NTT DATA
Trusted Global Innovator

Dataspace Connector Survey Report

～Overview of IDS-RAM and Eclipse Dataspace Connector～

NTT DATA INTELLILINK Corporation

July 27, 2022
(This document is based on the survey around March of 2022)

Table of Contents

1

Introduction

1. Introduction

In Europe, many organizations and companies support ideas and initiatives such as IDS (International Data Spaces) and Gaia-X in order to share data among a large number of companies. From the concept of data sovereignty to architecture, they actively discuss and publish concrete plans.

Among those efforts, the IDS Reference Architecture Model (IDS-RAM)[*1] has influenced the development of the Gaia-X concept[*2] and particular data sharing projects, such as Catena-X[*3].

This document describes the results of survay on "connectors," which are the core technology of data exchange including data access control, while referring to the IDS-RAM and other related materials.

For details on the relationship between IDS and Gaia-X and the latest information, please refer to the white paper[*4] published by the IDSA (The International Data Spaces Association).

1. Introduction

Sources

1. https://internationaldataspaces.org//wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf
2. https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/project-gaia-x.html
3. https://catena-x.net/en/vision-goals
4. https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf

2

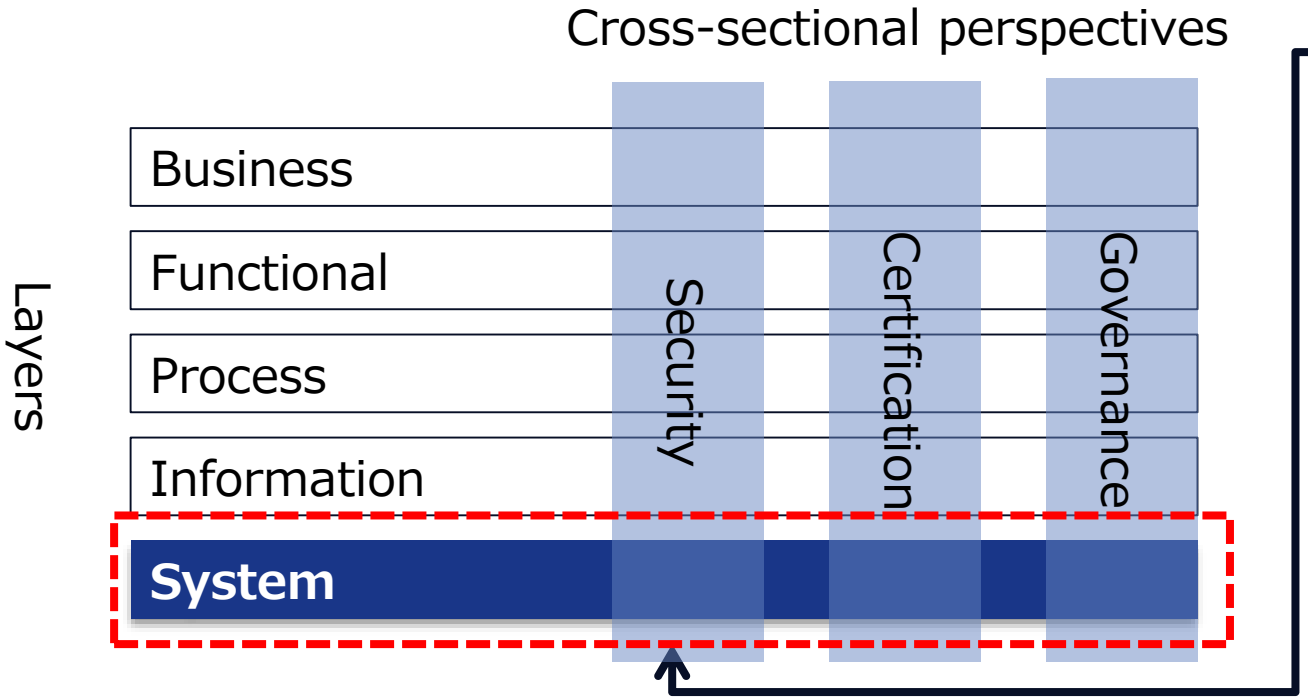
IDS Reference Architecture Model

2. IDS Reference Architecture Model

Overview

IDS-RAM is the highest abstract document that defines the components of IDS.

It defines the IDS in five layers and three perspectives. In the system layer, specific components including "connectors" are defined.

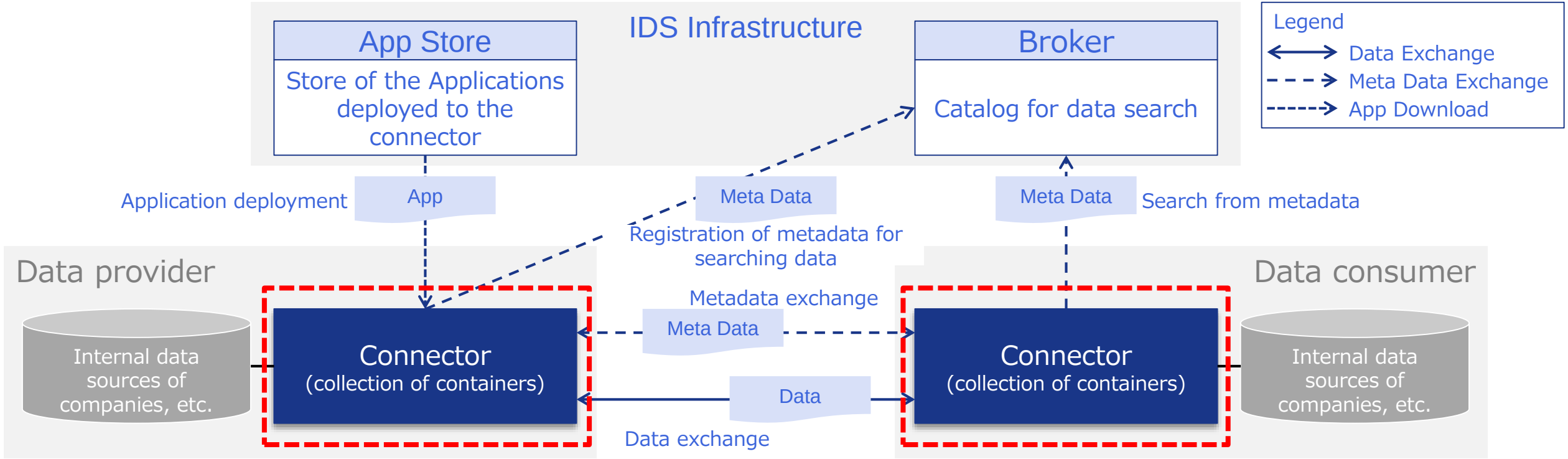


Classification	Items
Main components	Connector
	Broker
	App Store
Support components	Identity provider
	Vocabulary hub
	Update repository
	Trust repository

2. IDS Reference Architecture Model

Connector structure

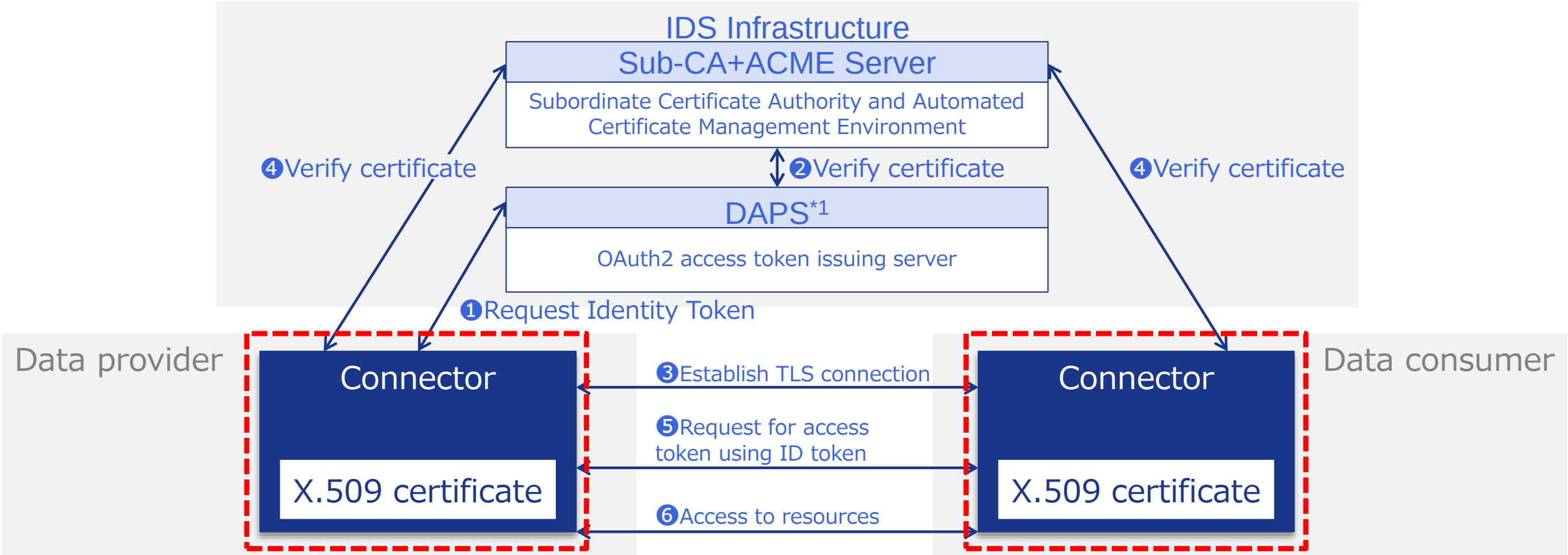
The connector acts as an entity or its proxy that performs the transfer of data in order to realize the process of data exchange between internal data sources among participating companies. Typically, the connector is composed as a collection of containers using container management technology. The diagram below, created by NTT DATA based on the contents of the IDS-RAM, shows the relationship between the major components.



2. IDS Reference Architecture Model

Communication between connectors

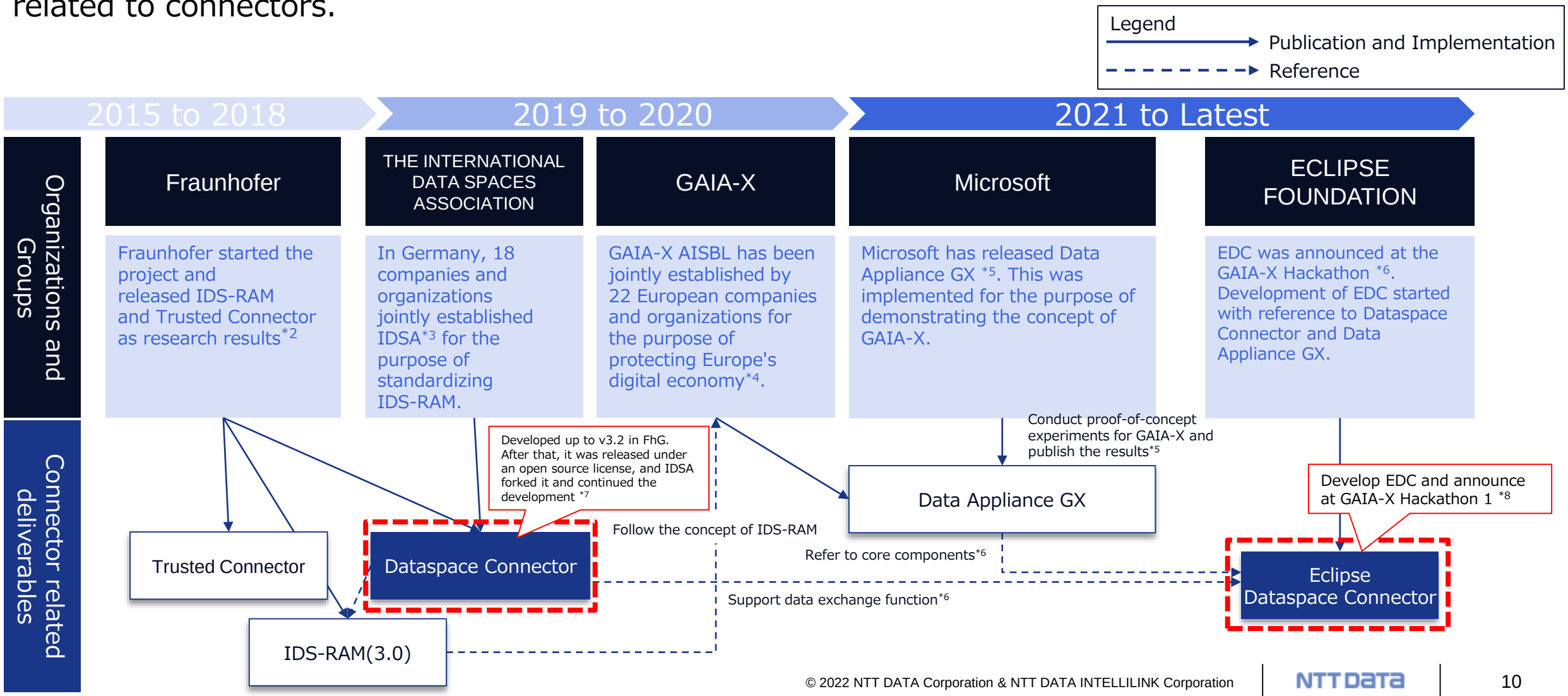
Since the connector acts as a data-exchanging entity or its proxy, some communication occurs during its execution. The diagram below, created by NTT DATA based on the contents of the IDS-RAM, shows the communication between the major components. Communication between connectors must go through an encrypted tunnel such as TLS.



2. IDS Reference Architecture Model

Organizations and Groups related to connectors

Below is a figure created by NTT DATA based on the survey results of organizations and groups related to connectors.



2. IDS Reference Architecture Model

Connectors to be surveyed

The following describes NTT DATA's interpretation of the characteristics of each connectors based on the survey results of organizations and groups related to the connectors.

Connector name	Brief description	Interpretation by NTT DATA
Dataspace Connector	The connector compliant with IDS-RAM standardized by IDSA, which was established by 18 German companies.	Reference implementation of IDS-RAM
Eclipse Dataspace Connector	The connector presented at the Gaia-X Hackathon of the Gaia-X AISBL, founded by 22 European companies, which incorporates the IDS-RAM concept.	Implementation that is being into practical use*

We investigated: the **Dataspace Connector** considered to be a reference implementation of IDS-RAM based on central element of GAIA-X, and the **Eclipse Dataspace Connector** considered to be a proof-of-concept implementation of GAIA-X.

*NOTE: It is being considered for adoption in Catena-X, a data sharing initiative in the European automobile industry*9.

2. IDS Reference Architecture Model

Sources

1. https://www.dataspaces.fraunhofer.de/en/software/identity_provider.html
2. https://www.fraunhofer.de/en/press/research-news/2015/september/Fraunhofer-initiative-for-secure-data-space-launched.html
3. https://www.fraunhofer.de/en/press/industrial-data-space-association-founded1.html
4. https://b5gnbsc.jp/wp-content/uploads/2022/04/R02-0049-0221_report.pdf
5. https://github.com/microsoft/Data-Appliance-GX
6. https://projects.eclipse.org/proposals/eclipse-dataspace-connector
7. https://github.com/FraunhoferISST/DataspaceConnector/releases/tag/v3.2.0
8. https://azure.microsoft.com/ja-jp/blog/gaiax-gets-new-support-with-european-eclipse-data-connector/
9. https://catena-x.net/en/angebote/edc-die-zentrale-komponente-fuer-die

3

Dataspace Connector

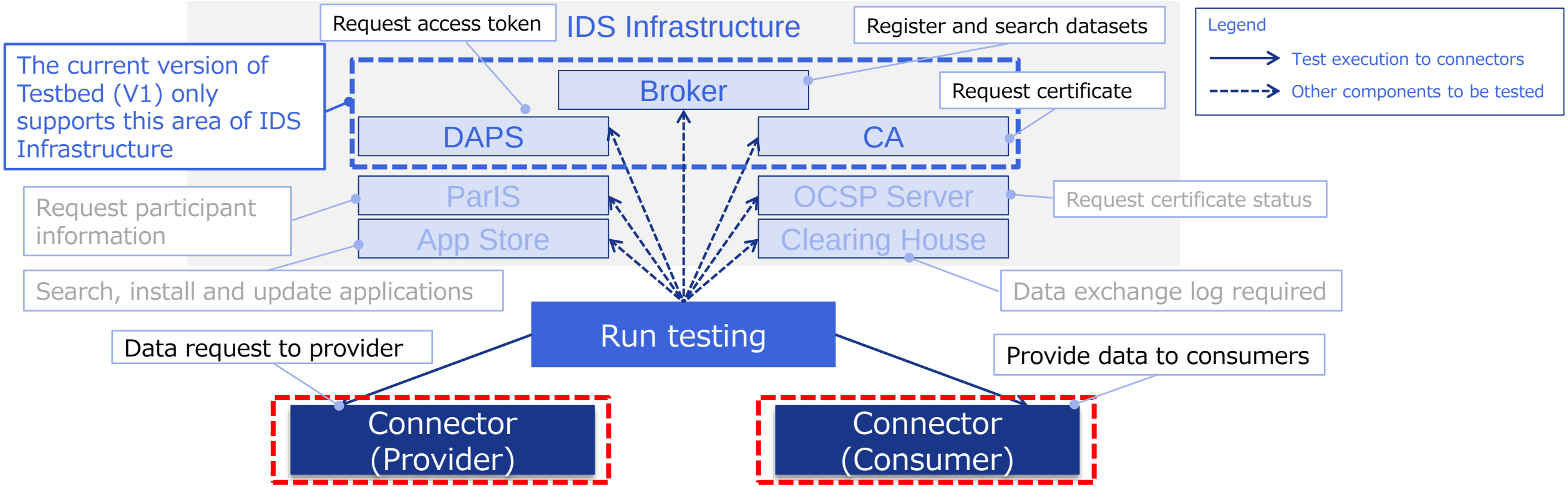
3. Dataspace Connector

Overview

IDS dataspace connector is a reference implementation of the IDS-RAM compliant Dataspace Connector standardized by IDSA, of which the Fraunhofer Society is a founding member.

It implements a data model called Information Model defined by IDS-RAM.

The diagram below, created by NTT DATA based on the component configuration published by IDSA as Testbed [*1], shows the overview of the Dataspace Connector.



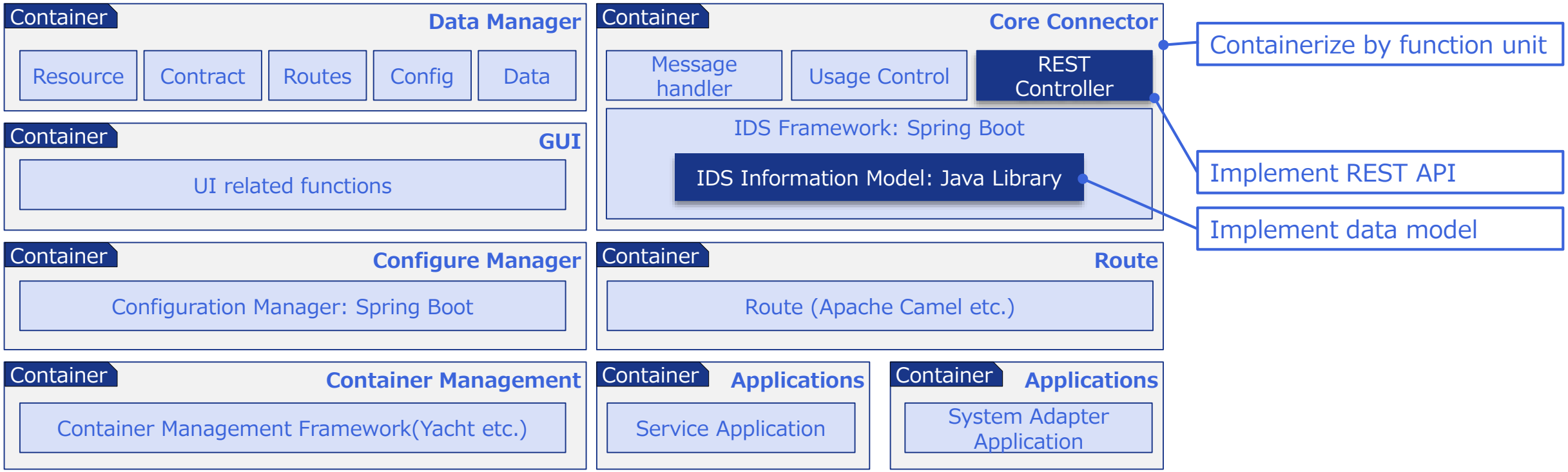
3. Dataspace Connector

Architecture

Dataspace Connector consists of core connector, route etc. and each of these functions are executed on a container. In order to keep the Dataspace Connector as lightweight as possible, each function is independent from each other, allowing it to be replaced by alternative technologies.

For example, instead of using Apache Camel for routing, Apache Airflow or Apache Kafka can be considered. In addition, from the IDSA's roadmap, it is believed that this structure is adopted in order to support Kubernetes or similar products in the future.

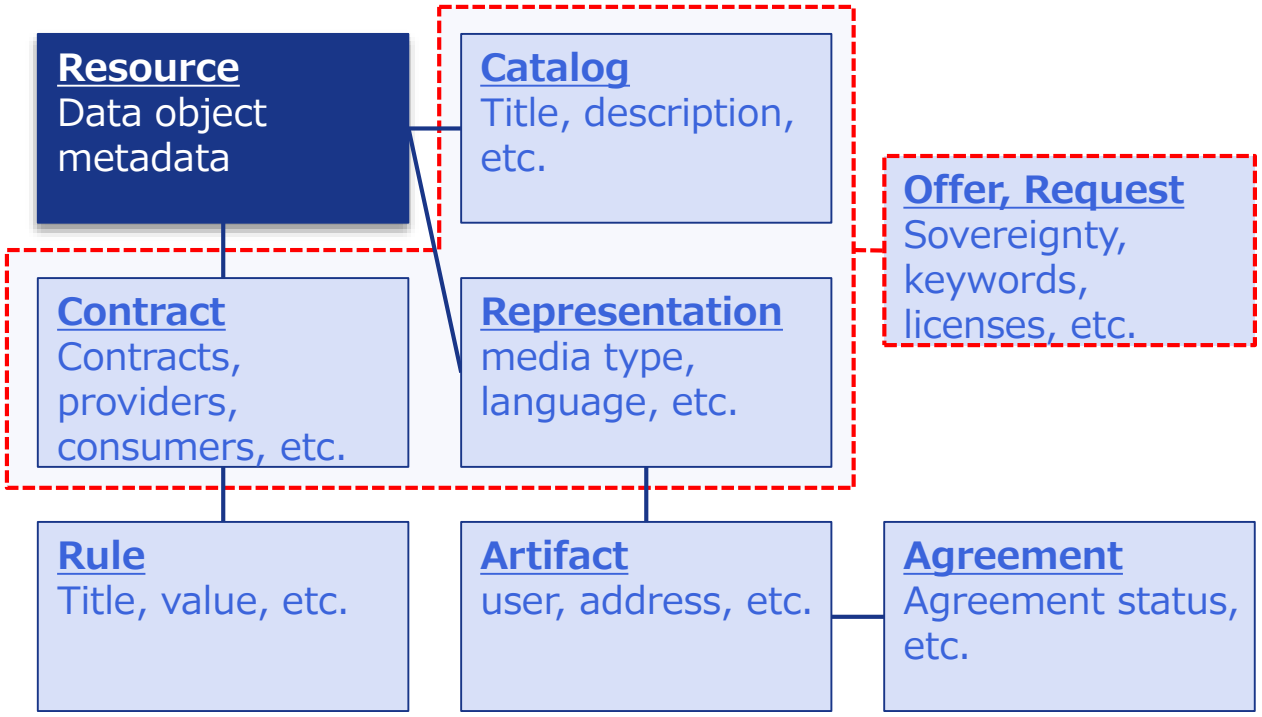
The diagram below shows the basic architecture of Dataspace Connector. It was created by NTT DATA based on the Dataspace Connector architecture diagram[*2].



3. Dataspace Connector

Data model

Dataspace Connector's data model is designed based on IDS-RAM. Connectors in the dataspace can handle unified message exchanges by returning data model-compliant responses to data requests. The diagram below, created by NTT DATA based on the IDS-RAM Information Model [*3], shows the basis of data model.



Resource

The top level of the data model. Defined as metadata for data objects.

Contract

Rules are defined for resource usage exchanged between resource requesters and resource providers.

Catalog

Holds metadata associated with this resource, such as connectors, participants, and resources.

Representation

Defines a description of what the resource is, such as resource type, language, title, etc.

Offer, Request

Refers to data providers and data requesters. For example, a contract issued by a data provider becomes a ContractRequest.

3. Dataspace Connector

Application Interfaces

APIs for accessing Dataspace Connector functions are available as OpenAPI documents. The table on the right side, created by NTTDATA based on the OpenAPI definition[*4] of Dataspace Connector, shows the APIs that are considered necessary for data exchange.

/api/connector

Acquire information about the connector itself, such as the title, information model version, and resource catalog.

/api/contracts

Handle contract negotiations to obtain data from data providers. Data cannot be obtained without agreement.

/api/catalogs

Obtain metadata about the requested data from the data provider.

/api/artifacts

Acquire actual data based on contractual agreement results.

Classification	PATH	Overview
Connector	/api/configurations	API for configuring connectors
	/api/connector	**API that provide its own self-description**
Data model	/api/representations	API for Representation of data model
	/api/contracts	**API for constructing data models**
	/api/catalogs	**API for data model catalog**
	/api/artifacts	**API for data model artifacts**
	/api/agreements	API for data model agreement
	/api/rules	API for data model rules
Resource	/api/subscriptions	API related to resource information reference
	/api/requests	API for resources (requests)
	/api/offers	API for resources (offers)
Linkage with other functions	/api/brokers	API for the Metadata Broker component
	/api/appstores	API for AppStore components
	/api/camel	API for Camel components
Camel	/api/routes	API for Camel routing
	/api/endpoints	API for endpoints on Camel routing

3. Dataspace Connector

Sources

1. https://github.com/International-Data-Spaces-Association/IDS-testbed
2. https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/v5/Architecture
3. https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/v6/DataModel
4. https://github.com/International-Data-Spaces-Association/DataspaceConnector/blob/main/openapi.yaml

4

Eclipse Dataspace Connector

4. Eclipse Dataspace Connector

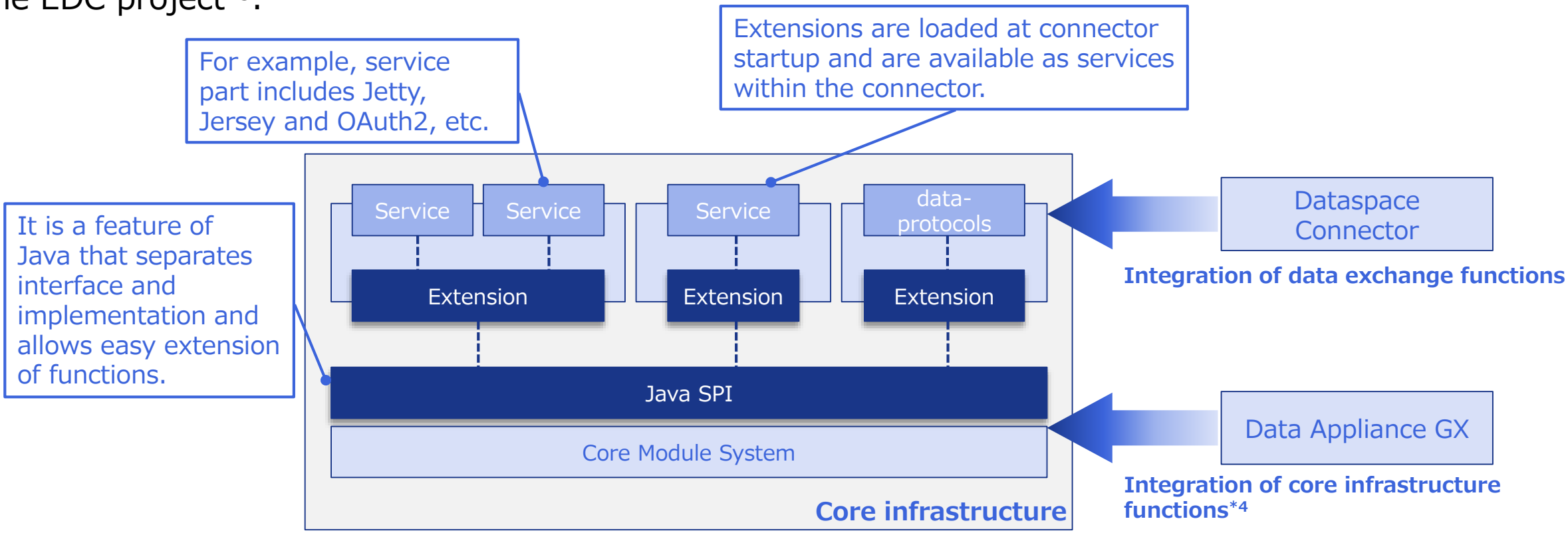
Overview

The Eclipse Dataspace Connector (EDC) implements technologies derived from the IDS dataspace standards and the Gaia-X's approach for data infrastructure. Also, this project will feed back their implementation and use cases to IDS and Gaia-X. The project is being implemented based on the planned technical milestones[*1], and it can be seen that the project is very active, with hackathons[*2] being held in collaboration with the Gaia-X project. Below is a figure created by NTT DATA based on the milestones and Hackathon materials.

4. Eclipse Dataspace Connector

Architecture

EDC is being implemented to integrate some of the functionalities of the Dataspace Connector by IDS.
And EDC is designed to allow for easy replacement or extension of functions.
The intent of its design appears to be to provide a flexible and extensible framework.
The following diagram was created by NTT DATA, based on the architecture diagram published by the EDC project[*3].



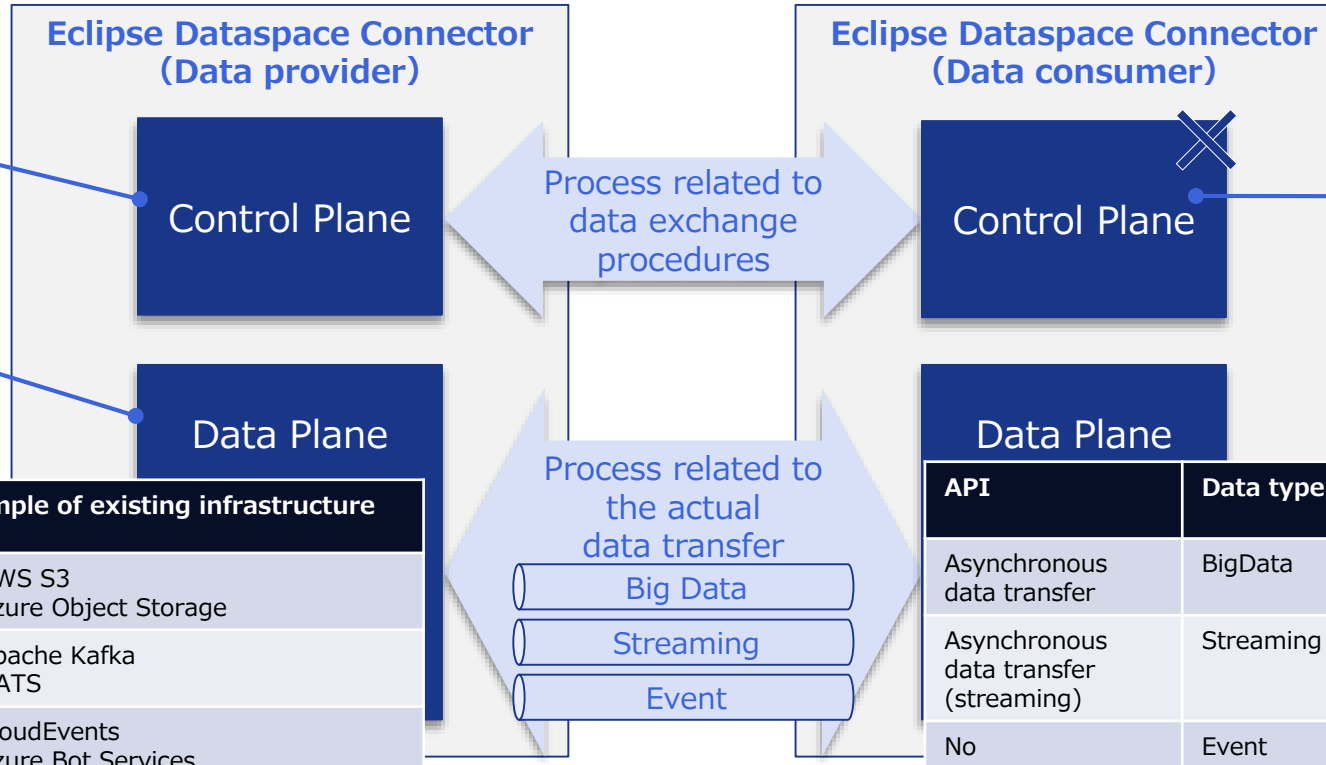
4. Eclipse Dataspace Connector

Data Plane Framework（DPF）

In EDC, a structure called the Data Plane Framework (DPF) divides the connector into two logical subsystems: the control plane and the data plane.
This mechanism seems to be aimed at leveraging existing infrastructure and increasing availability.
DPF is loaded into the connector as one of the extensions.
The following diagram was created by NTT DATA, based on the architecture diagram published at EDC Conference [*5].

The control plane is mainly responsible for processing related to data exchange procedures including policy execution monitoring such as contract negotiations and restrictions in data provision.

The data plane is responsible for processing related to actual data transfer including big data handling, streaming processing, event-driven etc.



For example, even if the Control Plane fails, the Data Plane continues processing related to actual data transfer and availability keep improving. Regarding the Control Plane, there are no part where it was clearly described in the source codes.

API	Data type	Example of existing infrastructure
Asynchronous data transfer	BigData	• AWS S3 • Azure Object Storage
Asynchronous data transfer (streaming)	Streaming	• Apache Kafka • NATS
No implementation of API at the time of survey	Event	• CloudEvents • Azure Bot Services

API	Data type	Example of existing infrastructure
Asynchronous data transfer	BigData	• AWS S3 • Azure Object Storage
Asynchronous data transfer (streaming)	Streaming	• Apache Kafka • NATS
No implementation of API at the time of survey	Event	• CloudEvents • Azure Bot Services

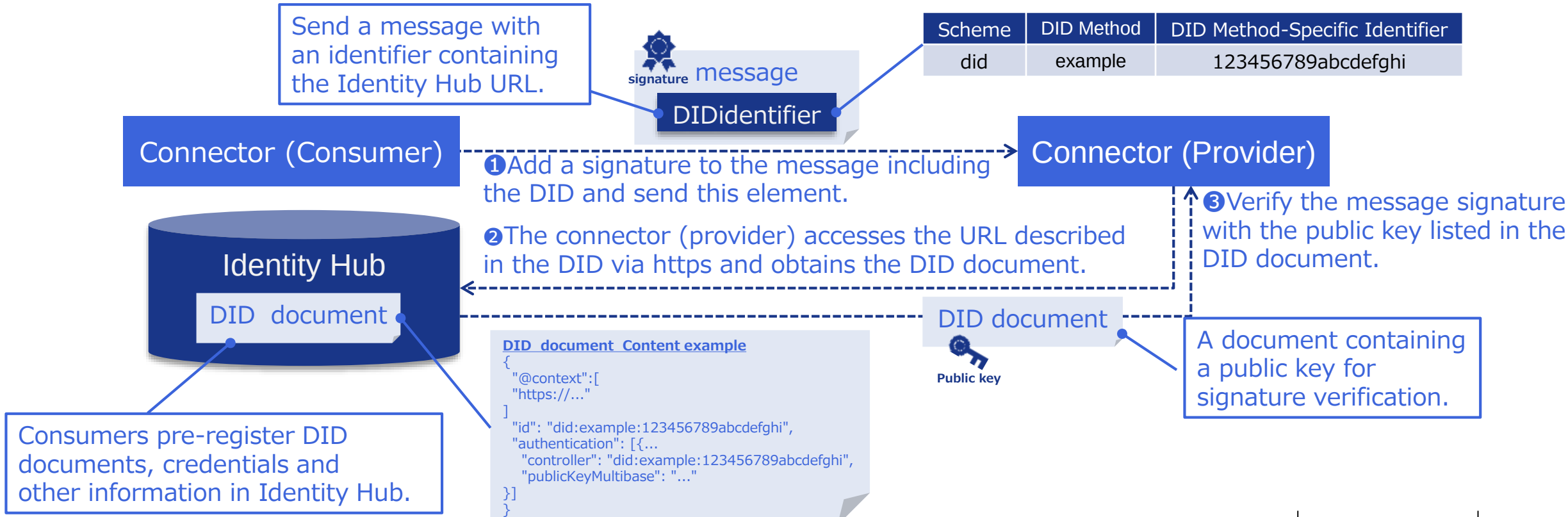
4. Eclipse Dataspace Connector

Decentralized Identity (DID)

Implementation of Decentralized Identity (DID), whose concept Gaia-X will adopt[*6], is underway at EDC.

DID uses a DID document (typically in JSON format) containing the DID identifier and public key to verify the signature of a message. the DID document is stored in a storage called Identity Hub. Currently, did:web is supported for this storage.

The following diagram was created by NTT DATA, based on the architecture diagram published at EDC Conference[*7], and shows an example of the procedure by which a consumer sends a request to a provider.



4. Eclipse Dataspace Connector

Application Interface

A set of APIs for accessing EDC functions is available as OpenAPI documents.

EDC's APIs are its own implementation, such as the distributed ID-related ones mentioned above, and separate from that of the IDS Dataspace Connector.

The table on the right is created by NTT DATA, based on EDC's OpenAPI documents[*8].

Among these, APIs that are considered to be necessary for data exchange are outlined below.

/contractnegotiations

Handle contract negotiations for obtaining data from data providers.

/contractagreements

Refer to the contractual agreements between data consumers and data providers.

/transferprocess

If agreement is reached as a result of contract negotiations, send a request to acquire the actual data.

/catalog

Access the catalog referenced by the connector and retrieve the metadata.

Classification	PATH	Overview
Connector	/check/health	API for connector health
	/check/liveness	
	/check/readiness	
	/check/startup	
	/instances	API for DataPlane selection
Data transfer	**/contractnegotiations**	**API for Contract negotiation**
	/contractagreements	
	/contractdefinitions	
	/transferprocess	**API for data transfer**
	/assets	API for assets
ID management	/identity-hub/collections	API for accessing IdentityHub
	/identity-hub/collections-commit	
	/identity-hub/query-commits	
	/identity-hub/query-objects	
Catalog	**/catalog**	**API for retrieving catalogs**
	/federatedcatalog	

4. Eclipse Dataspace Connector

Catalog

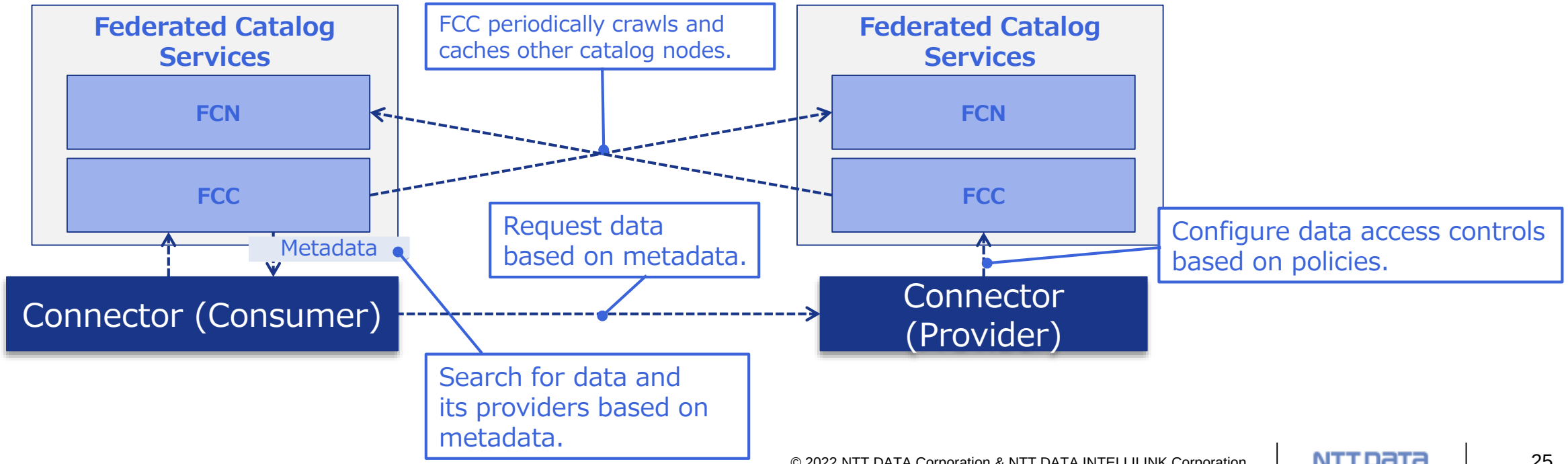
A data catalog facilitate access to data and the publication of owned data by storing metadata.

The Federated Catalog Service is being implemented in EDC to provide data catalog functions while ensuring data sovereignty, availability and scalability.

The Federated Catalog Service consists of the Federated Catalog Node (FCN) and Federated Catalog Crawler (FCC) functions.

The FCN handles catalog cache requests, and the FCC periodically crawls and caches catalogs in trusted relationships. This allows for a decentralized configuration of the data catalog[*5].

Based on the EDC conference materials, the figure created by NTT DATA is shown below.



4. Eclipse Dataspace Connector

Sources

1. https://github.com/eclipse-dataspaceconnector/Collateral/blob/main/Events/Conferences/2022-01%20EDC%20Conference/2022-01-31%20EDC%20-%20A%20peek%20into%20the%20future.pdf
   https://github.com/eclipse-dataspaceconnector/DataSpaceConnector/milestones
   https://github.com/eclipse-dataspaceconnector/DataSpaceConnector/milestones?state=closed
2. https://github.com/eclipse-dataspaceconnector/Collateral/tree/main/Events/Hackathons
3. https://github.com/eclipse-dataspaceconnector/Collateral/blob/main/Latest%20Presentations/2022-04-26%20Eclipse%20Dataspace%20Connector%20-%20Overview%20Deck.pdf
4. https://projects.eclipse.org/proposals/eclipse-dataspace-connector
5. https://github.com/eclipse-dataspaceconnector/Collateral/blob/main/Events/Conferences/2022-01%20EDC%20Conference/2022-01-31%20EDC%20-%20Architecture%20and%20Concepts.pdf
6. https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5
7. https://github.com/eclipse-dataspaceconnector/Collateral/blob/main/Events/Conferences/2022-01%20EDC%20Conference/2022-01-31%20EDC%20-%20DID.pdf
8. https://github.com/eclipse-dataspaceconnector/DataSpaceConnector/blob/main/resources/openapi/openapi.yaml

5

Discussions

Implications

EDC shares a common role with the IDS DataSpace Connector, a reference implementation based on IDS-RAM, but is a more scalable and independent implementation of the Gaia-X concepts.

Digital Sovereignty	GAIA-X idea[*1]	EDC support[*2]
The idea that parties have sovereignty over their own digital data. It includes digital processing, infrastructure, how data movement is organized and managed etc.	Facilitate customization of functionality, use of cloud-native services and avoid single-provider lock-in.	Leverage scalability, availability, and existing infrastructure with EDC architecture and DPF.
Decentralized design Decentralized control over their digital data. It includes interoperability of heterogeneous environments with many different technologies and operation models, improvement of resilience by utilizing distributed system architecture, and high availability.	Implement decentralized decision-making and regulation through a set of automatically enforceable rules on a distributed ledger.	Ensure availability and scalability in decentralized management of identities with DID and distributed catalogs with Federated Catalog Services.

Implications

EDC is expected to be developed based on the major policies of Digital Sovereignty and Decentralized Design. The core functions will continue to be actively developed with major changes. For example, as of June 2022, a number of improvement activities related to the Event Framework, a component which is responsible for event handling within the connector, have been proposed and are being implemented. Also, the IDS module is being cleaned up.

In terms of cooperation with other technologies, there is FIWARE[*3], an OSS product developed to promote data utilization. FIWARE for Data Spaces, which integrates FIWARE IAM (authentication function) and Dataspace Connector function, has been announced. In the future, it is expected that the interoperability of environments based on different technologies will become active.

The EDC project has test materials available for some scenarios[*4]. Also, for Catena-X, which is considering adopting EDC, construction materials[*5] for Docker Compose and construction materials[*6] for Kubernetes have been released. We have conducted verification using those materials and will also carry out improvement activities and contributions in collaboration with the development community.

5. Implications

Sources

1. https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5
2. https://github.com/eclipse-dataspaceconnector/Collateral/blob/main/Latest%20Presentations/2022-04-26%20Eclipse%20Dataspace%20Connector%20-%20Overview%20Deck.pdf
3. https://www.letsfiware.jp/fiware-for-data-spaces/
4. https://github.com/eclipse-dataspaceconnector/DataSpaceConnector/tree/main/samples
5. https://github.com/catenax/tractusx/blob/main/eclipsedataspaceconnector/api-wrapper/README.md
6. https://github.com/catenax-ng/catenax-at-home

NTT DATA

Trusted Global Innovator

NTT DATA

NTT DATA INTELLILINK Corporation