

Data-Sovereign Enterprise Collaboration using the Solid Protocol

Thorsten Kastner^{1,3}, Christoph H.-J. Braun², Andreas Both^{1,5}, Dustin Yeboah¹, Sebastian J. Schmid³, Daniel Schraudner³, Tobias Käfer² and Andreas Harth^{3,4}

¹DATEV eG, Nuremberg, Germany

²Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

³Friedrich-Alexander University (FAU), Nuremberg, Germany

⁴Fraunhofer Institute for Integrated Circuits IIS, Division Data Spaces and IoT Solutions, Nuremberg, Germany

⁵Leipzig University of Applied Sciences, Leipzig, Germany

Abstract

We demonstrate a system architecture for seamless and sovereign business-to-business (B2B) data sharing using the Solid Protocol. We highlight two core system components: The Rights Delegation Proxy allows organizations to internally manage and enforce access policies on requests from their employees to external data providers. The Data Provision Proxy allows organizations to share data from an external data provider in a privacy-preserving manner. Organizations define and enforce the policies for internal and external data sharing themselves, thereby maintaining sovereignty in enterprise collaboration.

Keywords

Solid Protocol, Dataspaces, Data Value Chains, Zero Trust, Data Sovereignty

1. Introduction

In today's connected world, data sharing between enterprises and among organizations is commonplace. Both providing and consuming businesses benefit from sharing data in collaboration. In their data-sharing processes, today's businesses often rely on centralized platforms or ad-hoc solutions like email attachments. As the requirements of data sovereignty, security, and separation of data and application [1] become more pressing, endeavors like the International Dataspaces (IDS)¹ or Solid Dataspaces (SDS) [2] are gaining momentum.

Building on the SDS approach, we present a system for sovereign data sharing along a chain of enterprises. We implement a use case for credit requests, where an enterprise requests a credit from a bank and is required to provide data made available by the enterprise's tax advisor. Only the directly communicating actors know each other, so the origin of data passed on must remain hidden while data is still processed along the business chain. In our solution, the data is not stored as a hard copy for each participant. Instead, data from the original source is passed along to involved parties on-demand if access is granted. The party owning the data source is in control of data access and – we assume here – has permitted re-distribution according to contractually defined purposes. We thus present an implementation of a *data value chain* [3].

SEMANTICS'24



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0

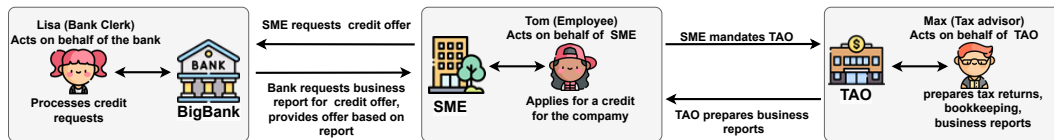


Figure 1: Use Case: a credit request.

In this paper, we highlight the core components of our demonstrator: a *Rights Delegation Proxy (RDP)* [4] to handle the actions of natural persons on behalf of legal entities, and a *Data Provision Proxy (DPP)* [3] to requests to concealed locations along the chain. Our contribution is the demonstration of a Solid-based system for sovereign B2B data sharing.

2. Use Case

Fig. 1 illustrates our use case where the banking institution BigBank, the tax advisory office TAO, and the company SME create a data value chain for a credit offer scenario. As a representative of BigBank, the bank employee Lisa acts as a natural person.

Upon a credit request, BigBank is to create a credit offer for the requesting organization SME. The bank’s employee Lisa is responsible for this task. To assess the financial situation of SME, Lisa needs to access accounting data and a business report of SME. This data is prepared and provided to SME by TAO. To get the credit, SME must provide that data to BigBank. In this data value chain, two requirements exist regarding business relationships and data sovereignty [3]:

- A:** The relation between SME and TAO must remain secret until SME is to disclose it.
- B:** The TAO is required to keep control over the aggregated business report data.

3. Related Work

The Solid Protocol is a bundle of specifications for read/write Linked Data under access control. The protocol builds on the RESTful HTTP specification of a Linked Data Platform (LDP)² for interacting with Web resources, and extends it with authentication and authorization using an extended version of OpenID Connect³ and Web Access Control⁴.

For handling the delegation of rights by real-world businesses, government-managed registers of commerce are used. Companies publicly and lawfully declare which individuals may act and negotiate on their behalf; see the German Commercial Register⁵ for example. In cyberspace, the long-standing vision of dataspaces aims to integrate different data sources [1]: Top-down approaches like the International Data Space (IDS)⁶ or GAIA-X⁷ aim for sovereign data exchange using governance models and standardized data exchange components. From a bottom-up

²<https://www.w3.org/TR/ldp/>

³<https://solidproject.org/TR/oidc>

⁴<https://solid.github.io/web-access-control-spec/>

⁵<https://www.handelsregister.de/>

⁶https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0

⁷<https://docs.gaia-x.eu/technical-committee/architecture-document/22.04/>

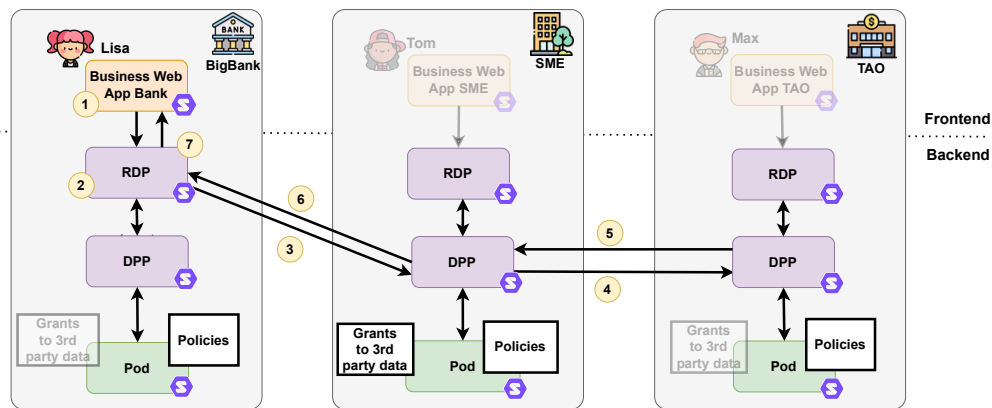


Figure 2: Component view of the system deployed at each organization. Components not relevant to the walkthrough described are greyed out in the figure.

perspective, Meckler et al. [2] identifies the Solid Protocol⁸ as one technological foundation for decentralized yet interoperable dataspaces, named *Solid Dataspaces (SDS)*.

For data exchange between collaborating enterprises, Henselmann et al. [5] present a Solid-based solution in a credit request scenario similar to ours. While their solution passes data along a chain of participants, their system relies on hard copies of data to be stored at each participant and does not include natural persons as actors for legal entities. In a machine-to-machine interaction use case, Wang et al. [6] present rule-based agents that negotiate a trade contract under German law using the Solid Protocol. In contrast to our approach, decision-making is automated and all participants, their needs, and data are known.

4. Demonstrator

Our demonstrator^{9,10} is composed of the following components for each company (see Fig. 2):

Solid Business Web Apps enable the employees to carry out their business activities. For the example of SME, the initiation of a credit inquiry, the provision of the business reports requested by the bank to prepare a credit offer, and the approval of the Bank's credit offer. In addition, an authorization application, as described in [7], is used to manage incoming, existing, rejected, and revoked requests for data sharing. A **Company's Solid Pod** stores the company's business data and makes it available under access control. Notably, it also contains rights delegation policies, i.e., rules that define which specific employees are allowed to interact internally and externally on behalf of the company. Further, it contains definitions for resources that are made available via proxied data provision. A **Rights Delegation Proxy (RDP)**, as described in [4], receives and logs requests made by the company's employee (e.g., Tom). It authenticates the employee using their WebId and checks if they were delegated the required rights to proceed

⁸<https://solidproject.org/TR/protocol>

⁹<https://github.com/mandat-project/hackathon-demo>

¹⁰<https://github.com/mandat-project/delegation-proxy>

with their request. To this end, the RDP retrieves and validates corresponding policies defined by the delegator (SME) from their Pod. If all requirements are fulfilled, the RDP proceeds with the delegatee's request but updates the authentication headers with the delegator's credentials. Any received response is logged and forwarded to the delegatee. **A Data Provision Proxy (DPP)**, as envisioned in [3], receives all data requests (e.g., from Lisa on behalf of BigBank) and checks on the Pod whether the requested resource is an actual Pod-stored resource or to be retrieved and passed along from an external data source. That is, it validates if the company's own data (SME) or data from a third party (TAO shared with SME) is requested. If the requested data originate from a third party, the DPP checks the sharing policy of the third party. If allowed, the DPP retrieves the resource (authenticated as SME) and provides it as if originating from the company (SME), masking the original data source (TAO).

A walk-through of our demonstrator:

1. Lisa logs in to the business Web App using her own WebID. To retrieve the accounting data on the SME's Pod, she sends an authenticated request to the RDP of BigBank.
2. The *BigBank RDP* authenticates Lisa and checks if she is authorized by the bank's policies to interact with SME. Then it requests the resource from SME, authenticated as BigBank.
3. The *SME DPP* receives the authenticated request coming from BigBank to access the accounting data. It checks whether the requested resource is an actual Pod-stored resource or an external one to be forwarded. Additionally, it checks if re-distribution is allowed.
4. On pass, the *SME DPP* performs an authenticated request as SME for the TAO's data.
5. The *TAO DPP* receives an authenticated request from SME to the resource containing the accounting data. As this resource is an actual Pod-stored resource, the request is forwarded. The Pod checks access control rules and returns the data via the proxy.
6. The *SME DPP* receives the response, logs it, and responds to the request of BigBank.
7. The *BigBank RDP* receives the response from SME and forwards it to Lisa's app.

A screencast of our demonstrator is available online¹¹.

Fulfillment of requirements:

Our system demonstrates how the requirements defined in Sec. 2 can be met:

- A:** BigBank and TAO are not disclosed to each other while Lisa (i.e., BigBank) still receives the desired data. SME and its RDP are acting as a broker facilitating the data sharing.
- B:** The accounting data remains stored on TAO's Pod and in its control, without hard copies being necessary at the SME's Pod.

Additionally, we highlight the following features of our approach:

- The *data value chain* could also continue further by the TAO without the SME knowing the additional upstream participants.
- The existence of participants is secret and only revealed on a "need to know"-basis. The existence of upstream or downstream business participants is obfuscated. Acting employees also remain private as their actions are attributed to the respective company.

¹¹<https://purl.archive.org/mandatb2b/Semantics2024>

- Policies for data access and re-use specify relations between two agents, e.g., the internal relation of Lisa and BigBank or the external relation of SME and TAO. The policies and thus the existence of these bilateral relations remain only known to the involved agents.

5. Conclusions and Future Work

In this paper, we presented a Solid-based system for sovereign data sharing between enterprises. Our main contribution is the demonstration of a *data value chain* that respects the privacy and data sovereignty of its participants. We emphasize that our demonstrator is comprised of reusable components based on Web standards and the Solid Protocol to build a Solid Dataspace. We highlighted the Rights Delegation Proxy (RDP), to let natural persons acting on behalf of their companies, and the Data Provision Proxy (DPP), to retrieve data from upstream data sources. We re-iterate that both of these components are built using the same Semantic Web standards and specifications as the rest our demonstrator.

In future work, we plan to address the process of interactively initiating a B2B collaboration while also protecting the currently often publicly accessible metadata of the involved enterprises.

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

References

- [1] M. Franklin, A. Halevy, D. Maier, From databases to dataspace: a new abstraction for information management, *SIGMOD Rec.* 34 (2005) 27–33.
- [2] S. Meckler, R. Dorsch, D. Henselmann, A. Harth, The Web and Linked Data as a Solid Foundation for Dataspace, in: *Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion*, Association for Computing Machinery, 2023, p. 1440–1446.
- [3] A. Both, D. Yeboah, T. Kastner, D. Schraudner, S. Schmid, C. Braun, A. Harth, T. Käfer, Towards Solid-based B2B Data Value Chains, in: *21st Extended Semantic Web Conference (ESWC 2024)*, 2024.
- [4] S. Schmid, D. Schraudner, A. Harth, The Rights Delegation Proxy: An Approach for Delegations in the Solid Dataspace, in: *Proceedings of the Second International Workshop on Semantics in Dataspace (SDS 2024) co-located with the 21st Extended Semantic Web Conference (ESWC 2024)*, 2024. URL: <https://ceur-ws.org/Vol-3705/paper02.pdf>.
- [5] D. Henselmann, K. Kolinsky, S. J. Schmid, D. Schraudner, A. Both, A. Harth, Solid Proof of Concept in an Enterprise Loan Request Use Case, in: *Proceedings of Poster and Demo Track and Workshop Track of the 18th International Conference on Semantic Systems (SEMANTiCS 2022)*, volume 3235, CEUR-WS, 2022.
- [6] X. Wang, C. H.-J. Braun, A. Both, T. Käfer, Using schema.org and solid for linked data-based machine-to-machine sales contract conclusion, in: *Companion Proceedings of the Web Conference 2022, WWW '22*, Association for Computing Machinery, 2022, p. 269–272.
- [7] A. Both, T. Kastner, D. Yeboah, C. Braun, D. Schraudner, S. Schmid, T. Käfer, A. Harth, AuthApp – Portable, Reusable Solid App for GDPR-compliant Access Granting, in: *International Conference on Web Engineering (ICWE 2024)*, 2024.