REFERENCE GUIDE FOR INTER AI DATA SPACE INTEROPERABILITY

Interconnecting multiple AI data space instances

Working group Data Sharing

NL AI Coalition

March 2023

Content

MANAGEMENT SUMMARY

The ambition of the Netherlands AI Coalition (NL AIC) is to position the Netherlands at the forefront of knowledge and application of AI for prosperity and well-being. To achieve this goal, it is deemed crucial to make data widely available to train and fuel the AI-algorithms. This is why the NL AIC working group Data Sharing has set the goal of the creation of trustworthy and interoperable AI data spaces. This will be done in alignment with the European data strategy, in short summarised as '*Towards a Federation of Interoperable (AI) Data Spaces*'. Therefore, in 2021, it has described the overarching AI data space reference guide [1], in which two developments lines are introduced: the intra and the inter AI data space interoperability development line. This report address the latter, i.e. the development line for inter AI data space interoperability. The former is addressed in the companion report [2].

The ambition of the reference guides for both intra and inter AI data space interoperability is to support organisations in developing interoperable AI data spaces to address the data sharing challenges to optimally support AI with its variation in collaboration models as introduced in the overarching AI data space reference guide [1]. They elaborate the architecture and building blocks, providing a rich set of capabilities to support data sharing and to ensure trust and interoperability within and between different AI data spaces.

The inter AI data space reference guides in this report address two distinct scenarios for inter AI data space interoperability, i.e.: (1) the '*homogenous*' inter AI data space interoperability scenario with full harmonisation in which the building blocks of the aligned AI data space architectures are '*federable*' over multiple AI data spaces, and (2) the '*heterogenous*' inter AI data space interoperability scenario with partial harmonisation in which data space proxies are used to translate differences in data space specific, non-aligned, building block implementations into their harmonised equivalents to absorb the variation in protocols.

The subsequent parts of this report address the ecosystem, the building block and the trust architecture for inter AI data space interoperability. With the ecosystem architecture describing the main strategic, organisational and data space ecosystem principles, the building block architecture defining and elaborating the individual building blocks, and the trust architecture addressing the interaction patterns and protocols to assure that data and AI processing services and resources are shared in a trustworthy manner. In addition, this report contains a part on the reference implementation, roadmap and conclusions for inter AI data space interoperability.

Both the work on the intra and inter AI data space interoperability development line report on work-in-progress. Based on the input, know-how and expertise of the participants of the NL AIC working group Data Sharing, the reports provide the foundation for data spaces for AI in the Netherlands. The collaborative development of the AI data space architecture, its building blocks, the sharing of best practices and the management of the roadmap from proofs-of-concept towards operationalisation paves the way to the successful introduction of a federation of operational and interoperable AI data spaces in the Netherlands.

Moreover, it is to be noted that the Netherlands with the NL AIC working group Data Sharing and adjacent data sharing initiatives has a good starting position to make data sharing for AI work and to take a leading role in Europe for realising the European Data Strategy [3]. As such, the work and knowledge of the NL AIC working group Data Sharing will be provided transferred to the various Dutch and EU initiatives working on a common goal and strategy for realising the European Data Strategy of the '*federation of interoperable data spaces*'. Specifically to (1) the Data Sharing and Cloud Centre-of-Excellence as joint follow-up effort of the work for the Data Sharing Coalition, the NL AIC working group Data Sharing and the Gaia-X Hub in the Netherlands and to (2) the EU Data Spaces Support Centre (DSSC) project as part of the Digital Europe program addressing the aligned development of data spaces for and across various sectors in Europe.

INTRODUCTION

Founded in 2019, the NL AIC has been set up to support well-being and welfare in the Netherlands by putting it in a front-runner position in terms of AI knowledge and applications. The NL AIC is a public-private partnership in which the government, business sectors, educational and research institutions, as well as civil society organisations collaborate to accelerate, implement, encourage and connect AI activities [4].

1. INTRODUCTION

One of NL AIC's building blocks is 'Data Sharing' [4], for which the NL AIC working group Data Sharing has been started in 2020. This introductory chapter describes the goal, scope and structure of this reference guide for inter AI data space interoperability in the context of the overarching work and deliverables of the NL AIC working group Data Sharing.

1.1 The NL AIC working group Data Sharing: goals and deliverables

A dedicated working group, the NL AIC working group Data Sharing, is tasked with providing the community knowledge, guidance and resources around responsible data sharing for AI, taking due note of Dutch and European developments and values.

As preparatory work, in 2020 the NL AIC working group Data Sharing has (1) identified the specific challenges for data sharing for advanced data analytics and provided an overview of technologies and architectures that can be used in addressing these challenges [5][6], (2) outlined the process of how companies can share data for AI, from experimental ("first-time engineering") phase to a phase of daily practice ("operationalisation") [7], (3) developed three proofs-of-concepts to demonstrate the architectural and technical concepts for controlled data sharing for AI, using three illustrative and representative cases from the sectors 'government', 'health' and 'energy' [8], (4) done a 'GAP-analysis' on the system operations gaps and the governance gaps to be bridged between the architectures and technology as demonstrated in the proofs-of-concepts and the large-scale deployment and adoption thereof [8], and (5) caried out a quick scan to validate that a data space approach is in line with international developments [9].

Starting in 2021, the NL AIC working group Data Sharing has initiated the structural work on the interrelated set of deliverables as graphically depicted in **Figure 1**.

As the figure shows, the NL AIC working group Data Sharing has provided the overarching reference guide for AI data spaces ('*Towards a Federation of AI Data Spaces*') in 2021 [1]. It sets the development direction towards federated and interoperable AI data spaces, aligning with the European data strategy and adhering to the European values of trust and

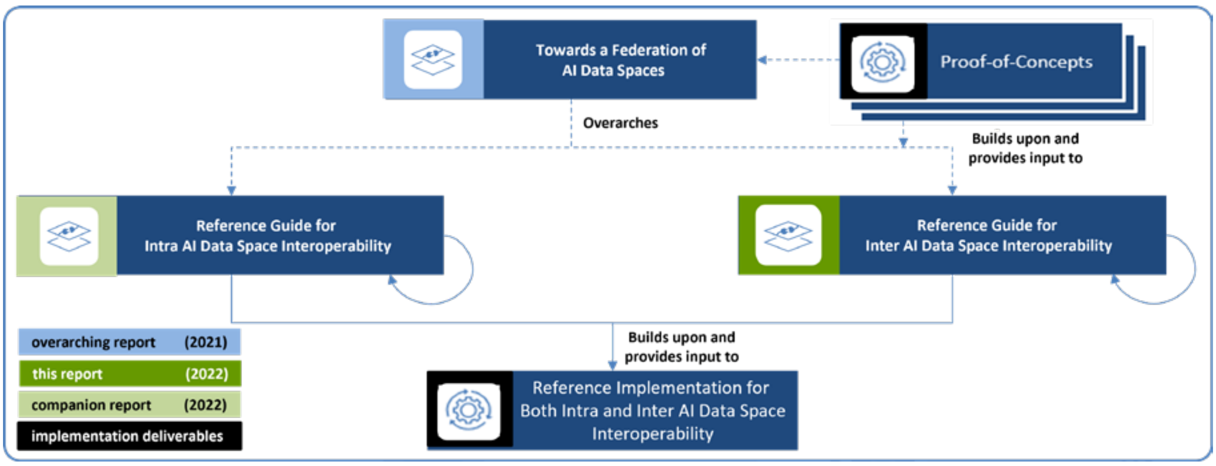

Figure 1 - Deliverables of the NL AIC working group Data Sharing and their interrelationship.

data sovereignty. It introduces the developments lines for intra and inter AI data space interoperability, which are reported in their corresponding reference guides.

Moreover, to show the potential and to identify lessons learned, their architectural concepts and technologies have been demonstrated by means of use cases and demonstrators in close collaboration with participants the NL AIC working group Data Sharing [10] and are further developed by means of illustrative and representative scenarios in the NL AIC reference implementation as described in chapter 8 and annex B of the companion report [2].

1.2 Reference guides for intra and inter AI data space interoperability: goals

As described in [1], AI data spaces provide the ecosystem and building blocks for sharing data and AI algorithms, for processing AI algorithms and data apps and for managing trust, data sovereignty and (legal) agreements. In view of the European ambition of federation of interoperable European data spaces, adequate governance is required to realize interoperability of the AI data space building blocks, both within individual AI data spaces and between multiple AI data spaces. Therefore, the NL AIC working group Data Sharing distinguishes two development lines for AI data spaces:

- *Intra AI data space interoperability*, focussing on a reference architecture, building blocks, guidelines and solutions for interoperability between building blocks within a single AI data space.
- *Inter AI data space interoperability*, focussing on a reference architecture, building blocks, guidelines and solutions for interoperability between multiple AI data space instances.

The work on the inter AI data space interoperability development line is reported on in this report. It elaborates the overarching architecture, building blocks and roadmap for inter AI data space interoperability. The goal is to serve as reference guide for realising interoperability between multiple data space instances, jointly providing overarching data sharing capabilities whilst ensuring trust and interoperability between AI data spaces.

The work on the intra AI data space interoperability development line is reported in the companion report '*Reference guide for intra AI data space interoperability*' [2].

1.3 Transfer of results

The results of the NL AIC working group Data Sharing on architecture, building blocks and roadmap (as described in the previous section) are transferred to and followed up both within the relevant data space development initiatives, both within the EU and within the Netherlands:

- *Within the EU context*: hand-over of the results is done to the Digital Europe programs (under the responsibility of EU DG Connect) addressing the aligned development of data spaces for and across sectors, specifically the EU Data Spaces Support Centre programme [11] aimed to facilitate common data spaces that collectively create an interoperable data sharing environment in Europe, executing from October 2022 until March 2026 and the EU SIMPL initiative [12] aimed at procuring the open-source development of the smart middleware building blocks that will enable cloud-to-edge federations and support all major data initiatives funded by the European Commission, such as the common European data spaces.

- *Within the Dutch context*: the Centre-of-Excellence Data Sharing and Cloud as currently being defined as joint effort in the Netherlands of the work of the Data Sharing Coalition [13], the NL AIC working group Data Sharing [14] and the Gaia-X Hub in The Netherlands [15].

With the transfer of the work of the NL AIC working group Data Sharing as described in this report, its results will be firmly embedded in strong national and international initiatives.

1.4 Structure of this report

The major part on the AI data space development lines has been described within the companion report on intra AI data space interoperability [2]. This report is restricted to elaborating the companion report only on those relevant aspects for interconnecting multiple AI data space instances.

The structure of this report is similar to the structure of the companion report. It has four parts, subsequently addressing the ecosystem architecture, the building block architecture and the trust architecture of the inter AI data space development line, followed by a concluding part describing the reference implementation, the roadmap and the overarching conclusions.

PART A: ECOSYSTEM ARCHITECTURE

The ecosystem architecture describes the main strategic and organisational principles that provide the foundation for developing AI data spaces. The ecosystem architecture encompasses the phases A and B as defined in the TOGAF Architecture Development Method (TOGAF ADM [16]): i.e. the architecture vision and the business architecture. These are addressed in chapter 2 and chapter 3, respectively.

2. ARCHITECTURE VISION

The architecture vision for inter AI data space interoperability (i.e. phase A in the TOGAF ADM [16].) builds (and extends) upon the architecture vision for intra AI data space interoperability as described in chapter 2 of the companion report [2]. As such, this vision is recapitulated in the following section of this chapter, after which the subsequent section introduces the business role model for inter AI data space interoperability.

2.1 European Data Strategy: federation of interoperable data spaces

As described in section 2.1 of the companion report [2], data sharing and data spaces are clearly on the radar of the European Commission. As expressed in European Data Strategy [3], the EU ambition on federative data sharing can be summarised as:

'Towards a federation of interoperable data spaces'.

As motivated and described in the overarching reference guide for AI data spaces *'Towards a Federation of AI Data Spaces'* [1] the development of AI data spaces as pursued by the NLAIC working group Data Sharing adheres and builds upon this ambition.

Furthermore, as described in section 3.2 of the companion report [2], both interoperability within individual data spaces and interoperability between multiple data spaces need adequate architectures and governance. These are referred to as intra and inter data space interoperability, as illustrated in **Figure 2**.

- *Intra data space interoperability*: Individual data spaces have a high degree of autonomy in developing and deploying their own internal agreements and architecture. Intra data space interoperability focusses on the alignment of the various capabilities (building blocks) within an individual data space.
- *Inter data space interoperability*: Interoperability between multiple data spaces is key for the federation of data spaces as expressed in the ambition of the EU Data Strategy. Inter data space interoperability requires alignment and guidelines for individual data spaces to ensure interoperability between them.

This report address inter AI data space interoperability. Intra AI data space interoperability is addressed in the companion report [2].

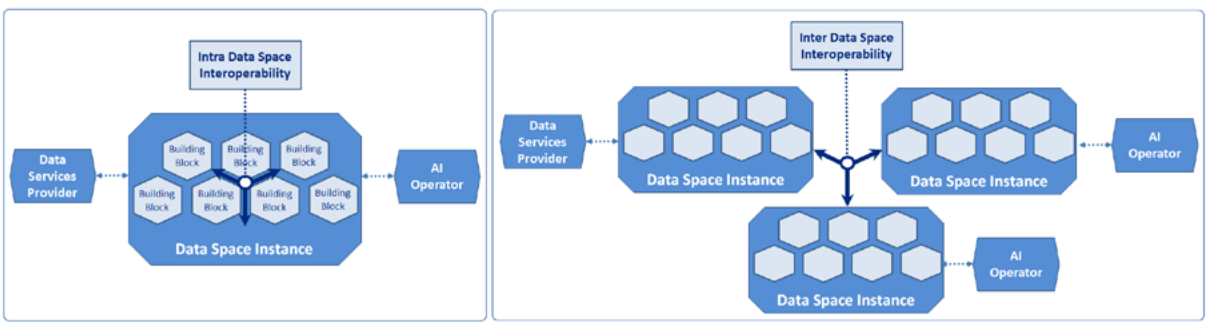

Figure 2 - Intra (l) and inter (r) data space interoperability.

2.2 Business role model for inter AI data space interoperability

The business role model for intra AI data spaces has been defined in the overarching reference guide report [1] and has been elaborated in the companion report [2], chapter 2. The business role model for inter AI data space interoperability extends (only slightly) upon the business role model for intra AI data space interoperability by including roles for governing and enabling inter data space interoperability as depicted in **Figure 3**.

As the figure shows, three categories of business roles can be distinguished for interconnectivity between data spaces (similar as for the business role model for intra AI data space interoperability): (1) the data space interconnectivity core roles, (2) the data space interconnectivity intermediary roles, and (3) the data space interconnectivity governance roles. The description of the business roles for each of the categories is included in Table 1.

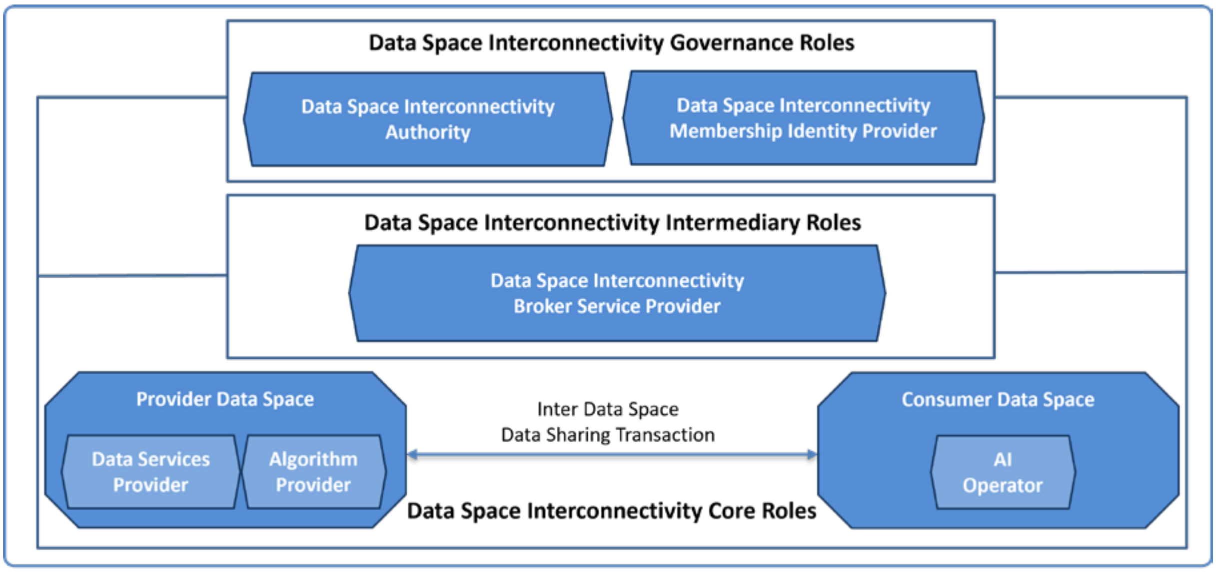

Figure 3 - Business role model for inter AI data space interoperability.

Table 1: The three categories of business roles for inter AI data space interoperability and their individual roles
Data Space Interconnectivity Core Roles The data space interconnectivity core roles represent the data space instances with stakeholders between which data sharing transactions are executed.
Provider Data Space Provider Data Spaces host a Data Services Provider or an Algorithm Provider that provide data services or an AI-algorithm to a stakeholder in an external data space, i.e. an AI Operator.
Provider Data Space Consumer Data Spaces host an AI Operator that requests data services or an AI-algorithm from an external data space, i.e. a Provider Data Space.
Data Space Interconnectivity Intermediary Roles The data space interconnectivity intermediary roles enable the processes for interaction between the stakeholders within different data space instances by providing metadata support services.
Data Space Interconnectivity Broker Service Provider A Data Space Interconnectivity Broker Service Provider manages information (metadata) about individual data spaces, e.g. on the business roles they support and data services providers and consumers they contain. The activities of a Broker Service Provider mainly focus on making the data services provided by stakeholders in various data spaces findable and available.
Data Space Interconnectivity Governance Roles The data space interconnectivity governance roles coordinate the set of commonly agreed principles between the data spaces and manage the compliance of data spaces to these agreed principles. As such, the data space interconnectivity governance roles manage the *'agreement framework' over data spaces'*, which can also be referred to as the 'trust framework'.
Data Space Interconnectivity Authority In larger ecosystems of data spaces, the Data Space Interconnectivity Authority is responsible for the (legal and operational) agreements between individual data spaces and for certification of participating data spaces.
Data Space Interconnectivity Membership Identity Provider The Data Space Interconnectivity Membership Identity Provider offers a service to create, maintain, manage, monitor, and validate identity information on participating data spaces. This is imperative for secure operation of the interconnectivity between data spaces and to avoid unauthorised access to data. The Data Space Interconnectivity Membership Identity Provider includes a Certification Authority, managing digital certificates for the participating data spaces.

3. BUSINESS ARCHITECTURE

The business architecture (corresponding to phase B in the TOGAF ADM [16]) encompasses a translation of the architecture vision architecture vision for AI data spaces (as described in the previous chapter) into business architecture principles and the inter data space interoperability modes as addressed in the following sections, respectively.

3.1 Business architecture principles

The business architecture principles for inter AI data space interoperability elaborate the architecture vision as described in the previous chapter. They align with the business architecture principles as defined for the individual data spaces in the companion report [2].

BA.1. *A single point of entry can provide access to each data service in the federation of AI data spaces*

To prevent the major integration efforts from having to connect to multiple data sharing environments, a single entry point gives participants access to each data service provided by a participant within any AI data space instance of the overarching federation. The single point of entry improves user friendliness and lowers the barriers for adoption.

BA.2. *Data may be a valuable asset for which data sovereignty and full stack integrity must be managed across the federation of interoperable AI data spaces*

Data is a valuable asset to be protected from unauthorised use and disclosure, required management of data sovereignty for the entitled parties. Hence, both access control (managing which participants are allowed access to the data) and usage control (managing what participants are allowed to do with the data) capabilities must be supported across the federation of AI data spaces. Moreover, full stack integrity must guarantee that data policies can be technically enforced whilst data sharing processing are adequately secured, e.g. using encryption, isolation, certification and remote attestation.

BA.3. *The inter AI data space interoperability modes apply to each of the levels of the new European Interoperability Framework (EIF)*

The new European Interoperability Framework (EIF) [17] is used as interoperability framework for AI data spaces, as described in section 3.2, depicted in **Figure 5** and elaborated in the companion report [2], section 3.2. To achieve inter AI data space interoperability the inter AI inter AI data space interoperability modes apply to each of these four interoperability levels of the EIF: technical, semantic, organisational and legal.

BA.4. *Minimal dependence and reliance on trusted third parties fulfilling data space interconnectivity intermediary and governance roles may be aimed for*

The business role model for inter AI data space interoperability has been defined in section 2.2 and depicted in **Figure 3**. It consists of the combination of data space interconnectivity intermediary roles and data space interconnectivity governance roles. The current EU developments on reference architectures for federative data sharing and data spaces (IDSA, GAIA-X, DSBA, ...) are developing towards fully distributed data space and trust framework capabilities, including distributed capabilities for IAA, contract negotiation and usage control. These developments may further minimize the (dependence and reliance on) the data space interconnectivity intermediary and governance roles provided by trusted third parties.

3.2 Inter data space interoperability modes: full and partial harmonisation

Interoperability between multiple data spaces (i.e. inter data space interoperability) is a key aspect of the EU Data Strategy as recapitulated in the previous section. The Data Sharing Coalition (DSC) addresses interoperability between multiple data spaces in its Data Sharing Canvas [18]. It introduces the concept of 'harmonisation', which is defined as '*the establishment of agreements, standards, and requirements between participants to enable data sharing between them*'. As the Data Sharing Canvas describes two basic inter data space interoperability modes can be distinguished: full and partial harmonisation, as depicted in **Figure 4**.

In case of *full harmonisation* of data spaces, the data space instances to be federated adhere to the same (already 'harmonised') principles, architecture, building blocks and protocols. Full harmonisation between data spaces provides major advantages for inter data space interoperability, both functionally and on ease and efficiency in realisation. However, in reality, full harmonisation may often not be feasible in practice and will also be an utopia for all newly formed data spaces. For existing data spaces for example, going for full harmonisation with other data spaces may have a big impact in terms of alignment and migration efforts and costs. The Data Sharing Canvas [18] therefore introduces *partial harmonisation* through a new module, called a data space proxy, that absorbs the complexity of harmonisation 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 capability of translating data space specific transactions to their harmonised equivalents, thereby facilitating interoperable transactions and creating an understanding of concepts like trust and security across data spaces.

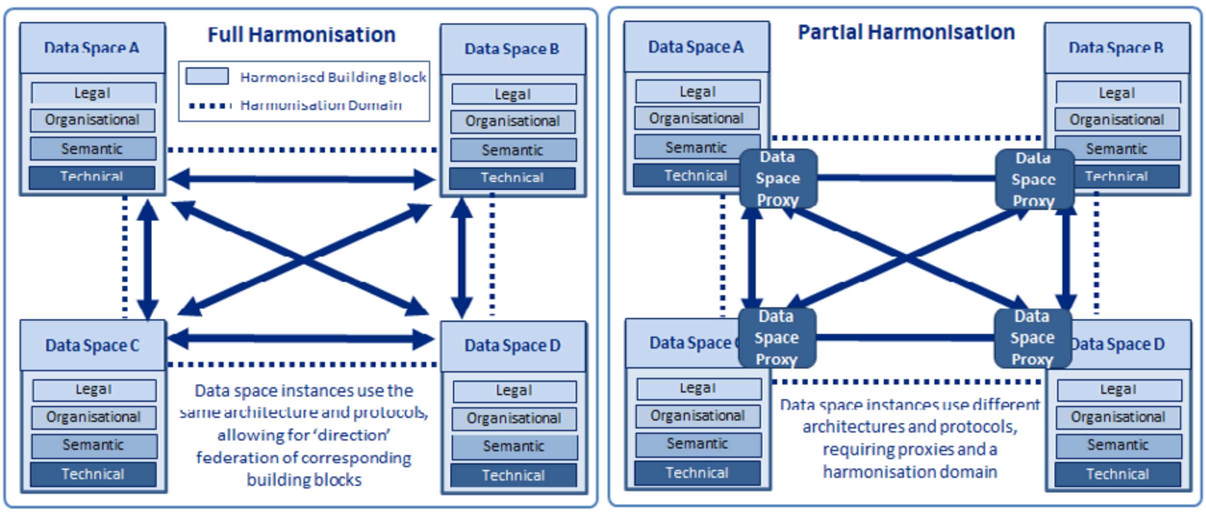

Figure 4 - Full (l) and partial (r) harmonisation mode for inter AI data space interoperability.

It is to be noted that both full and partial harmonisation are expected to be required to support the diversity of existing and emerging data space implementations. Therefore, both are addressed as inter AI data space interoperability mode in the remainder of this report.

Harmonisation is applicable at various levels and aspects requiring interoperability. An approach to systematically categorize the interoperability aspects is provided by the new European Interoperability Framework (EIF) as developed by the European Commission [17]. As **Figure 5** depicts, the EIF distinguishes four interoperability levels (technical, semantic, organisational and legal) under an overarching integrated governance approach.

Each of the four EIF interoperability levels needs to be addressed in developing the interoperability architecture for data spaces, both for intra and inter data space interoperability. For each of the levels, these aspects have been identified and described in the section 3.2 of the companion report [2]. Both full and partial harmonisation apply to the various levels and aspects requiring interoperability. The data space proxies for the partial harmonisation mode may separately operate at the individual interoperability levels and aspects of the EIF.

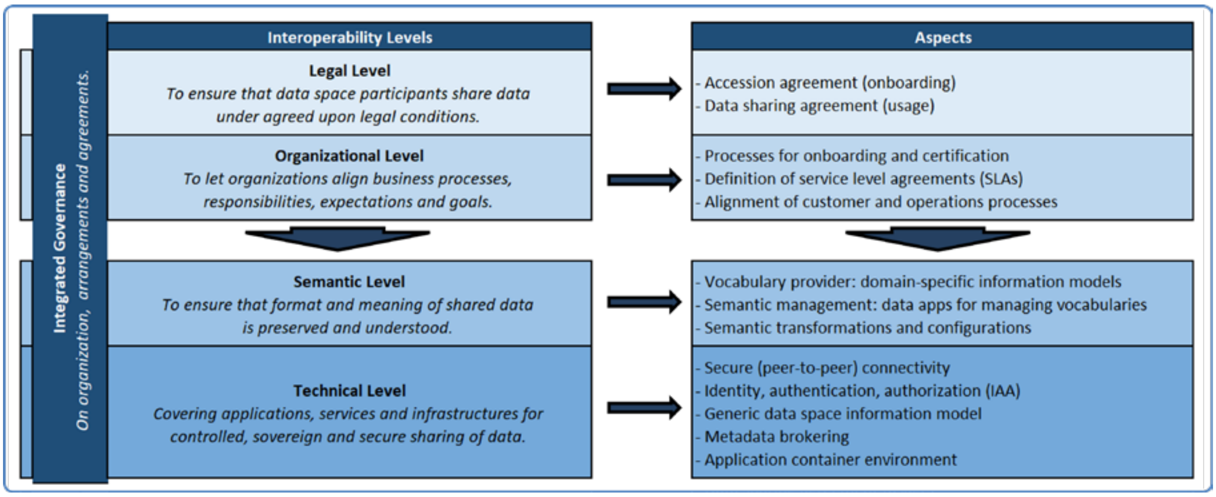

Figure 5 - The new European Interoperability Framework (EIF) [17]

PART B: BUILDING BLOCK ARCHITECTURE

The Information Systems Architecture (ISA) gives a breakdown of the inter AI data space interoperability architecture into building blocks, jointly implementing the capabilities for realising interoperability between multiple AI data spaces (as elaborated in the architecture vision in chapter 2) and the associated business architecture principles (as defined in chapter 3).

4. INFORMATION SYSTEM ARCHITECTURE PRINCIPLES

The Information Systems Architecture (ISA) gives a breakdown of the inter AI data space interoperability architecture into building blocks, jointly implementing the capabilities for realising interoperability between multiple AI data spaces as elaborated in the architecture vision and the associated business architecture principles as described in the previous chapters.

4.1 Information System Architecture principles

The Information System Architecture (ISA) deploys a set of principles for realising the business vision and business architecture principles for inter AI data space interoperability as defined in the previous chapters.

ISA.1. *Both the full and partial harmonisation inter data space interoperability modes may need to be supported for inter AI data space interoperability*

As described in section 3.2, full harmonisation between data spaces provides major advantages for inter data space interoperability, both functionally and on ease and efficiency in realisation. Its building blocks are relatively easy to develop as federable versions/extensions of the basic data space building blocks as will be addressed in section 5.2. However, full harmonisation is often not feasible in practice and may also be an utopia for newly formed data spaces. Therefore, partial harmonisation also needs to be developed for interconnecting heterogeneous AI data spaces.

ISA.2. *Emerging, fully distributed, reference architectures for federative data sharing will minimize the need for centralised data space interconnectivity building blocks*

A fully federated approach for inter AI data space interoperability allows a high level of autonomy for each of the individual AI data spaces in (1) shaping its own internal agreement framework and technical infrastructure, and (2) being minimally dependent on external stakeholders (to be trusted) providing capabilities for realising inter data space interoperability. However, a minimal set of (centralised) data space interoperability capabilities will need to be provided by the combination of data space interconnectivity intermediary roles and data space interconnectivity governance roles, see section 2.2 and **Figure 3**. The current EU developments on reference architectures for federative data sharing and data spaces (IDSA, GAIA-X, DSBA, ...) are developing towards fully distributed data space and trust framework capabilities, including distributed capabilities for IAA, contract negotiation and usage control, which may minimize the need for centralised data space interconnectivity building blocks.

ISA.3. *For the full harmonisation inter data space interoperability mode, federable versions/ extensions of the data space building blocks are foreseen*

Standardised interoperability for the various instances of AI data space building blocks over multiple data spaces will both minimize loss of functionality and minimize efforts for implementation thereof. Hence, for full harmonisation the goal is to develop federable versions/extensions of the standardised data space building blocks and aim for adoption thereof by the international data space reference architecture initiatives. Federability can for instance be applied to the IDS data space building blocks, i.e. a federable IDS DAPS, a federable metadata broker, a federable app store and a federable vocabulary hub, as will be addressed in section 5.2.

ISA.4. *For the partial harmonisation inter data space interoperability mode, multiple harmonisation profiles are foreseen*

The Data Sharing Coalition has developed an (initial) Use Case Implementation Guide (UCIG) [19] defining a harmonisation profile that can directly be implemented by means of proxies to interconnect data sharing domains. The harmonisation profile in the UCIG focusses on the identification, authentication and authorisation (IAA) functions and builds upon an OAuth2.0 IAA-protocol. However, as described in the companion report [2], the architecture for AI data spaces builds upon the IDS Reference Architecture Model (IDS RAM) [20][21], which uses an IAA protocol based on policy negotiation and policy enforcement, which is essentially different from the OAuth2.0 protocol.

Hence, for interoperability between AI data spaces that adopt the reference guides for AI data spaces and partial harmonisation for inter data space interoperability, a harmonisation profile that is tailored to the specific needs and protocols of IDS-based data spaces is expected to be both (1) more suitable for retaining the advanced capabilities that IDS provides to the individual AI data spaces, i.e. with minimal loss of capability, and (2) most effective in the number and complexity of transformations needed in the harmonisation domain. This implies that multiple harmonisation profiles may need to be defined, better matching with the requirements for the variety of trust frameworks and protocols being deployed in the various types of data spaces. An initial approach on how to identify and define an adequate set of harmonisation profiles has been presented in [22].

Furthermore, for publishing the ICT-resources within AI data space instances to other AI data space instances, a federated catalogue approach based may be used in the partial harmonization mode, e.g. using the DCAT-AP as harmonization profile, as will be further elaborated in paragraph 5.3.3.

4.2 Demarcation between AI data spaces

As described in section 3.2 and expressed ISA.1 in the previous section, both the full and the partial harmonisation modes may be considered for inter AI data space interoperability. Both have their own demarcation points and APIs to support inter AI data space interoperability, as depicted in **Figure 6**.

The figure depicts both the interoperability of the domain of the data services provider with an AI data space with (left side of the figure, see also the companion report [2], **Figure 6** and **Figure 13**) and the interoperability between multiple AI data spaces (right side of the figure).

As the figure shows, the following demarcation points are distinguished as APIs for interoperability between multiple AI data spaces:

- *Federable building block API for full harmonisation.*

  In case of full harmonisation, the federable building block API provides the 'external' point-of-connection per building block of the data space. For the federable building APIs various patterns of federation may be used, as will be described in paragraph 5.2.

- *Data space proxy API for partial harmonisation*

  In case of partial harmonisation, the data space proxy API provides the 'external' point-of-connection to a data space. A data space proxy itself is part of the individual data spaces. The data space proxy APIs expose the individual data spaces. The APIs conform to the protocols as defined for the individual harmonisation profiles for the harmonisation domain.

- *Federation API for data space interconnectivity capabilities.*

  To enable interoperability between data spaces, a minimal set of data space interconnectivity capabilities is needed (see section 4.2), which may be accessed by means of the Federation API.

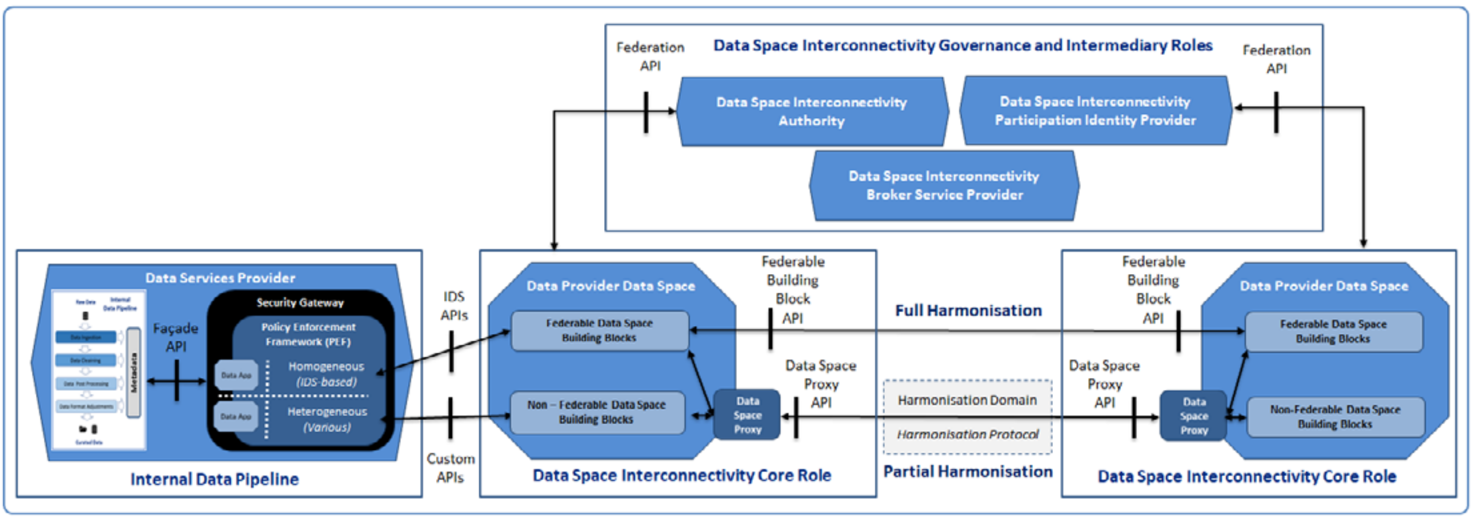

Figure 6 - Demarcation points and APIs to support inter AI data space interoperability with full and partial harmonisation.

4.3 Building blocks

In the ISA, the building blocks required for realising the various roles in the business role model are defined. A building block provides a (software) implementation of a capability to be performed by a role in the business role model. Table 2 provides the overview of the building blocks in the ISA for inter data space interoperability, categorised into building blocks for the data space interconnectivity core roles, intermediary roles and governance roles, respectively.

Figure 7 shows how the building blocks can be attributed to each of the roles in the business role model for inter AI data space interoperability for deployment.

Table 2: Building Blocks in the ISA for Inter Data Space Interoperability.
Data Space Interconnectivity Governance Role Building Blocks Capabilities to manage the various types of identities for multiple data spaces.
Data Space Interconnectivity Membership Certificate Authority (DSIM CA) Provides (X.509) certificates for participants and systems involved in data sharing within the AI data space and is used for verifying AI data space membership in data sharing transactions.
Dynamic Data Space Attribute Provisioning Service (DDSAPS) Manages and registers the dynamic attributes of the participating data spaces, including the certification status, data space interconnectivity membership status, applicable legal agreements and reference to data space building blocks (e.g. data space proxies and federated catalogues).
Data Space Interconnectivity Intermediary Role Building Blocks Capabilities to expose, find and connect to the various data spaces.
Data Space Interconnectivity Metadata Broker (DSI MB) Manages, registers and publishes the features of the various data spaces.
Data Space Interconnectivity Core Role Building Blocks Capabilities to be in control over the sharing of data and AI-algorithms, enabling data sovereignty for the entitled party of data or AI-algorithms
Federable Building Blocks The enabling building blocks within a data space that have the capabilities to be federated with the corresponding building blocks in other data spaces based on a full harmonisation mode. Federable building blocks can for instance be applied for the federable DAPS, the federable metadata broker, the federable app store and the federable vocabulary hub.
Non-Federable Building Blocks The enabling building blocks within a data space need capabilities of the partial harmonisation mode to interact with corresponding building blocks in other data spaces.
Data Space Proxy Translates between specifications and requirements from a data sharing domain and harmonised specifications and requirements (and vice versa) to achieve interoperability and trust across domains.
Harmonisation Profile The harmonised (technical) protocols used within the harmonisation domain, i.e. to communicate between data space proxies.

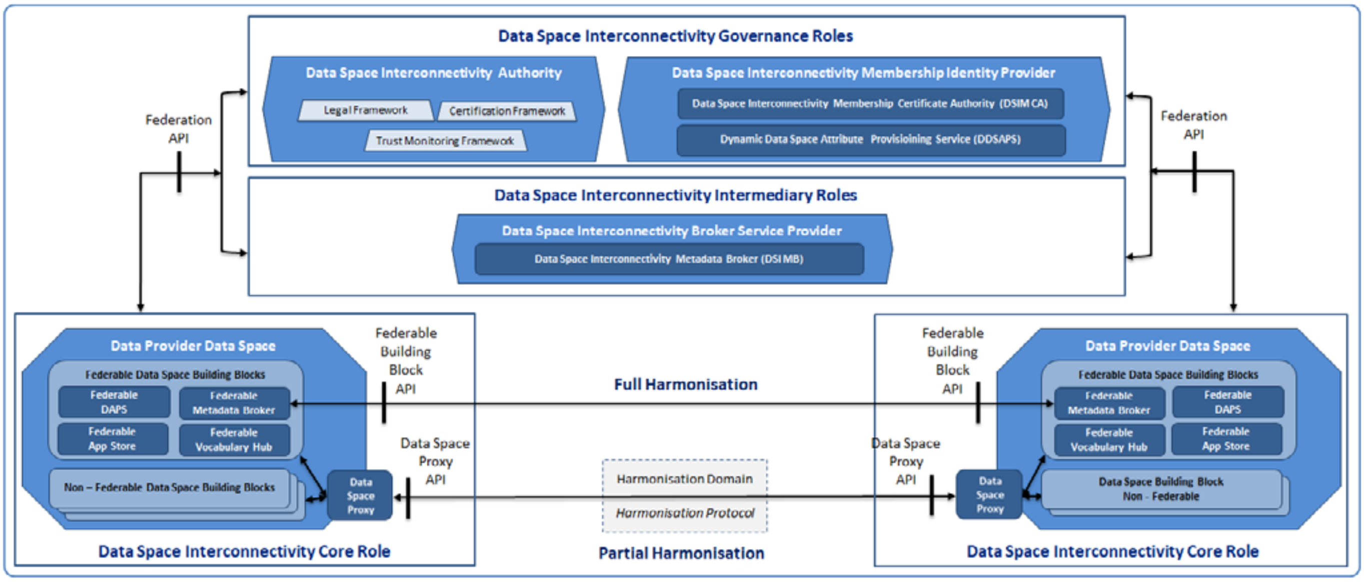

Figure 7 - ISA: break-down of business roles for inter AI data space interoperability into building blocks.

5. TECHNOLOGY ARCHITECTURE

The technology architecture describes the architecture principles and design of the buildings blocks as defined in the ISA and listed in section 4.3. After the technology architecture development approach (in the following section), the subsequent sections in this chapter elaborate each of the building blocks as listed in section 4.3.

5.1 Technology architecture development approach

The technology architecture development approach is elaborated in this section in terms of the technology architecture principles.

To a major extend, the technology architecture principles for intra AI data space interoperability as defined in the companion report [2], section 5.1, apply for inter AI data space interoperability as well. In addition, the following technology architecture principles specifically hold for inter AI data space interoperability:

TEC.1. For the various technical interoperability aspects, independent design decisions can be made for implementing (1) the full and/or partial harmonisation inter data space interoperability modes and (2) either in a centralised or distributed manner for the associated enabling building blocks

Inter data space interoperability applies to each of the four levels of the EIF as depicted in **Figure 5** (section 3.2) and as elaborated into further interoperability aspects per level in the companion report [2], section 3.3. For the different aspects of technical interoperability, an independent design decision can be made on using a full and/or partial harmonisation mode to enable interoperability between data spaces. Moreover, for the associated enabling building blocks the decision can be made to implement it either in a centralised or distributed manner. For the full harmonisation mode, this applies to the federable building blocks as will be addressed in paragraph 5.2. For the partial harmonisation modes, this applies to the data space proxy as will be addressed in paragraph 5.3.1. A distributed manner may be preferred in case of sharing of sensitive or valuable primary data or metadata. For instance for the data space proxy in the partial harmonisation mode, peer-to-peer connectivity between the data provider and data consumer may be preferred. Hence, at the secure connectivity level the associated data space proxy capability may be implemented in a distributed manner, i.e. by each of the individual data services providers, e.g. as separate data app within the security gateway.

TEC.2. By default and where applicable API definitions are based generally accepted standards

By adhering to generally accepted standards the adoption of the technology architecture will be stimulated and the efforts of integration will be minimised. However, as also indicated in section 4.1, the development of the architecture, building blocks, interfaces and the standardisation thereof are still in their infancy. Their development is expected to take considerable time to mature.

In the following sections, each of the individual building blocks as listed in 4.3 is elaborated, addressing both the architecture, the APIs and (open-source) implementations, respectively. 5.2

5.2 Full harmonisation: federable building blocks

As the Data Sharing Canvas [18] describes, full harmonisation between data sharing domains exist when the domains use or follow a shared cross-domain design, i.e. follow the same technical protocols and speak the same language.

Full harmonisation has implications for the interactions between associated instances of a building block in the various data spaces. These separate instances of a building block for various data spaces have to interact such that they jointly act as a single instance towards the users thereof, i.e. they should be able to be federated. As such, they are referred to as 'federable building blocks'. Federability applies to various building blocks in the data space architecture, as will be addressed in paragraph 5.2.3.

Various interaction scenarios for federation between building blocks can be distinguished:

- *Building block initiated federation*, further distinguishing between *federation at publish-time* and *federation at query-time*, and
- *Connector initiated federation*, further distinguishing between *service consumer initiated federation* and *service provider initiated federation*.

The following paragraphs in this section elaborate both interaction scenarios for federation and their sub-categories.

5.2.1 Building-block initiated federation

In case of building block initiated federation, the data space building blocks take care of federation of their capabilities between their instances in multiple data spaces. A further distinction can be made between federation at publish-time and federation at query-time, which are described in the following subparagraphs, respectively.

5.2.1.1 Federation at publish-time

In publish-time federation, the federation between the associated building block instances in different data spaces requires regular synchronisation of the stored meta data across each other. In **Figure 8** publish-time federation is depicted for the case of federation of the metadata broker building block across data spaces.

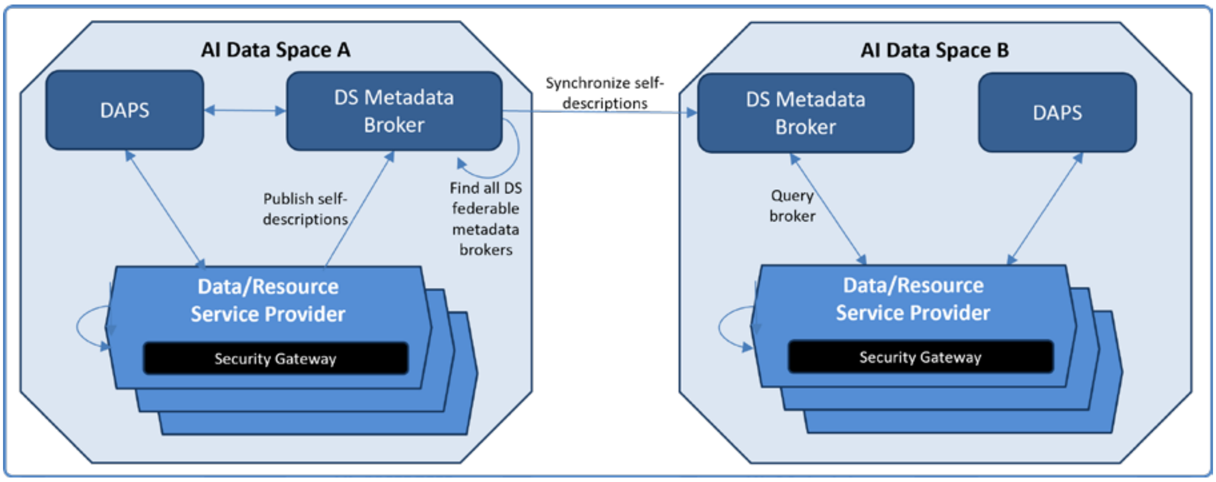

Figure 8 - Federation at publish-time for the case of federation of the metadata broker building block.

For federation at publish-time, all stored metadata in the building block instances is available to all other associated building block instances in joint network thereof. Data space participants (especially data services consumers and data services providers) can find and access the metadata from the building block instance in their own data space, which contains a replica of the same metadata stored at any of the associated building block instances in another data space.

5.2.1.2 Federation at query-time

In query-time federation, the associated building block instances in different data spaces do not regularly synchronize their metadata. Instead, when from a data services consumer to a building block in its own data space does not yield the requested information, that specific building block propagates the request to all other (previously configured) associated building block instances in other data spaces. In **Figure 9**, query-time federation is depicted for the case of federation of the metadata broker building block across data spaces.

For federation at query-time, all stored metadata remains in the building block instance of its own data space until explicitly requested for. Each building block maintains metadata sovereignty within its own data space, responding only with useful metadata to other building block instances when requested. Requesting data services consumers get an aggregated response to their queries that abstract away from the interaction structure between building block instances interconnected in the network of federated building blocks. All access to the metadata in building blocks in other data spaces is funneled through the local associated building block instance

5.2.2 Connector initiated federation

In case of connector initiated federation, the (security gateway/connector of) individual data space data services providers or data services consumers take care of federation of the capabilities between the building block instances in multiple data spaces. A further distinction can be made between 'service consumer initiated federation' and 'service provider initiated federation', which are described in the following subparagraphs, respectively.

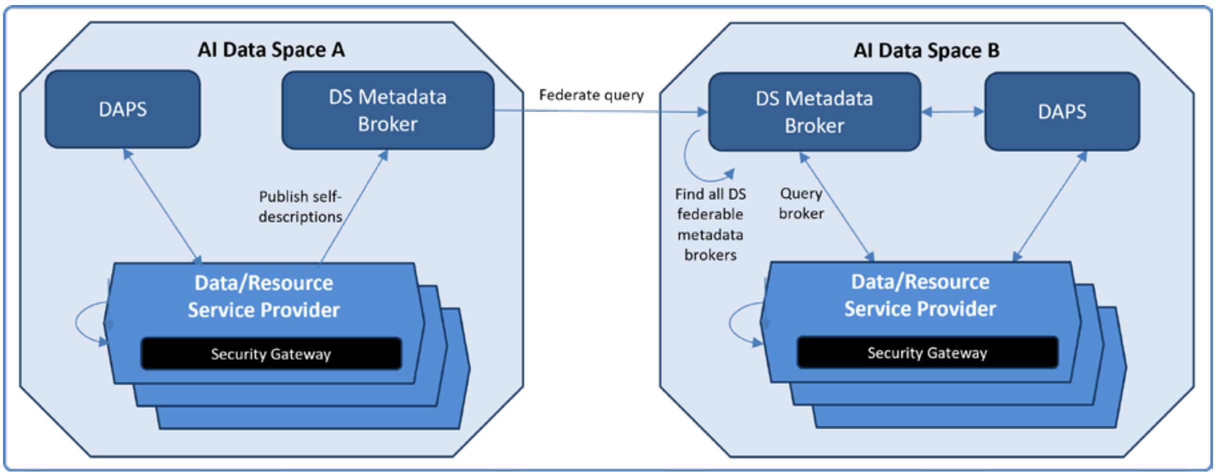

Figure 9 - Federation at query-time for the case of federation of the metadata broker building block.

5.2.2.1 Service consumer initiated federation

Service consumer initiated federation provides a mechanism for a metadata discovery flow that is directed and initiated by a data services consumer. The data services consumer queries the building block in its own data space for metadata regarding other associated building blocks in other data spaces. It then uses context-specific logic to send queries to a subset of the received data space instances for the required metadata. Querying the associated building block in another data space uses the same querying mechanism as the local building block instance in its own data space. The data services consumer has to manage aggregation of all the responses from the various building block instances. In **Figure 10**, service consumer initiated federation is depicted for the case of federation of the metadata broker building block across data spaces.

For service consumer initiated federation, all stored metadata remains in the building block instance of its own data space until explicitly requested for. Hence, each building block maintains metadata sovereignty within its own data space, responding only with useful metadata to data services consumers in (other) data spaces when requested. Requesting data services consumers get multiple replies for which they have to do the aggregation.

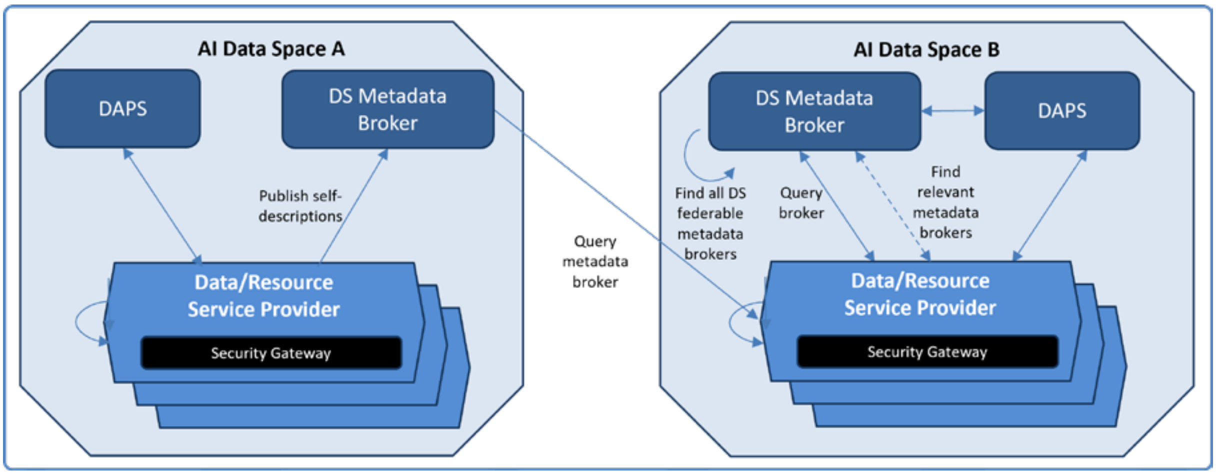

Figure 10 - Service consumer initiated federation for the case of federation of the metadata broker building block.

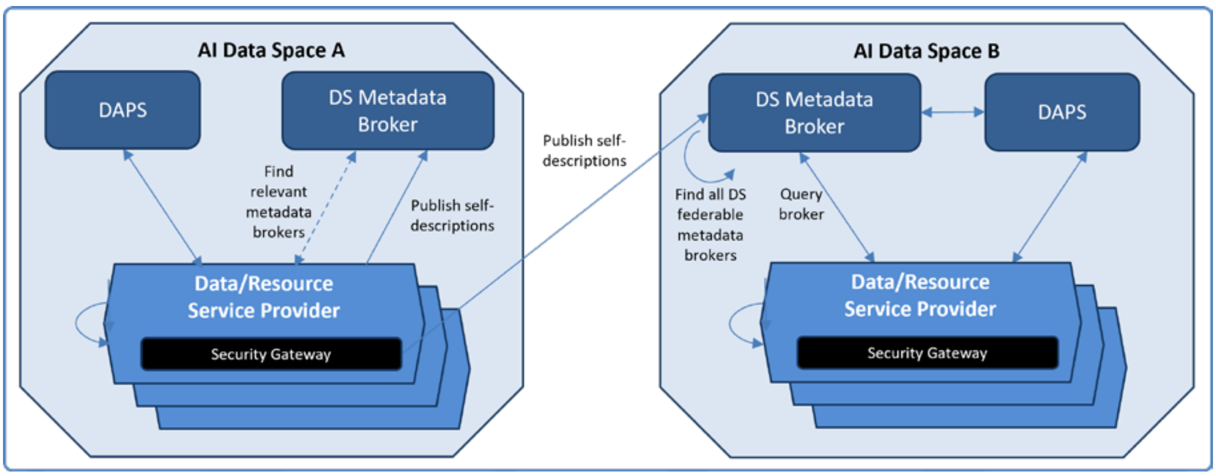

Figure 11 - Service provider initiated federation for the case of federation of the metadata broker building block.

5.2.2.2 Service provider initiated federation

In case of service provider initiated federation, the data services providers drive the publication of metadata to one or more individual building block instantiations. The building blocks do not have interaction with the associated building blocks in other data spaces. Data services providers may be aware of the building block instances in their own data space and in other data spaces, and may choose to publish metadata to each of them based on context-specific logic. Publishing metadata to the associated building block in another data space uses the same mechanism of publishing as that for the local building block instance in its own data space. In Figure 11, service provider initiated federation is depicted for the case of federation of the metadata broker building block across data spaces.

For service provider initiated federation, the building blocks (e.g. the metadata brokers) may only be aware of the metadata published to them by the data services providers. This is the proposed software architecture for the federated catalogue building block as defined in the Gaia-X Federated Services (GXFS) [23][24], where data services providers drive the publication of resources to one or more individual catalogues[1]

5.2.3 Federable building blocks: assessment of interaction scenarios

Criteria that are taken into account for assessing the various interaction scenarios for federation are:

- *Scalability*, estimated by a combination of how frequent publication of metadata may be, as well as how frequent queries to those resources may be.
- *Latency*, estimated by a combination of traffic generated by a single query or publication of metadata, as well as the speed of a response.
- *Sovereignty*, estimated by how much data or metadata is shared across entities or outside data spaces.

Table 3 provides a high-level assessment of the various interaction scenarios for federation on these three criteria.

Applicability for meta brokering and for identity management is defined in the context of the requirements that drive the design process, which are defined by different business models and technical circumstances. In practice, we may have hybrid solutions that achieve a balance between the requirements of all the stakeholders.

Table 3: Assessment of the various interaction scenarios for federation on the criteria: scalability, latency, sovereignty.

	Scalability	Latency	Sovereignty
Publish-time Federation	Very low	Very low	Low
Query-time Federation	Good	Moderate	High
Service-consumer Initiated Federation	High	Low[1]	High
Service-provider Initiated Federation	Good	Low[2]	Moderate [3]

1 The Software Requirements Specifications for the GXFS Federated Catalogue are a more detailed standard that define some additional planned capabilities (such as enabling subscription to updates in self-descriptions), but the core aspects of the service provider initiated federation are as described.

1, 2, 3 Depends upon the frequency of updates, and targets of queries/updates.

5.2.4 Federable building blocks: applicability to building blocks

Paragraph 5.2.1 and paragraph 5.2.2 have described the various interaction scenarios for federation of building block instances between multiple instances in a generic manner, i.e. equally well applicable for various types of building blocks. In this paragraph, the federable building blocks are considered for various types of data space building blocks as described in the companion report [2]:

- *Federable metadata broker*

  The goal of metadata brokers is to enable the matching of data service offerings between data services providers and data services consumers. Data services providers register their data services in a metadata broker and the data services consumers are able to query the broker. Interoperability between AI data spaces requires inter data space discovery of - and accessibility to – data services metadata across data spaces. This, in turn, requires federation of metadata brokers across data spaces, with federable metadata broker indicating the feature of a data space metadata broker to be able to federate with associated metadata brokers in other data spaces according to (a subset of) the interaction patterns for federation as described in paragraph 5.2.1 and paragraph 5.2.2, which already considered the case of federation of metadata brokers in its illustrative and representative figures.

  A federable metadata broker is currently being developed as part of the interoperability initiative between IDS-based data spaces in the industry sector. Its architecture and interfacing are being considered for inclusion in the IDSA RAM and Open DEI framework.

  Furthermore, it is noted that a 'network of interconnected (federable) metadata brokers' sometimes is also referred to as a '*Federated Catalogue*', although in this report we reserve this term for the interconnectivity of metadata brokers by means of a harmonisation profile in the partial harmonisation mode as will further be elaborated in paragraph 5.3.3.

- *Federable DAPS*

  For interoperability between multiple data spaces mutual trust is a key requirement. For trust in the identities as used in the independent data spaces, the federation of the DAPS instances of the various data spaces can be applied.

  A federable DAPS has the capability to exchange certificates (i.e. public keys) of the DAPS and identity manager with a federable DAPS in other data spaces. Once these public keys have been exchanged, a metadata broker can accept connections with data space access tokens from another data space. This takes care that identities and identity management remain within the originating data space and there is no need to exchange user identities. Moreover, it also omits the necessity for security gateway to connect to a DAPS in another data space.

  A federable DAPS is currently being developed as part of the interoperability initiative between IDS-based data spaces in the industry sector and is being considered for inclusion in the IDSA RAM and Open DEI framework

- *Federable app store*

  TNO currently develops an app store implementation, see the companion report [2], paragraph 5.4.2. Although the app store could also be federated to enable interoperability across data spaces by offering a single point of entry to find and use data apps, a version of the federable app store is not (yet) being developed.

- *Federable vocabulary hub*

  TNO currently develops a vocabulary hub implementation, see the companion report [2], paragraph 5.4.3. Although the vocabulary hub could also be federated to enable interoperability across data spaces by offering a single point of entry to find and use semantic models, a federable version of the vocabulary hub is not (yet) being developed.

For accessing federable building blocks, i.e. the federable metadata broker or the federable app store in an external data space, the federated DAPS in each of the individual data spaces must be configured with each other. This is a pre-requisite for interoperability between IDS-based data spaces (as adopted for the AI data spaces).

5.3 Partial harmonisation: building blocks

As described in section 3.2, the Data Sharing Canvas [18] introduces partial harmonisation through a new module, called a data space proxy. The data space proxy absorbs the complexity of harmonisation of data spaces. Data space proxies allow Data Service Consumers and Data Service Providers to simply connect to other data spaces via the data space proxy within their own data space.

Data space proxies provide the main capability of translating data space specific protocols and transactions to their harmonised equivalents. As such, they facilitate interoperable transactions and create a common understanding of concepts like trust and security across data spaces. To this end (and as depicted in Figure 6) the concept of a harmonisation domain is introduced. The harmonisation domain is the (virtual) domain between data spaces and data space proxies for which common 'intermediary' protocols are defined for harmonisation. The common 'intermediary' protocols are referred to as 'harmonisation profile'. Therefore, the data space proxy and the harmonisation profiles are addressed as building block in the following paragraphs of this section, respectively.

It is noted that the work on partial harmonisation by means of data space proxies and harmonisation profiles are in the early phases. They still have to prove their technical and market viability and are not yet sufficiently mature to be deployed at a large scale on the shorter term.

The Data Sharing Coalition has defined an initial harmonisation profile in its Use Case Implementation Guide (UCIG) [25]. Based on this UCIG, a representative proof-of-concept (PoC) was realised for controlled sharing of privacy-sensitive geriatric data between data spaces in the health domain as part of the NL AIC reference implementation as described in the companion report for intra AI data space interoperability [2], chapter 8 and annex B.

5.3.1 Data space proxy

Data space proxies allow Data Service Consumers and Data Service Providers to connect to other data spaces via the data space proxy within their own data space. The data space proxy absorbs the complexity of harmonisation of data spaces. A data space proxy can operate at each of the four interoperability levels (technical, semantic, organisational and legal) and their and their various aspects as of the European Interoperability Framework (EIF) as described in section 3.2 and depicted in Figure 5.

For each of the interoperability aspects, an independent design decision can be made to implement its associated data space proxy capability as either 'centralised' or 'decentralised':

- *Centralised data space proxy capability*, in which the data space proxy capability is positioned/implemented as generic intermediary software building block of a data space servicing multiple/all participants in the data space.
- *Decentralised data space proxy capability,* in which the data space proxy capability is positioned/implemented at the edge of a data space, i.e. within security domain of a specific participant servicing a single participant (e.g. as data app running within its security gateway).

The 'centralised' and 'decentralised' architectural implementation options for data space proxy capabilities are visually depicted in Figure 12.

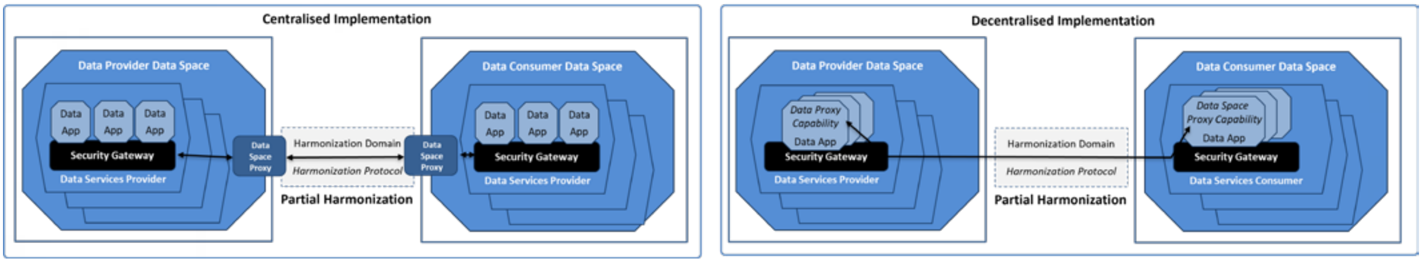

Figure 12 - Visualisation: categorisation of data space trust interaction patterns ([2], Figure 13).

5.3.2 Harmonisation profiles

The harmonisation domain (i.e., the domain between proxies, see Figure 7) uses a technical protocol, defined as a harmonisation profile. Ideally, this is a single protocol that supports all required capabilities 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 harmonisation profile can be used as a technical 'lingua franca' in the harmonisation domain. More likely, multiple harmonisation profiles are required to facilitate the interoperability for specific types of data spaces.

An initial exploration and structuring of-harmonisation profiles to support interoperable data sharing between a variety of data spaces has been presented in [22]. It was identified that a number of different harmonisation profiles were required to enable the required capabilities for data sharing between a variety of data spaces, based on two main trust aspects need to be addressed: policy management and the trust ecosystem.[2]

Moreover, the preliminary conclusion has been drawn that data space interoperability may be realised by a limited set of harmonisation profiles, for which two trust aspects (policy management and trust ecosystem) can be independently developed as part of the harmonisation profiles.[3]

5.3.3 Federated catalogue

In the partial harmonisation mode, the federated catalogue provides the capabilities for inter data space interoperability with respect to registration and findability of the data services across multiple data spaces. As such, it provides:

- *the catalogue capability*, i.e., how resources can be registered and exposed and made findable for users, and
- *the federation capability*, i.e., how individual catalogues for various data space can be federated into a 'virtually single' catalogue.

As the federated catalogue applies to the partial harmonisation mode, it implements the data proxy building block specifically for the cataloguing capability in which it translates the data space internal metadata brokering capability into the harmonised equivalent in the harmonisation domain to absorb the variation in protocols as used by various data spaces.

2 Policy management encompasses access- and usage policies. 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.

Independence of the aspects of policy management and the trust ecosystem implies that they may be developed independently, and their impact on the number of required harmonisation profiles is additive rather than multiplicative. The latter would -theoretically- reduce the required number of harmonisation profiles drastically.

For both trust aspects there is a limited number of varying protocols that act as a distinguishing factor for harmonisation profiles, of which the two main ones have been identified in this paper. Independence of both trust aspects suggests that the overall number of harmonisation profiles to be developed in the future remains limited and manageable.

3 For the trust aspects on policy management and the trust ecosystem, the UCIG [25] 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 have been addressed in [22]: (1) validation of independence of trust aspects, (2) validation of adequateness of individual harmonisation profiles and (3) validation of completeness of set of harmonisation profiles.

The DCAT standard [26] has been developed to standardize the exchange of metadata between catalogues. However, as the DCAT standard was considered to be too general, and European Commission has develop an application profile: DCAT-AP [27], which is a variant that is compatible with DCAT, but contains guidelines on which data are mandatory and which data are optional. As such, DCAT-AP is the European specification of the metadata that European data portals use for the exchange of metadata about data sets between the various catalogue instances. For instance, the central European data portal [28] as managed by the EU Publication Office, is based on DCAT-AP.

5.4 Data Space Interconnectivity Governance and Intermediary Role Building Blocks

As described in section 4.3 and Table 2 provides the building blocks for data space interconnectivity (for both governance and intermediary capabilities) consist of:

- the Data Space Interconnectivity Metadata Broker (DSI MB),
- the Data Space Interconnectivity Membership Certificate Authority (DSIM CA), and
- the Dynamic Data Space Attribute Provisioning Service (DDSAPS).

The development (and standardisation) of the building blocks and their capabilities is still in their infancy in the international reference architecture initiatives. For several of the capabilities, the iSHARE initiative currently provides an initial set of trust framework capabilities to support (interconnectivity between) data spaces [29][30].

Due to their immature status of development, the building blocks for data space interconnectivity (for both governance and intermediary capabilities) are not further elaborated in this report.

PART C: TRUST ARCHITECTURE

The trust architecture addresses the collection of agreements, policies, architecture and technical measures to assure that both the primary (potential sensitive and valuable) data and the metadata being shared within and between AI data spaces are trustworthy. The trust architecture enables sovereignty by entitled parties over their data, services and assets. Chapter 6 and chapter 7 extend the corresponding chapters in the companion report for intra AI data space interoperability [2] with specific aspects relevant for inter data space interoperability. As such, chapter 6 addresses the trust agreement framework and chapter 7 elaborates the trust interaction framework.

6. TRUST FRAMEWORK

Chapter 6 in the companion report [2] addresses the trust agreement framework for intra AI data space interoperability, distinguishing data space authority trust management (encompassing the legal framework, the certification framework and the system monitoring framework), data space identity management and data space policy management. For inter AI data space interoperability, the scope of these aspects needs to be extended to both applicability to participants over multiple data spaces and to applicability to multiple data spaces themselves.

Applicability of the the trust agreement framework to participants over multiple data spaces is associated to the harmonisation of data spaces, which has been defined in the Data Sharing Canvas [18] and in section 3.2 as 'the establishment of agreements, standards, and requirements between participants to enable data sharing between them', for which the two basic inter data space interoperability modes of full and partial harmonisation have been distinguished and elaborated chapter 5.

Applicability of the the trust agreement framework to multiple data spaces themselves implies that to trust the individual participants within each of the data spaces, there needs to be trust in the individual data space authorities that are responsible for managing, controlling and monitoring the trust framework within their associated data space. Hence, a 'new level' in the overarching trust framework arises for managing trust between data space authorities, i.e. on the specific topics for data space authority trust management: the legal framework, the certification framework and the system monitoring framework. The development of such inter data space authority trust management aspects are still in its infancy. It is noted that the iSHARE foundation has taken initial steps and provides capabilities for inter data space trust management [29][30]. Development of these capabilities are only starting to be addressed within the international reference architecture initiatives.

7. TRUST INTERACTION FRAMEWORK

The trust interaction framework ensures trustworthiness of the metadata being exchanged between the various building blocks in the AI data space, specifically the metadata being exchanged for identification, authentication and authorisation (IAA) and to define and ensure data sovereignty.

Chapter 7 in the companion report [2] addresses various AI data space trust interaction patterns for both intra and inter AI data space interoperability, specifically the homogeneous and heterogeneous trust interaction patterns:

- For intra AI data space interoperability, the homogeneous interaction pattern based on uniform (aligned) security gateways and the heterogeneous interaction pattern based on a hybrid security gateway have been distinguished.
- For inter AI data space interoperability, the homogeneous interaction pattern based on full harmonisation by means of federable building blocks and the heterogeneous interaction pattern based on partial harmonisation by means of data space proxies have been distinguished.

This categorisation is visually depicted in Figure 13.

For both intra and inter AI data space interoperability the (homogeneous and heterogeneous) trust interaction patterns have been elaborated in the companion report [2] (section 7.2) to which the reader is referred.

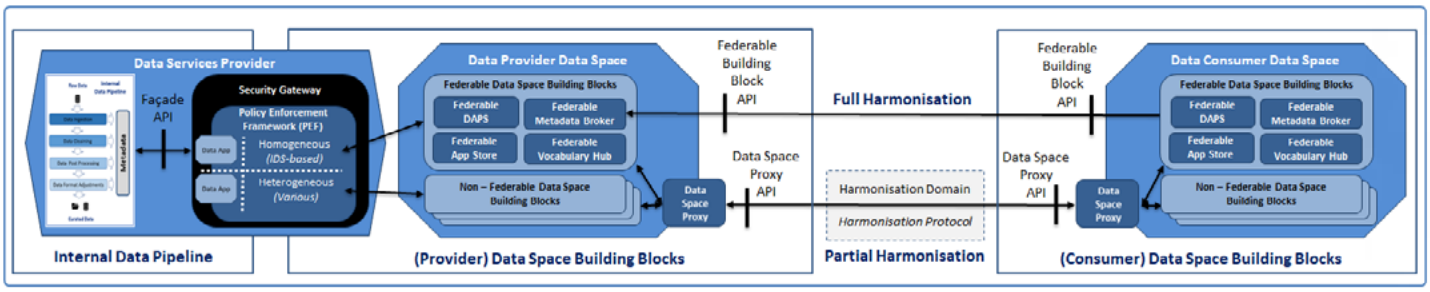

Figure 13 - Visualisation: categorisation of data space trust interaction patterns ([2], Figure 13).

PART D: REFERENCE IMPLEMENTATION, DEVELOPMENT ROADMAP AND CONCLUSIONS

The architectures, concepts and building blocks for realising inter AI data space interoperability as described in this report are still in their infancy. Further architectural development and guidance on adoption re needed. As starting point, chapter 8 describes the reference implementation to demonstrate the potential for the inter AI data space interoperability approach as described in this report. Subsequently, chapter 9 provides the further development roadmap for inter AI data space interoperability, after which chapter 10 provides the overarching conclusions.

8. REFERENCE IMPLEMENTATION

To demonstrate the potential and to identify lessons learned for developing towards large scale adoption, the architectural concepts and technologies for intra and inter AI data space interoperability (as described in this report and in the companion report [2]) are demonstrated by means of an illustrative and representative reference implementation.

The reference implementation shows how the various building blocks for intra and inter AI data space interoperability work together and can be integrated to implement the overarching architecture for a federation of interoperable AI data spaces as pursued by the NL AIC working group Data Sharing in alignment with the EU Data Strategy.

The scenario and story lines for the reference implementation focus on geriatric health care. Geriatric health care is used as it covers the various complexities and concepts of data sharing for AI, both applicable to intra and inter AI data space interoperability. It is considered both illustrative and representative due to:

- the privacy and sensitive nature of the data needed as input for AI processing, and
- the diversity in participants in providing and processing of geriatric data.

Various story lines for the reference implementation apply to the topic of inter AI data space interoperability as they address topics on technical or semantic interoperability to enable the controlled sharing of data between stakeholders in multiple data spaces. Nevertheless, for providing a complete and overarching view over the reference implementation, its scenario and its story lines, both those that are applicable to intra and inter AI data space interoperability aspects are jointly elaborated in the companion report on intra AI data space interoperability [2], chapter 8 and annex B. Specifically the story line on data sovereignty and technical (trust) interoperability and the story line on semantic interoperability (as elaborated in annex B.4 and annex B.5 in the companion report [2]) apply to inter AI data space interoperability.

9. DEVELOPMENT ROADMAP

As for the development roadmap for intra AI data space interoperability as addressed in the companion report [2], the development roadmap for inter AI data space interoperability distinguishes the three main views as used within this report: the ecosystem architecture, the building block architecture and the trust architecture. Figure 14 shows the overarching development roadmap for inter AI data space interoperability for the time period 2023 - 2025.[4]

The figure shows how the individual activities are grouped in various topics. The following sections address the specific development activities and their topics for the ecosystem architecture, the building block architecture and the trust architecture view, respectively.

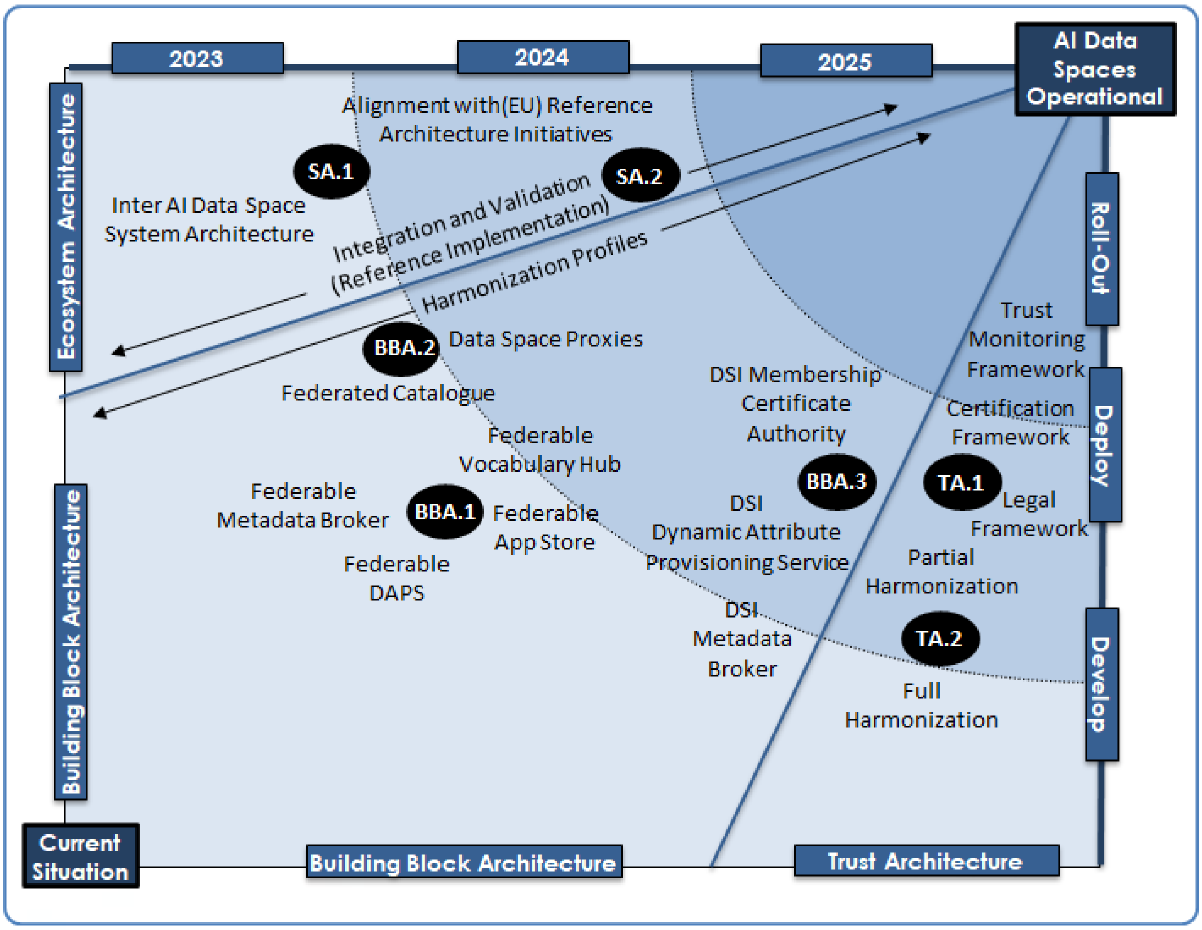

Figure 14 - Visualisation: categorisation of data space trust interaction patterns ([2], Figure 13).

9.1 Developing the ecosystem architecture

The ecosystem architecture view includes the activities for the inter AI data space system architecture (SA.1) and for integration and validation (SA.2).

The activity for the inter AI data space system architecture (SA.1) describes the decomposition of the architecture for inter AI data space interoperability into business roles and technical building blocks, as have been addressed in section 2.2 and in section 4.3, respectively. The development of the architectures, building blocks, interfaces and the standardisation for inter AI data space interoperability are still in their infancy and is expected to take considerable time to mature. Alignment with international reference architecture initiatives is therefore required, especially those as described in the companion report [2], paragraph 5.1.2: Open DEI, International Data Spaces Association (IDSA), Gaia-X, Data Space Business Alliance (DSBA), iSHARE and FIWARE.

The activity for integration and validation (SA.2) must be a continuously ongoing activity in which the architectures, building blocks, interfaces and standards for inter AI data space interoperability are demonstrated and assessed by the NL AIC working group Data Sharing on technical and market viability, e.g. by means of use case development and reference implementations.

9.2 Developing the building block architecture

The activities in the building block architecture view address the further development of the building blocks for the full harmonisation architecture (BBA.1), for the for the partial harmonisation architecture (BBA.2), data space trust architecture on identity management (BBA.2) and for the interconnectivity enabling architecture (BBA.3), respectively:

- The activity for the further development of the building blocks for the full harmonisation architecture (BBA.1) extends the set of (open source implementations of) federable building blocks as described in section 5.2. This may include the further development of the federable building blocks as enumerated paragraph 5.2.4, i.e. the federable DAPS, the federable metadata broker, the federable app store and the federable vocabulary hub. Additional federable building blocks may be considered. Their development is to be aligned with the relevant international reference architecture initiatives.

- The activity for the further development of the building blocks for the partial harmonisation architecture (BBA.2) should further explore the potential of the partial harmonisation mode for inter AI data space interoperability. As described in section 5.3, it is noted that the work on partial harmonisation by means of data space proxies and harmonisation profiles is in the early phases. It still has to prove their technical and market viability. The (limits to the) contribution of the NL AIC Data Sharing Working Group should therefore carefully considered.

  Furthermore, and as Figure 14 indicates it is noted that a 'network of interconnected (federable) metadata brokers' by means of a partial harmonisation approach with a harmonised federation profile (as also referred to as a '*Federated Catalogue*') may be considered an initial step to be elaborated on the shorter term.

- The activity for the further development of the building blocks for the interconnectivity enabling architecture (BBA.3) elaborates the building blocks for data space interconnectivity (for both governance and intermediary capabilities) as described in Table 2. Also the development (and standardisation) of these building blocks and their standardisation are still in their infancy in the international reference architecture initiatives. Alignment with these initiatives needs to be sought for further development thereof.

9.3 Developing the trust architecture

The activities in the trust architecture view address the further development of the inter data space authority trust management (TA.1) and the trust interaction framework (TA.2).

The inter data space authority trust management (TA.1) develops (in coherence) the three topics for inter data space authority trust management with applicability to multiple data spaces (as identified and described in chapter 6), i.e. the legal framework, the certification framework and the trust monitoring framework. The work on inter data space authority trust management is only starting to be addressed in the international reference architecture initiatives and is therefore foreseen later on in time.

The trust interaction framework for inter AI data space interoperability (TA.2) extends the work and scope of the trust interaction framework for intra AI data space interoperability as described in the companion report [2] (chapter 7) for applicability to data space participants being member of different data spaces. The work on the trust interaction framework based on a full harmonisation mode is mainly addressed by means of the federable data space building blocks as described in section 5.2 and needs to be continued and to be aligned and adopted by the international reference architectures on federative data sharing. The work on the trust interaction framework based on a partial harmonisation mode is still in its infancy and will continue over a longer time period.

10. CONCLUSIONS

This reference guide report has elaborated overarching architecture, building blocks and roadmap for the inter AI data space interoperability development line. It builds upon the lessons learned from the work of the NL AIC working group Data Sharing as done in 2021/2022, pursuing the goal of evolving towards a '*federation of interoperable data spaces*' as being the ambition of the EU Data Strategy. It has addressed both the ecosystem architecture, the building block architecture and the trust architecture.

It is noted that the major part on the AI data space interoperability architectures has been described within the companion report on intra AI data space interoperability [2]. This report has been restricted to further elaborating the companion report on those aspects relevant for interoperability between multiple AI data spaces.

As stated in the concluding chapter of the companion report [2], it is to be realised that reference architectures and standards are still in development. Alignment with the main (inter-) national reference architecture initiatives on federative data sharing should be a focal point for further development. These reference architecture initiatives are developing towards more fully distributed trust framework capabilities for identity, authentication and authorisation (IAA), contract negotiation and usage control. As such, this reference guide may help organisations with their initial steps towards AI data spaces, whilst being aware that the environment, architecture, concept and standards are evolving.

The adoption of AI data spaces is still in in its infancy. Data sharing communities and organisations can contribute to their further development by implementing proof-of-concepts and use cases for (federated) AI data spaces providing feedback and input for extension and improvement to the reference guides. The know-how and expertise of the participants of the NL AIC working group Data Sharing can provide a major contribution to this collaborative development and operationalisation of a (federation of) AI data spaces in the Netherlands and the EU.

REFERENCES

[1] The Netherlands AI Coalition working group Data Sharing (2021). "Towards a federation of AI data spaces - NL AIC reference guide to federated and interoperable AI data spaces". URL: https://nlaic.com/wp-content/uploads/2021/11/NL_AIC_Towards_a_federation_of_AI_data_spaces.pdf.

[2] The Netherlands AI Coalition working group Data Sharing (2022). "Reference guide for intra AI data space interoperability - Intra data space development line". URL: https://nlaic.com/wp-content/uploads/2023/04/reference-guide-for-intra-ai-data-space-interoperability.

[3] European Commission (2020). "A European strategy for data". URL: https://digital-strategy.ec.europa.eu/en/policies/strategy-data.

[4] NL AIC (2021). "About NL AIC". 2021. URL: https://nlaic.com/en/about-nl-ai.

[5] The Netherlands AI Coalition working group Data Sharing (2020). "Verantwoord datadelen voor AI". URL: https://nlaic.com/wp-content/uploads/2020/03/Verantwoord-datadelen-voor-AI.pdf.

[6] The Netherlands AI Coalition working group Data Sharing (2020). "Responsible data sharing in AI". URL: https://nlaic.com/wp-content/uploads/2020/10/Responsible-data-sharing-in-AI.pdf.

[7] The Netherlands AI Coalition working group Data Sharing (2020). 'Van First-time-Engineering naar Operationalisatie". URL: https://nlaic.com/wp-content/uploads/2020/08/NL-AIC-Naar-First-time-Engineering-en-Operationalisatie.pdf.

[8] The Netherlands AI Coalition working group Data Sharing (2020). "GAP-analysis- From data sharing proofs-of-concept towards operationalisation of the system architecture". URL: https://nlaic.com/wp-content/uploads/2021/03/NL-AIC-GAP-Analysis.pdf.

[9] The Netherlands AI Coalition working group Data Sharing (2021). "AI Ecosystem & Market Analysis - Quick scan of data sharing market to validate blueprint of the NL AIC". URL: https://nlaic.com/wp-content/uploads/2021/02/AI_Ecosystem_and_Market_Analysis_Data_Sharing_4-feb-2021.pdf.

[10] The Netherlands AI Coalition working group Data Sharing (2022). "Key insights AI use cases and demos". Available for NL AIC Community Members. URL: https://community.nlaic.com/groep-media/data-delen/221223-publication-version-nlaic-use-case-summaries.

[11] EU Digital Europe Programme. "Data Spaces Support Centre (DSSC)". URL: https://dssc.eu.

[12] EU Digital Europe Programme, "SIMPL: cloud-to-edge federations and data spaces made simple". URL: https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple.

[13] The Data Sharing Coalition. “Unlocking the true value of data”. URL: https://datasharingcoalition.eu.

[14] The Netherlands AI Coalition working group Data Sharing. “Data Sharing Building Block”. URL: https://nlaic.com/en/building-blocks/data-sharing.

[15] Gaia-X. “Gaia-X Hub - The Netherlands”. URL: https://gaia-x.nl/en.

[16] The Open Group, “TOGAF 9.1”. URL: https://pubs.opengroup.org/architecture/togaf91-doc/arch.

[17] European Union (2017). “New European Interoperability Framework (EIF) – Promoting seamless services and data flows for European public administrations”. URL: https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf.

[18] Data Sharing Coalition (2021). “Data Sharing Canvas - A stepping stone towards cross-domain data sharing at scale”. URL: https://datasharingcoalition.eu/app/uploads/2021/04/data-sharing-canvas-30-04-2021.pdf.

[19] Data Sharing Coalition (2022). “Use Case Implementation Guide”. URL: https://datasharingcoalition.eu/app/uploads/2022/03/data-sharing-coalition-use-case-implementation-guide-ucig.pdf.

[20] International Data Spaces Association (IDSA) (2019). “International Data Spaces: Reference Architecture Model Version 3”. URL: https://www.internationaldata spaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf.

[21] International Data Spaces Association (IDSA) (2022). “International Data Spaces: Reference Architecture Model Version 4”. GitHub URL: https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0.

[22] A. Stoter, B. Rietveld, V. Jansen and H. Bastiaansen. “Harmonisation Profiles for Trusted Data Sharing Between Data Spaces - Striking the Balance between Capability and Complexity”. Interoperability for Enterprise Systems and Applications Conference 2022 (I-ESA’22), Valencia, Spain. http://ceur-ws.org/Vol-3214/WS6Paper2.pdf.

[23] EU Gaia-X Initiative. “A Federated and Secure Data Infrastructure”, URL: https://www.gaia-x.eu.

[24] EU Gaia-X Initiative. “Gaia-X Federation Services - FXFS”. URL: https://www.gxfs.eu/specifications.

[25] Data Sharing Coalition (2022). “Use Case Implementation Guide”. https://datasharingcoalition.eu/app/uploads/2022/03/data-sharing-coalition-use-case-implementation-guide-ucig.pdf.

[26] W3C (2022). “Data Catalog Vocabulary (DCAT) - Version 2 “. URL https://www.w3.org/TR/vocab-dcat.

[27] European Commission. “DCAT Application Profile for Data Portals in Europe”. URL: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe.

[28] European Commission. “The Official Portal for European Data”. URL: https://data.europa.eu.

[29] iSHARE Foundation. “iSHARE – Trust Framework for Data Spaces”. URL: https://ishare.eu.

[30] iSHARE Foundation. “(Benefits) For Data Spaces”. URL: https://ishare.eu/ishare/benefits/for-data-spaces.

Colophon

This report is a result of the work of the NL AIC working group Data Sharing. In conjunction with the overarching AI data space reference guide '*Towards a Federation of AI Data Spaces*' [1] and the companion report 'Reference Guide for Intra AI Data Space Interoperability' [2]. This report provides guidelines for realising the ambition of the NL AIC working group Data Sharing of providing the data sharing foundation for AI in the Netherlands in alignment with the interoperable data space approach as pursued by the EU Data Strategy [2].

Both the intra and the inter data space reference guide report on work-in-progress.

Contributors
Participants in the NL AIC working group Data Sharing
TNO (editors)

Review
Advisory Board of the NL AIC working group Data Sharing

Contact:
The Netherlands AI Coalition
E-mail — info@nlaic.com
Website — nlaic.com

March 2023