particular, Solid provides a standardized machine-processable process for data exchanged based on Linked Data. In addition, the separation of data, applications, and identities facilitates the simple reuse of software components. The standardization of process steps in the form of semantic data representations and linked-data-driven implementations is therefore not limited to one use case but has a leverage effect for the entire industrial landscape.

Here, we focus on B2B data collaborations as a central use case of all businesses. Figure 1 shows an example of typical data exchange operations in B2B collaborations, where a small or medium enterprise (SME) requires data from its tax advisory office (TAO) to hand it over to the bank for creating a credit offer (that would finally be handed as data to the SME). There, only the directly communicating actors know about each other and are not aware of any other hidden participants in the data value chain. In general, in such an ecosystem, the collaboration of organizational actors is expressed while providing data to a service provider to perform computations (i.e., contract data processing) and provide data to the original data provider (process and return) or hand the data

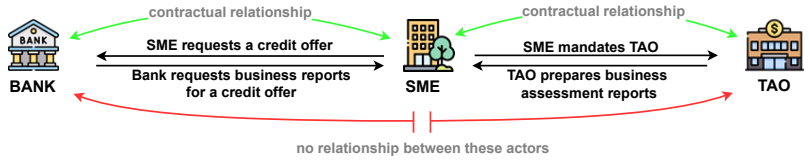

Fig. 1: Use case: Personas and contractual relationships.

safely to another partner (data transmission). The automation of such processes is crucial for businesses. Therefore, several technical challenges have to be fulfilled to be able to fully exploit the potential of data-driven services: (C_1) All data exchange process data has to be represented machine-readable, s.t., the process can be automated (e.g., while asking automatically for a data instance of a particular data format); (C_2) For the exchanged data additional metadata needs to be provided, s.t., the legal requirements are fulfilled (e.g., to provide the purpose of the data exchange). Solid technologies already provide some standards to address both challenges. In this paper, we describe our solution to extend the process by a standardized process driven by Linked Data and enable a universal and integrated solution supporting end-to-end data value chains. Hence, we provide here a (possible) cornerstone for a data-driven B2B ecosystem in the form of requirements, modeling, and a Solid app implementation.

2 Solid-based Data Value Chains

Linked Data is a profound solution to represent semantic data on the web and in enterprises. However, while publishing and interlinking data works well in the context of public data (e.g., public company register [3]), the same approaches cannot be applied to B2B environments where business partners cannot share all data nor are allowed to reveal information about their business partners that are part of their data providers (actually, this strong data protection requirements often apply to units inside companies too, e.g., in case of a data trustee scenario). Therefore, a technology is needed that enables the exchange of data on demand, s.t., the B2B potential of Linked Data can be exploited. The W3C Solid Community Group's Application Interoperability [1] (`INTEROP`) provides a method to describe data representations and data access within so-called Solid Pods (Web storage) with information about the access grants and the affected data items. Following [7] all data processing needs a valid data processing purpose P (cf. C_2) that represents the minimum possible and necessary rights; typically that is also true within business environments extended with additional constraints (e.g., data re-sharing only within a particular business domain or administrative regions); the `INTEROP` vocabulary is not providing such a resolvable property.

Given our exemplary scenario (cf. Figure 2), the TAO has provided data to the SME for a particular purpose of data processing P_0. The SME has provided the data to the Bank also defining a purpose P_1. Additionally, re-sharing data of the TAO to the Bank is not allowed to reveal information about the TAO as this might uncover business secrets of the SME. `INTEROP` is standardizing only the direct data exchange processes.

Fig. 2: Chain of sharing business assessment reports.

We can derive two technical requirements for future data value chains. (RQ1) The data purpose needs to be represented in an unambiguous, machine-readable form. Accordingly, the following must always apply to all data shared with P_i and passed on with P_j: P_j is equal to or more restrictive than P_i. (RQ2) Providing the received data to another business partner has to hide the original source (i.e., each identifier needs to be concealed), s.t., such data is treated in the same way as data produced by the data provider internally. However, the data receivers need to document if and how data was re-distributed to third parties (traceability, e.g., as evidence for external auditors of the data value chain).

To fulfill the aforementioned requirements, we model (see the complete modeling in our online appendix[5]) the following data representations and extend the process with the intention of establishing end-to-end data value chains within a B2B environment. For RQ1 we have extended the INTEROP classes for data sharing and data access with the predicate `gdprp:purposeForSharing` linking to a purpose class, in order to facilitate documentation and passing on purposes throughout the data chain. RQ2 is addressed by the classes `delegatedAccessAuthorization` and `delegatedDataAuthorization` to indicate that the given access and data grants are based on grants given by a third party. Further, the additional class `FacadeDataRegistration` enables hiding the Data Registration of the original data provider when passing on data and access authorizations. As shown in Figure 3 (cf. online appendix), the modeled data enables businesses to fulfill the requirements and enables automatic processing as all metadata of the data value chain is represented in a machine-readable, semantic form. The implementation of the extended Solid-based data sharing is available online[6] and can be used to establish conform B2B data value chains.

3 Discussion and Conclusions

In this paper, we addressed the challenges of establishing B2B data-driven ecosystems. Our modeling and implementations aim to enable end-to-end processing of B2B data sharing/re-distributing while leaving the data granting control to the users (cf. Data Sovereignty). Following our approach, data value chains would be represented completely by Linked Data, enabling businesses to automatically process data while still providing crucial features like traceability and protection of business secrets. While doing so, we address a core blocker for applying Linked Data in distributed data-driven ecosystems with a strong interest in data protection and business secrets (e.g., supply chains, data value-added service). Our goal is to grow this solution into a global B2B ecosystem of data stores

[5] https://doi.org/10.6084/m9.figshare.25424635
[6] https://purl.archive.org/mandatb2b/ESWC2024

Fig. 3: Solid compatible chain of sharing business assessment reports.

(i.e., Solid Pods) with linked data that can be connected on demand to help companies thrive through collaboration based on their semantic data.

Although our approach is applicable and represents a cornerstone for such ecosystems, further research is needed to achieve this goal, e.g., the reasoning on instances of purposes ontology needs to be safe and sound to establish completely automatable processes. Another issue is the redirect chain that is enforced by hiding (cf. `FacadeDataRegistration`) the original source of data items, which might lead to implementing a data trustee for better scalability and availability.

Acknowledgments This work has been supported in part by the German ministry BMBF under grant 16DTM107B (*MANDAT*).

References

1. Bingham, J., Prud'hommeaux, E., Pavlik, E.: Solid Application Interoperability. W3C Editor's Draft. Nov. 2023
2. Kano, L., Tsang, E.W.K., Yeung, H.W.c.: Global value chains: A review of the multi-disciplinary literature. J. of International Business Studies **51**(4) (2020)
3. Roman, D., Alexiev, V., Paniagua, J., Elvesæter, B., von Zernichow, B.M., Soylu, A., Simeonov, B., Taggart, C.: The euBusinessGraph ontology: A lightweight ontology for harmonizing basic company information. Semantic Web **13**, 41–68 (2022), 1
4. Sambra, A.V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., Zagidulin, D., Aboulnaga, A., Berners-Lee, T.: Solid: a platform for decentralized social applications based on linked data. MIT CSAIL & Qatar Computing Research Institute, Tech. Rep. (2016)
5. Shore, M., Zeadally, S., Keshariya, A.: Zero trust: The what, how, why, and when. Computer **54**(11), 26–35 (2021)
6. Stafford, V.: Zero trust architecture. NIST special publication **800**, 207 (2020)
7. The European Parliament and the Council of the European Union: Regulation (EU) 2016/679 (General Data Protection Regulation) GDPR, https://gdpr-info.eu/