Harmonization Profiles for Trusted Data Sharing Between Data Spaces: Striking the Balance between Functionality and Complexity

Arjan J.R. Stoter[1], Bauke Rietveld[2], Vincent Jansen[2] and Harrie J.M. Bastiaansen[1]

[1] *TNO, Eemsgolaan 3, Groningen, 9727 DW, The Netherlands*
[2] *Innopay, WTC, F-Tower, Strawinskylaan 381, Amsterdam, 1077 XX, The Netherlands*

Abstract

The ambition of the EU Data Strategy can be summarized as a '*federation of interoperable data spaces*'. Currently, a multitude of architectures, frameworks and protocols is used by various data spaces. The Data Sharing Coalition has provided an architecture framework for interoperability between data spaces, making use of a harmonization domain and data space proxies as key architecture concepts. Complete interoperability between a wide variety of data spaces presents a challenge for the harmonization domain. To enable interoperability between a variety of data spaces, a set of harmonization profiles are required in the harmonization domain to provide the necessary functionalities. However, implementing each harmonization profile comes with additional complexity. Therefore, it is desirable to identify a minimal set of harmonization profiles to provide interoperability between an adequate variety of data spaces. This paper addresses the identification of harmonization profiles, presents a framework for structuring harmonization profiles and explores the impact of key trust aspects (policy management and trust ecosystem) on harmonization profiles.

Keywords

European data strategy, data space interoperability, federation, harmonization profile, proxy model, access and usage policies, trust ecosystem, IDS.

1. Introduction

Data and data sharing are clearly on the radar of the European Commission. The release of the European Data Strategy [1], the Data Governance Act [2] and the additional input sought on data spaces through OPEN DEI [3] illustrate the importance the EU attributes to data sharing for our society and economy. The goal of the EU Data Strategy can be summarized as a '*federation of interoperable data spaces*', for which interoperability of data sharing is vital, both within and across data spaces. In practice however, there is no single architecture, (legal) framework or protocol stack that is used by all data spaces. Sectors and communities are currently deploying or developing data spaces using a variety of approaches [4-6], posing a major challenge for interoperability between data spaces. An overarching framework to address interoperability is provided by the new European Interoperability Framework (EIF) [7], distinguishing the levels of technical, semantic, organizational, and legal interoperability. Establishing trust is a key component within the interoperability framework and the focus of this paper.

The Data Sharing Coalition (DSC) addresses inter data space interoperability in its Data Sharing Canvas [6], which describes an architecture framework to enable the trust required for interoperability between data spaces. The Data Sharing Canvas introduces the concept of 'harmonization', which is

Proceedings of the Workshop of I-ESA'22, March 23–24, 2022, Valencia, Spain
EMAIL: arjan.stoter@tno.nl (A.J.R. Stoter); rietveld.bauke@gmail.com (B. Rietveld); vince.jansen@innopay.com (V. Jansen); harrie.bastiaansen@tno.nl (H.J.M. Bastiaansen)
ORCID: 0000-0003-1673-4140 (H.J.M. Bastiaansen)

defined as 'the establishment of agreements, standards, and requirements between participants to enable data sharing between them'. The trust required for data sharing is enabled through harmonized, overarching trust framework agreements, which data spaces should adhere to. Compliance with these overarching agreements can be achieved via full or partial harmonization [6].

Full harmonization of data spaces means that data spaces (internally) adhere to the same requirements and principles, thereby enabling inter data space interoperability. Given the impact in terms of alignment effort and costs, full harmonization is often not feasible in practice. Therefore, the Data Sharing Canvas introduces partial harmonization through a new component, called a data space proxy, that absorbs the complexity of harmonization of data spaces. Proxies allow data consumers and providers within a data space to simply connect to other data spaces via their proxy. Proxies have the main functionality of translating data space specific transactions to their harmonized equivalents, thereby facilitating interoperable transactions and creating an understanding of concepts like trust and security across data spaces.

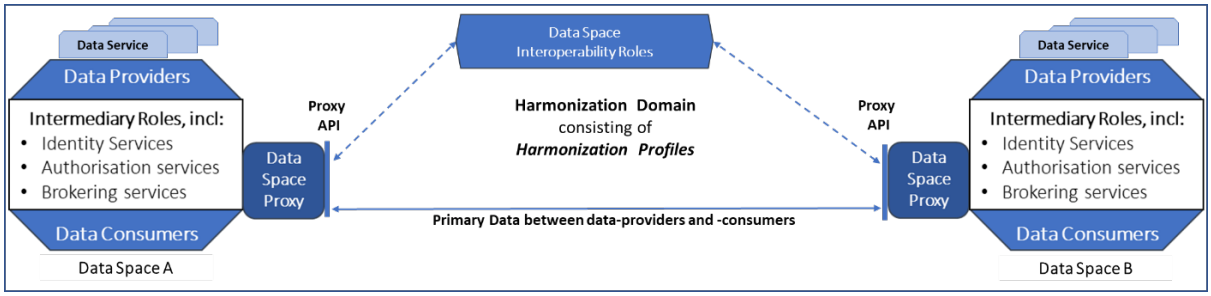

Figure 1: Data space interoperability roles and architecture, based on partial harmonization.

The harmonization domain (i.e., the domain between proxies, see Figure 1) uses a technical protocol, defined as a harmonization profile. Ideally, this is a single protocol that supports all required functionalities to facilitate trust and data sharing capabilities between all types of data spaces. In practice, given the variety of possible data space architectures, frameworks, and protocols, it is not feasible that a single harmonization profile can be used as a technical 'lingua franca' in the harmonization domain. More likely, multiple harmonization profiles are required to facilitate the interoperability for specific types of data spaces.

The current paper focusses on the trust aspects that are to be supported by the harmonization profiles. More specifically, it addresses the research question: *"How to structure and identify a minimal set of harmonization profiles that enables trust for interoperability between an adequate variety of data spaces?"*. This research will be addressed in an exploratory manner.

The following section describes main trust aspects of data space interoperability and explores their interdependencies. The subsequent sections address the main distinguishing features within trust aspects that lead to different harmonization profiles. The findings section considers the assessment of the defined harmonization profiles and provides initial validation results, after which the final section describes conclusions and follow-up work.

2. Trust aspects for defining a harmonization profile

As indicated, this paper focusses on the trust aspects for interoperability of data spaces by means of the partial harmonization approach, with data space proxies and harmonization models as key architectural concepts. Two main trust aspects need to be addressed: *policy management* and the *trust ecosystem*. Policy management encompasses access- and usage policies **Error! Reference source not found.**. Both express business and regulatory policies. Access policies define which participants are allowed access to data services, whilst usage policies define what participants are allowed to do with the data. A trust ecosystem ensures that all interactions between participants in a participant chain are trustworthy, both within- as well as between data spaces. A trust ecosystem is a prerequisite for data sharing transactions of (sensitive) primary data.

As a working hypothesis for the remainder of this paper, it is assumed that aspects of policy management and the trust ecosystem are independent. This implies that they may be developed

independently, and their impact on the number of required harmonization profiles is additive rather than multiplicative. The latter would -theoretically- reduce the required number of harmonization profiles drastically. This assumption is further addressed in the findings section. The following sections address the main distinguishing features of usage policy management and the trust ecosystem, respectively.

3. Policy management

Governance of usage- and access policies encompasses interactions between a data provider and data consumer for defining- and agreeing upon policies, as well as the capabilities to enforce them. Two commonly used approaches are identified here, each with its own main associated implementation technology, namely policy management with access tokens and policy management with contract negotiations.

In policy management with access tokens, a two-stage approach is followed in which (1) an access token is obtained from the data provider, based on approval provided by the entitled party, with which (2) the data can be retrieved from the data provider. Policy enforcement capabilities are in this case only required for the data provider, and only short-living, unidirectional sessions between data provider and data consumer are needed. The OAuth2.0 protocol [8] is commonly used as an implementation technology for policy management with access tokens, based on generic web service calls in the form of APIs using access tokens for authentication of data consumers by the data provider. It is important to note that access tokens can be used to enable access policies only, which means that potential usage control needs to be implemented separately.

Policy management with contract negotiation is applicable for data spaces with both access and usage control policies. As in policy management with access tokens, a two-stage approach is followed in which (1) a data sharing contract is negotiated between a data provider and a data consumer, based on which (2) the data provider shares the data with the consumer. Policy enforcement capabilities to support usage policies are required both at the data provider and consumer requiring long-living, bi-directional sessions between them.

An implementation technology that supports contract negotiation and policy enforcement to manage usage policies is currently provided by the International Data Spaces (IDS) initiative **Error! Reference source not found.** as a secure connectivity protocol, i.e., the IDS Communication Protocol (IDSCP) **Error! Reference source not found.**. IDS is attracting major attention as a pillar for the European Data Strategy [1] and its design principles as being developed by the EU OPEN DEI initiative [3]. IDS It is being defined by the IDS Association (IDSA) and is standardized **Error! Reference source not found.**.

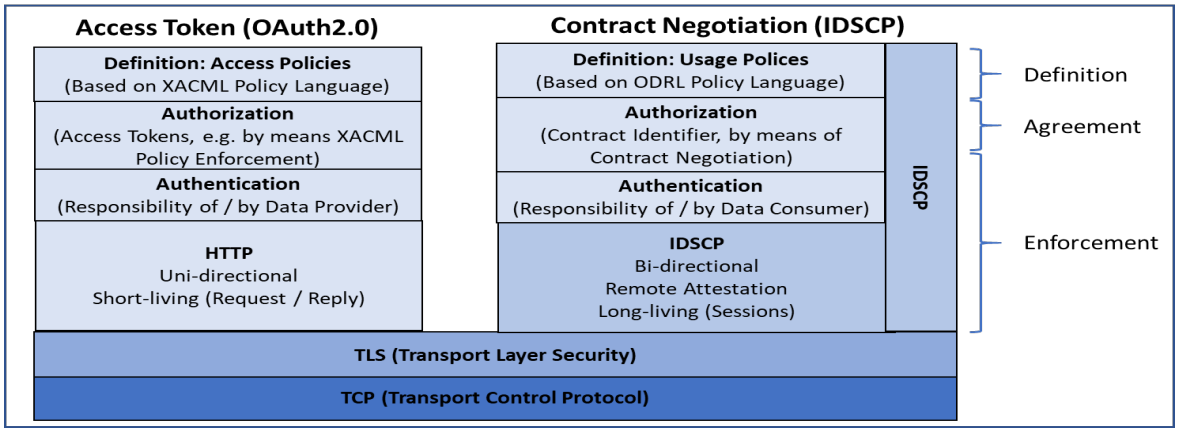

Figure 2: Policy management using access tokens (OAuth) (l) and contract negotiation (IDSCP) (r).

Figure 2 depicts the protocol stack for access- and usage policy management using access tokens and contract negotiation. A distinction is made between protocols for definition-, agreement- and enforcement of policies. Interoperability between data spaces requires specific harmonization profiles

for both access- and usage policy management that are tailored to the needs and capabilities of each data space. The two approaches and their corresponding protocols are the foundation for developing the associated harmonization profiles.

4. Trust ecosystem

An essential functionality of a harmonization profile is to facilitate trust in data sharing between participants in different data spaces. Compared to data sharing within a data space, facilitating trust becomes increasingly complex for inter data space interoperability, as it involves a chain of participants in the various types of roles as depicted in Figure 1. There are core roles (e.g., data consumers and data providers), intermediary roles (e.g., identity services, authorization services, brokering services and proxies) and interconnectivity roles. The identity and authenticity must be established and validated for each participant fulfilling a specific role. Additionally, authorization for each action of these participants must be verified. Depending on data space implementations, these mechanisms for identification, authentication, and authorization (IAA) can either be 'transparent' or 'opaque'.

Transparent IAA mechanisms provide a data space with all the required information to do IAA assessments. In this situation, IAA is the direct responsibility of the participant performing the IAA (e.g., a service provider). This requires complete (or at least sufficient) transparency of IAA information across data spaces. A harmonization profile should, in this case, enable communication and understanding of IAA information across data spaces, that is, to enable information transfer and mapping between a data space specific format and a harmonized equivalent. Furthermore, the IAA information must provide sufficient assurance so that it can be regarded as trustworthy, e.g., through signed claims. It is likely that all forms of transparent IAA could be handled by a single harmonization profile. Even though functional IAA (that is, how IAA related information is actually processed) can vary across data spaces, this would not result in *fundamental* differences in IAA data transport or translation protocols.

Opaque IAA mechanisms work via delegated decisions, where IAA authorities basically provide a 'yes' or 'no' answer. Here, IAA decisions can be made on behalf of others, without the need to explicitly share IAA information across data spaces. Technical trust enforcement measures include encryption and use of public key infrastructures (PKI) to be able to trust communications and claims. Due to the high level of international standardization that exists for these technical security measures (such as TLS/SSL and X.509 certificates), these can be sufficiently covered with a single harmonization profile.

Based on the above assessment, no wide variety of harmonization profiles seem to be required to enable trust in the ecosystem and support both transparent and opaque IAA. In fact, a single harmonization profile for each seems to be sufficient to enable all possible needs and capabilities for trust across a variety of data spaces.

5. Findings and status

The Data Sharing Coalition has defined an initial harmonization profile in its Use Case Implementation Guide (UCIG) **Error! Reference source not found.**. Based on this UCIG, a representative proof-of-concept (PoC) was realized for controlled sharing of privacy-sensitive geriatric data between a hospital and a municipality **Error! Reference source not found.**. For the trust aspect on policy management and the trust ecosystem, the UCIG and the PoC implement the *policy management with access tokens approach* and *the opaque approach,* respectively. Based on the results of this implementation, three validation perspectives are addressed below.

Validation of independence of trust aspects. Previously, the hypothesis was made that the trust aspects of policy management and the trust ecosystem were independent, resulting in a limited number of harmonization profiles. In the UCIG these elements were independently described, and the PoC implementation gave a practical confirmation of this independence. Further validation of the independence of the remaining policy management- and trust ecosystem aspects is to be provided by

representative data space interoperability cases that include trust aspects for the contract negotiation approach and the transparent IAA mechanisms.

Validation of adequateness of individual harmonization profiles. The functional adequateness of harmonization profiles can be validated by means of functioning implementations. The UCIG and its initial PoC explored the functional adequateness of the policy management with access tokens and transparent IAA mechanisms. The PoC resulted in a fully functioning harmonization profile and data sharing across data spaces, through which adequacy was confirmed. Findings in the implementation of the PoC have led to minor updates of the UCIG. Moving forward, additional illustrative and representative data space interoperability scenarios need to be defined and validated. A harmonization profile for the policy management with contract negotiation approach between data spaces will be developed as a next step, resulting in an implementation guide and associated PoC as upcoming for interoperability developments **Error! Reference source not found.**.

Validation of completeness of set of harmonization profiles. Policy management and trust ecosystem were identified as the two most relevant functional aspects of the harmonization domain which could require a variety of harmonization profiles. Initial results suggests that other aspects (such as digital identities and metadata) do not require multiple harmonization profiles to cover the variety of required functionalities for alternative implementations. Within both aspects covered in this paper, two main distinguishing cases leading to individual harmonization profiles have been described. Market research [4] and experience (i.e. professional judgement of the authors) in the data sharing context conclude that interoperability between current data space developments are covered by this initial set of harmonization profiles. Whether additional harmonization profiles will be required will become apparent in the future developments and projects.

6. Conclusions and future work

A primary objective of this paper was to provide a framework for - and initial exploration and structuring of - harmonization profiles to support interoperable data sharing between a variety of data spaces. It was identified that a number of different harmonization profiles were required to enable all required functionalities for data sharing between a variety of data spaces. The preliminary conclusion is that data space interoperability may be realized by a limited set of harmonization profiles, for which two trust aspects (policy management and trust ecosystem) can be independently developed as part of the harmonization profiles. For both trust aspects there is a limited number of varying protocols that act as a distinguishing factor for harmonization profiles, of which the two main ones have been identified in this paper. Independence of both trust aspects suggests that the overall number of harmonization profiles to be developed in the future remains limited and manageable.

Future work includes (i.) the identification of the (need for) additional harmonization profiles and the specification thereof, (ii.) developing the additional data spaces interconnect roles and architecture as depicted in Figure 1, (iii.) elaborate the data space interoperability architecture at other interoperability levels of the EIF [7], especially at the semantic and legal level, and (iv.) assess the scalability (with respect to the number of data spaces) for implementations of the identified harmonization profiles.

7. Acknowledgements

The work as presented in this paper extends upon the architectural foundation for interoperability between data spaces as laid by the Data Sharing Coalition. It has been supported and co-financed by the Data Sharing Working Group of the Dutch AI Coalition (NL AIC, https://nlaic.com/en/building-blocks/data-sharing/). In addition, this paper builds upon the work done within the Dutch Research project DASLOGIS, supported by the Dutch Top consortia for Knowledge and Innovation Institute for Advanced Logistics (TKI Dinalog, www.dinalog.nl) of the Ministry of Economy and Environment.

8. References

[1] European Commission, A European strategy for data, 2020. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_273, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN.
[2] European Commission, Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act), 2020. URL: https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767.
[3] OPENDEI, Design Principles for Data Spaces – Position Paper, 2021. URL: https://design-principles-for-data-spaces.org/.
[4] The Netherlands AI Coalition, AI Ecosystem & Market Analysis: Quick scan of data sharing market to validate blueprint of the NL AIC, 2020. URL: https://nlaic.com/wp-content/uploads/2021/02/AI_Ecosystem_and_Market_Analysis_Data_Sharing_4-feb-2021.pdf.pdf.
[5] TRUSTS, Trusted Secure Data Sharing Space, 2022. URL: https://www.trusts-data.eu/.
[6] Data Sharing Coalition, Data Sharing Canvas - A stepping stone towards cross-domain data sharing at scale, 2021. URL: https://datasharingcoalition.eu/app/uploads/2021/04/data-sharing-canvas-30-04-2021.pdf.
[7] European Commission, Directorate-General for Informatics, New European Interoperability Framework (EIF): Promoting seamless services and data flows for European public administrations, 2017. URL: https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf.
[8] IDSA. Data Sovereignty: Updated Position Paper on Data usage Control in the IDS, 2021. URL: https://internationaldataspaces.org/data-sovereignty-updated-position-paper-on-data-usage-control-in-the-ids/.
[9] D. Hardt, The OAuth 2.0 Authorization Framework, 2012. URL: https://datatracker.ietf.org/doc/html/rfc6749.
[10] International Data Spaces Association, Reference Architecture Model. Version 3, 2019. URL: https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf
[11] Data Sharing Coalition, Use Case Implementation Guide, 2022. URL: https://datasharingcoalition.eu/app/uploads/2022/03/data-sharing-coalition-use-case-implementation-guide-ucig.pdf
[12] The Netherlands AI Coalition, Dataspace Proxy PoC, 2022. URL: https://gitlab.com/nlaic/dataspace-proxy-poc
[13] IDSA, TNO Will Set Up Next-Generation Data Spaces with NTT, 2022. URL: https://internationaldataspaces.org/tno-will-set-up-next-generation-data-spaces-with-ntt/