Dataspace connector in TEE and beyond: making a dataspace connector confidential 09:30 – 10:00

tech-x

Miki Kanno, Manager, NTT DATA Yuji Hagiwara, NTT DATA Koki Mitani, NTT DATA

#GaiaX #TechX24

This session's Presenters:

• Miki Kanno

- Manager / NTT DATA
- Leading global data spaces R & D activities.

Yuji Hagiwara

- Platform engineer / NTT DATA
- Research and development of technologies to promote data collaboration between enterprises

Koki Mitani

- Senior Research Engineer / NTT
- Leading open and collaborative innovation for building global infrastructure for data sharing between businesses











Today's presentation: Online and in-person Presentations and Discussion style!

1. Importance of the confidentiality of Dataspace connector

Presenter: Miki Kanno (manager from NTT DATA)

2.Deep dive: Analysis of threats and countermeasures in the use of EDC-based connectors Presenter: Yuji Hagiwara (platform engineer from NTT DATA)

3.Insights: Dataspace connector in TEE and beyond

Presenter(Online): Koki Mitani (Senior Research Engineer from NTT)

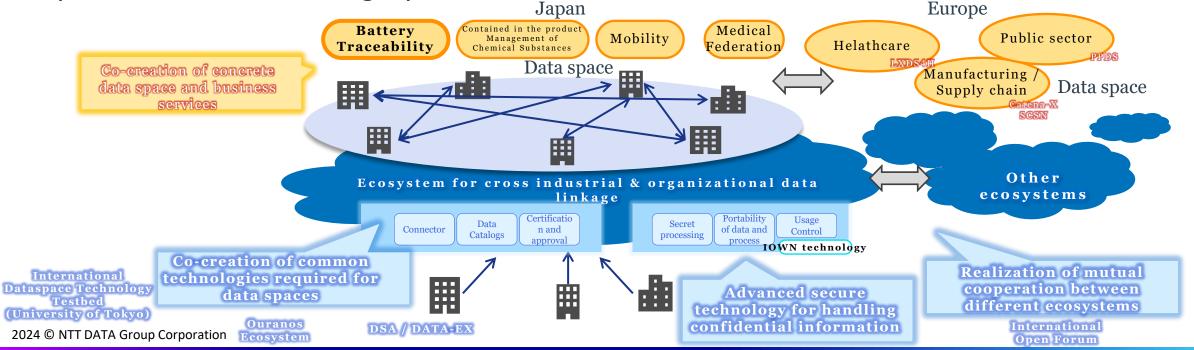
4. Discussion (With the audience in the room)

tech-x

NTT Group is working on the **development of international data-sharing eco-system** that interconnects data spaces.

We, NTT Group also **deploy connectors in the cloud for ease of use** and **provide assets** for ease of use by enterprises.

To exchange data, Dataspace connector (DC) is a software to connect to data spaces and is also a key to realize data sovereignty.

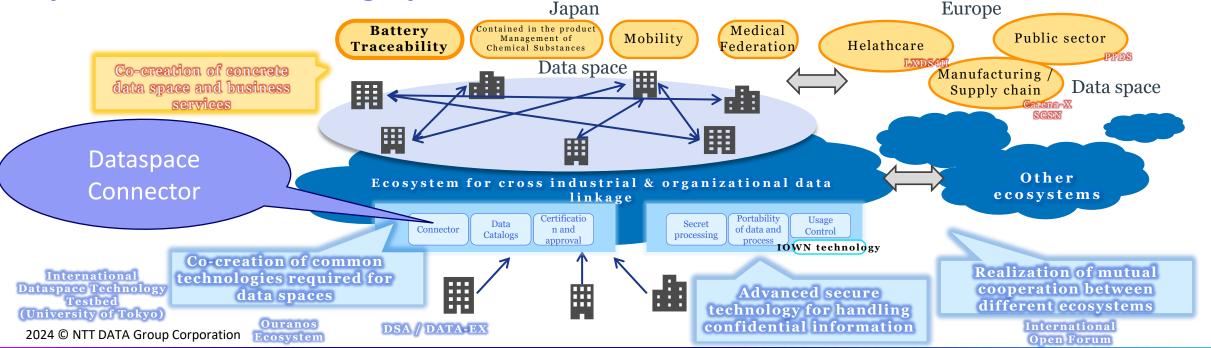


NTT Group is working on the development of international data-sharing eco-system that interconnects data spaces.

We, NTT Group also deploy connectors in the cloud for ease of use and provide assets for ease of use by enterprises.

tech-x

To exchange data, Dataspace connector (DC) is a software to connect to data spaces and is also key to realize data sovereignty.





In protecting data sovereignty, there are two broad patterns of protecting connectors while protecting the control authority of data sovereignty. (Logical, not physical connector placement)



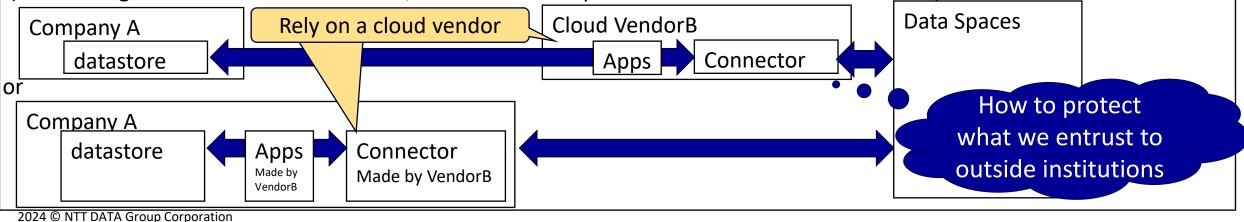


In protecting data sovereignty, there are two broad patterns of protecting connectors while protecting the control authority of data sovereignty. (Logical, not physical connector placement)



2. Case of not self-enforcement of data sovereignty control partly or completely

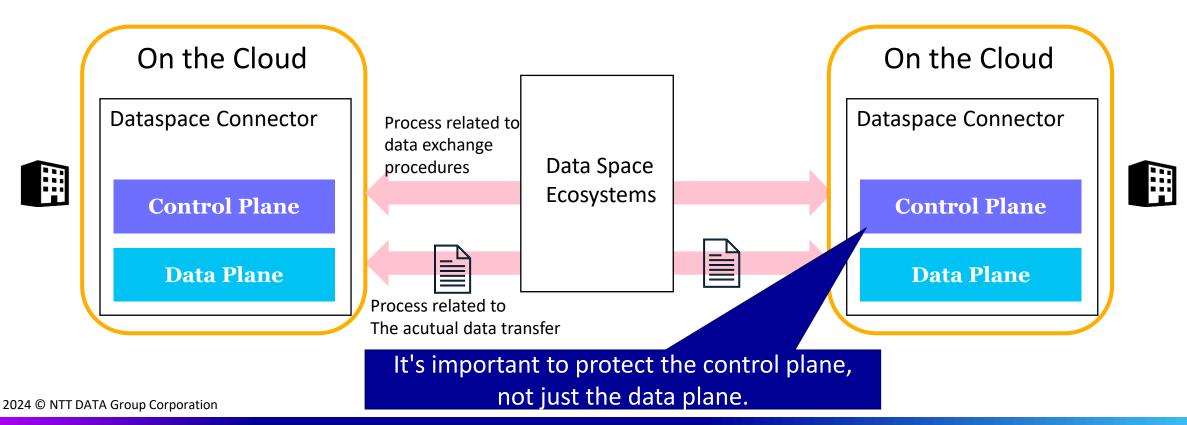
(ex. Running the Connector on the Cloud, Use vendor's easy-to-use connector connection tools)



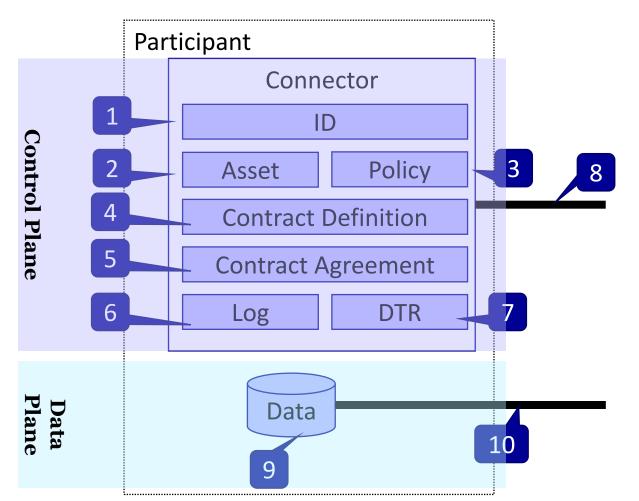
Cloud Computing Service's Possibilities



Data providers and consumers exchange data using Dataspace connector(DC). The cloud computing service is useful to realize flexibility and to make DC easy to use. The confidentiality of DC will be more and more important as the number of cases that multiple stakeholders are involved to exchange sensitive data increases.



Information which a connecter is handling with



These are our observations based on the implementation of Eclipse dataspace connector. We make no guarantee of any kind about the accuracy or completeness. They may be changed in future. 1: ID – The ID identifying who has the connector

2: Asset – A unit of sharing containing an endpoint of data

tech-x

3: Policy – A set of rules that define the terms of use for data or contracts

4: Contract Definition – A set of an access policy, a contract policy and an asset.

5: Contract Agreement – An agreement between two participants containing policy, derived from a contract definition.

6: Log – History of activities such as negotiation, agreeing, etc.

7: DTR(Digital Twin Registry) - Metadata

8: Control plane communication – Communications between connectors for negotiation, etc.

9: Data – Actual data to be shared.

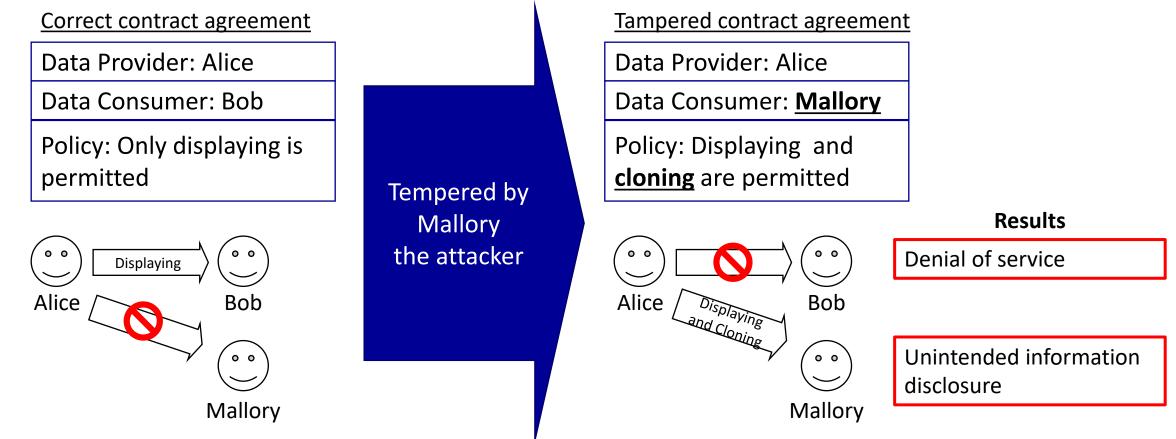
10: Data plane communication – Communications between participants for data transfer.

Ex) Tampering with a contract agreement



Not only data, but also other information in a connector MUST be kept secure.

Example:



Security risks we're considering



	Categories			Evaluation		
Information	Threats	Potential Adversary	Technique	Detectability	Result	
ID	Tampering	Core Service Provider A		Impossible		
Asset		Core Service Provider B			Damage to the trustiness of	
Policy	Repudiation	Onboarding Service Provider		Detectable by auditing	system	
Contract	Information	Advisory Provider]] []	
Definition	Disclosure	Enablement Service			Affect to Data	
Contract	Devial of	Provider		Detected		
Agreement	Denial of Service	Business Application		immediately	Affect to	
Log	JEIVICE	Provider			information other	
DTR		Data Provider			than Data	
		Data Consumer				
Control Plane Communication		Conformity Assessment				
Communication		Body				
Data		Connector Operator				
Data Plane		Insider				
Communication		Third party				

2024 © NTT DATA Group Corporation

Security risks we're considering



Categories			Evaluation		
Information	Threats	Potential Adversary	Technique	Detectability	Result
ID	Tampering	Core Service Provider A		Impossible	Domogo to the
Asset	Dereudietier	Core Service Provider B			Damage to the trustiness of
Policy	Repudiation	Onboarding Service Provider		Detectable by auditing	system
Contract	FromSTRIDE	Advisory Provider			
Definition From our	Disclosure	From datents X Roles		Detected	Affect to Data
Contract observation Agreement of EDC Log	Denial of Service	Provider definition Business Application Provider	:	Detected immediately	Affect to
implementati		Data Provider			information other than Data
DTR		Data Consumer			
Control Plane Communication		Conformity Assessment			
		Body	Research	ing real projects and	
Data		Connector Operator	extra	cted from them,	
Data Plane		Insider	but they m	hay not be complete	
Communication		Third party			

2024 © NTT DATA Group Corporation

Security risks we're considering



Categories			Evaluation		
Information	Threats	Potential Adversary	Technique	Detectability	Result
ID	Tampering	Core Service Provider A		Impossible	
Asset		Core Service Provider B			Damage to the trustiness of
Policy	Repudiation	Onboarding Service Provider		Detectable by auditing	system
Contract	Information	Advisory Provider			
Definition	Disclosure	Enablement Service	-	Detected	Affect to Data
Contract	Denial of	Provider			
Agreement	Denial of Service	Business Application		immediately	Affect to
Log		Provider			information other
DTR		Data Provider			than Data
		Data Consumer			
Control Plane Communication		Conformity Assessment			
		Body			
Data		Connector Operator			
Data Plane		Insider			
Communication		Third party			

2024 © NTT DATA Group Corporation

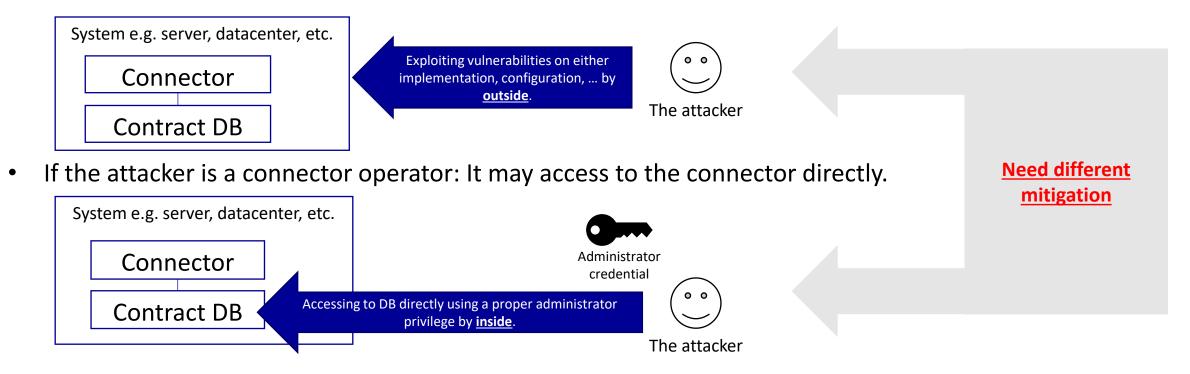
Importance of analyzing potential adversary



It's important to consider "Who'll be adversary?" to analyze potential techniques and mitigation.

Example: Tampering with a contract agreement by the attacker

• If the attacker is just a 3rd party: It may attempt to find and exploit vulnerabilities.



It is important to assess and deal with security risks properly.

Basic concepts of the impact evaluation criteria we're considering:

- Threats causing Data Leakage are more critical.
- Threats which are hard to detected are more critical.
- Threats exploiting the dataspace mechanism are more critical.

Contract is a key concept of data spaces and should be trusted by participants.

tech-x



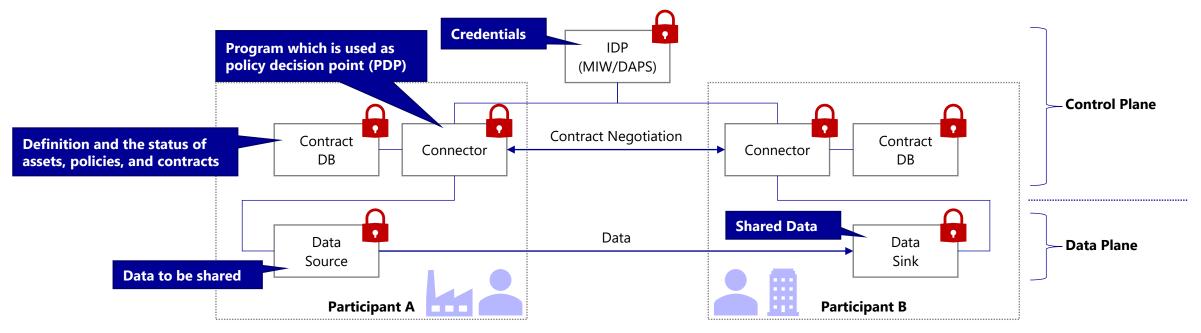
Risk assessment

Proposal: Protection by Trusted Execution Environment (TEE)

• From this slide, we talk about a simple use case for mitigating some risks related to key features of Connectors.

tech-x

- For preventing data from being leaked when the contract information in the Connector is tampered with, we need to consider protecting not only the data plane, but also the control plane.
- We are considering a concept to protect Connectors by using Trusted Execution Environment (TEE).
- TEE can be used for setting up hardware-supported isolated runtime environments (SEV-SNP, SGX, TDX etc.).
- TEE offers remote attestation with proof of the integrity of the initial software components, to prove the correct setup.



Scenario 1: Protect Your Connector by TEE

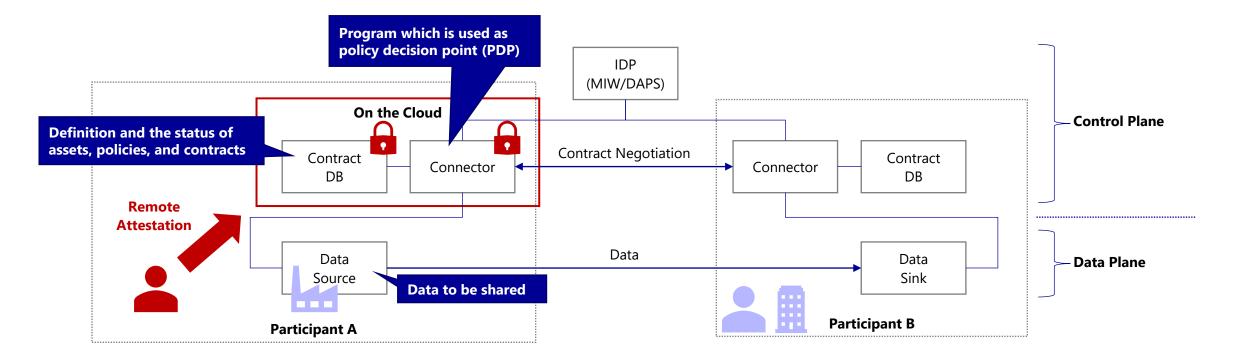
When using a third-party Connector on the cloud, the user may need additional protection for preventing data from being provided illegally.

111

11111

tech-x

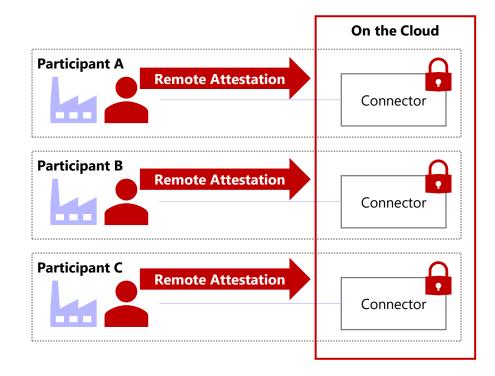
• By using the remote attestation feature of TEE, the user can verify whether the control plane configuration of the third-party Connector on the cloud is correct.



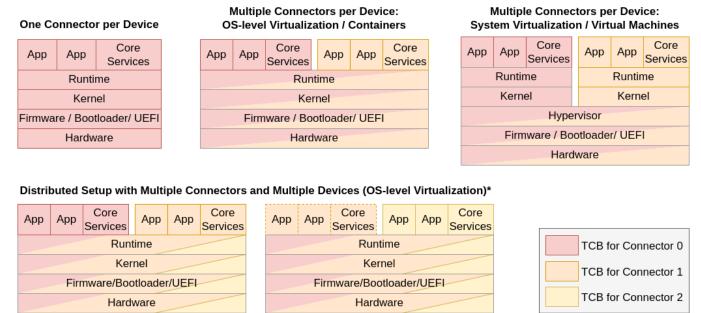
•

Scenario 1 - Additional Consideration: Multi-Tenancy Scenario

• For use cases where Connectors used by multiple users may be operated on the same cloud infrastructure, remote attestation for Connectors in multi-tenant configurations need to be considered.



(Reference) Related discussion in the IDS RAM4.



tech-x

*assuming each connector service may be moved to each device

https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/blob/main/documentation/4_Perspectives_of_the_Reference_Architecture_Model/4_1_Security_Pers pective/4_1_3_Securing_the_Platform.md

Copyright 2024 NTT CORPORATION

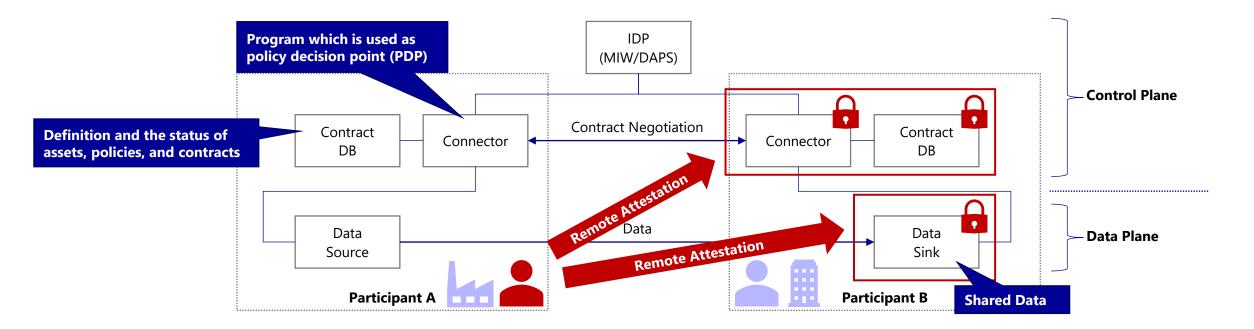
Scenario 2: Protect Your Shared Data in Use by TEE

• When the user shared data to the other participant, the user may need enforcement of data usage control to prevent the data from being handled improperly.

1 1 1

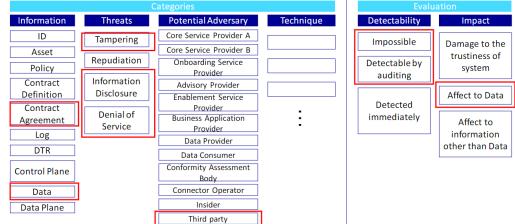
tech-x

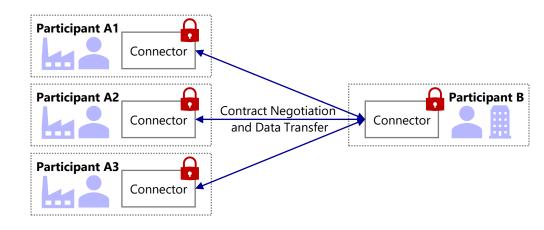
• By using the remote attestation feature of TEE, the user can verify whether the configuration of the control plane and the data sink of the other party's Connector are correct.



Key Findings

- In multi-stakeholder case, remote attestation by TEE can be used to mitigate some risks.
- Depending on system structure and operational structure, there are several variations of potential adversaries.
- Service providers need to consider to provide an option to apply TEE for adding value to their services.
- Due to the cost of TEE, we assume that TEE will be applied to some use cases such as financial use cases, and the use cases which use extremely sensitive data.
- In Scenario2, international standardization of remote attestation between stakeholders and Connectors may needed.
- TEE related scenario can also be applicable to use cases where multiple data providers and application providers participate.







Let's Have an Open Discussion about Use Cases of TEE

We would like to continue open discussion about the following topics.

- Are you working on use cases that involve sensitive data?
- Are you working on use cases that utilizes data from multiple companies?
- Are you already providing Connector services? Are you aware of the need for TEE?
- Do you think we need some standardization of remote attestation regarding the use of Connectors with TEEs?

tech-x

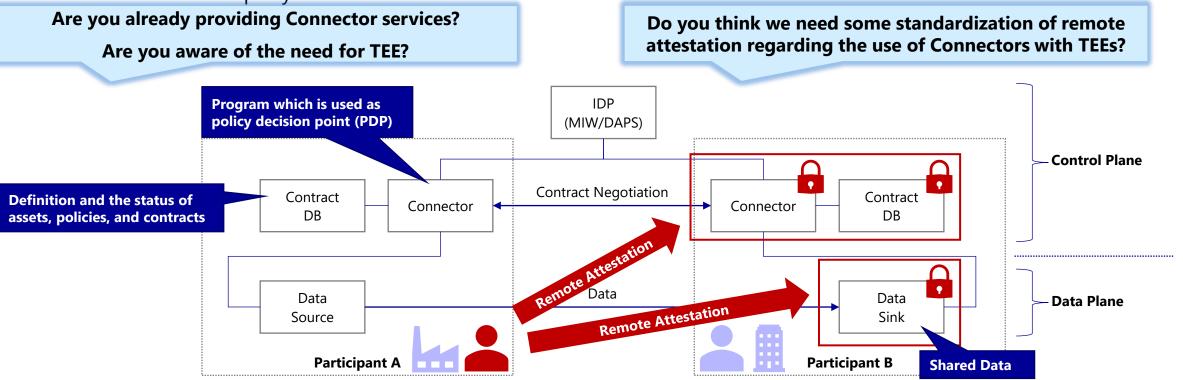
• Do you see any potential in the future of Connector-TEE integration?

Scenario 2: Protect Your Shared Data in Use by TEE

• When the user shared data to the other participant, the user may need enforcement of data usage control to prevent the data from being handled improperly.

tech-x

• By using the remote attestation feature of TEE, the user can verify whether the configuration of the control plane and the data sink of the other party's Connector are correct.



Thank you!

Miki Kanno, Yuji Hagiwara, Koki Mitani

#GaiaX #TechX24

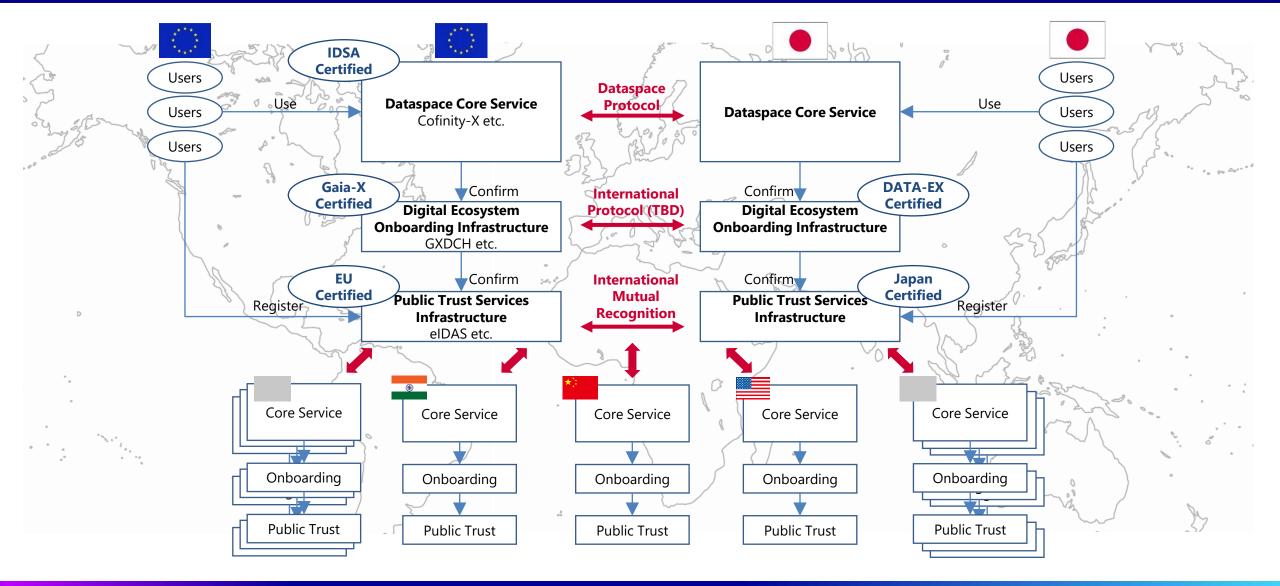
<u>Miki Kanno <Miki.Kanno@nttdata.com></u> <u>Yuji Hagiwara <Yuji.Hagiwara@nttdata.com></u> <u>Koki Mitani <koki.mitani@ntt.com></u> <u>Masaru Dobashi <Masaru.Dobashi@nttdata.com></u>

Names of products, companies and services in this document are trademarks or registered trademarks of their respective companies.

tech-x

We Need an International Interoperability Framework





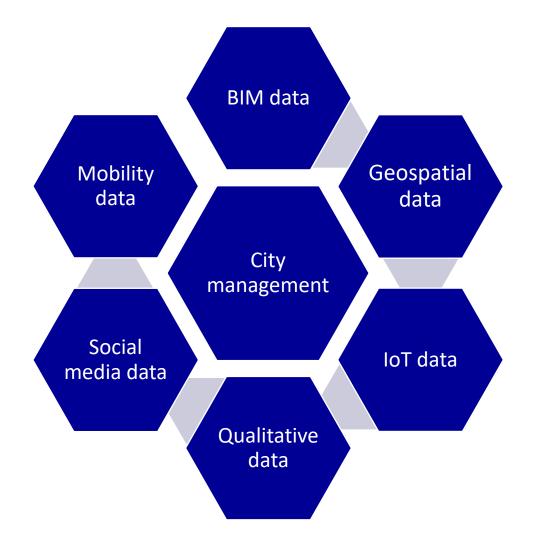
Achieving a higher degree of interoperability between data space APIs and components which use NGSI-LD, OGC, and IFC models or services 10:00 – 10:30

tech-x

Frédéric Lé, Youragileway Michael Mulquin, Open & Agile Smart Cities (OASC) Jean Brangé, AFNeT Services

#GaiaX #TechX24

A lot of data is gathered in a city or community



- All these different types of data are important to support city management
- Each is structured to enable specific types of insight
- Aligning them is important and difficult

Three families of standards

NGSI-LD Developed by Fiware and ETSI to bring context to sensor data

BIM and IFC Developed by BuildingSmart International to provide digital descriptions of buildings and urban infrastructure Geospatial standards developed by the Open Geospatial Consortium (OGC) to enable precise descriptions of locations and movements

All of them aim to handle the same set of issues

- Data about locations and movements
- Data about urban infrastructure buildings, roads, bridges
- How to link different data sources together to provide insight

However, because of their different focus, they each have their own strengths and weaknesses

High-level comparisons

- Fiware/NGSI-LD is particularly good at enabling IoT data to be linked with valuable context data to show its significance. It can handle geospatial and building data but only to a certain level
- OGC standards allow geo-spatial data to be handled to a high degree of sophistication, but can only provide a certain degree of context and building related information
- BIM/IFC standards provide a rich and detailed way of describing buildings and urban infrastructure, but struggle to indicate precise geographic location and wider context

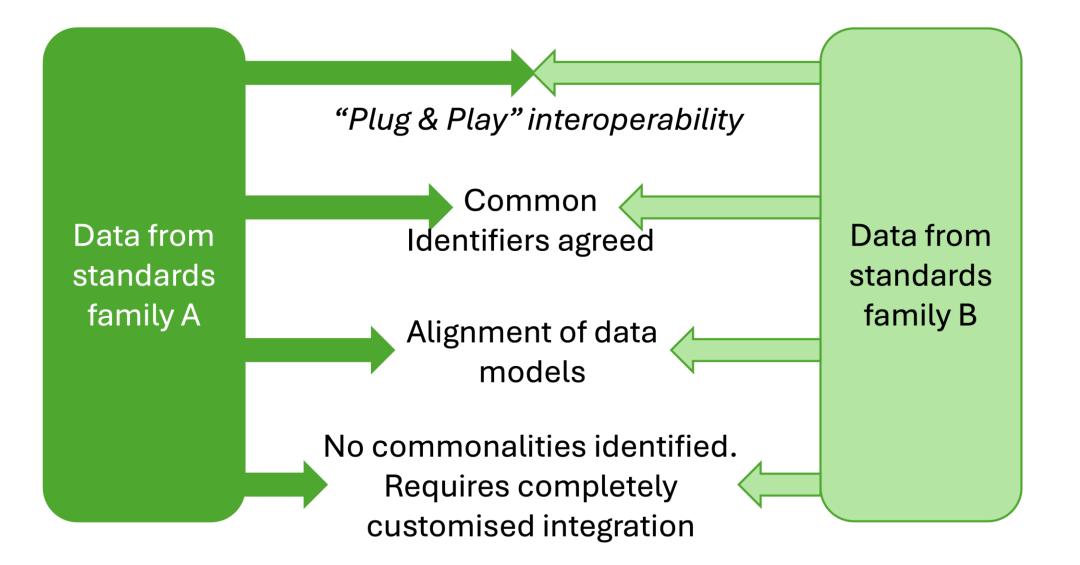
Minimal Interoperability

- To tackle key urban issues, cities need to be able access detailed sets of specific information about location, urban infrastructure and context
- They don't need all possible information just the minimal but "good enough" for their purposes,
- Cities need "work arounds" to help them gather the information they need from data collected using the different standards
- This is the focus of the OASC/Living-in.eu Minimal Interoperability Mechanisms

The MIMs being developed by OASC and Living-in.eu

MIM	Function
MIM1: Context	Data sets/streams can be linked according to context
MIM2: Data Models	All data sets/streams use consistent data models
MIM3: Contracts	Appropriate data sets/streams can be found, and agreement can easily be reached for their appropriate use
MIM4: Trust	Citizens can take charge of how data about them is used so that it can benefit themselves and their community
MIM5: Transparency	Decision making algorithms will use data appropriately to make fair and transparent decisions
MIM6: Security	Data can be held and shared securely
MIM7: Places	Geo-temporal information can be accurately described in consistent ways
MIM8: Indicators	KPIs can rely on consistent data from across the ecosystem to enable reliable measurement of progress
MIM9: Analytics	Models and analytics used within the ecosystem can work well with other models and analytics
MIM10: Resources	Information about city related resources can be appropriately shared

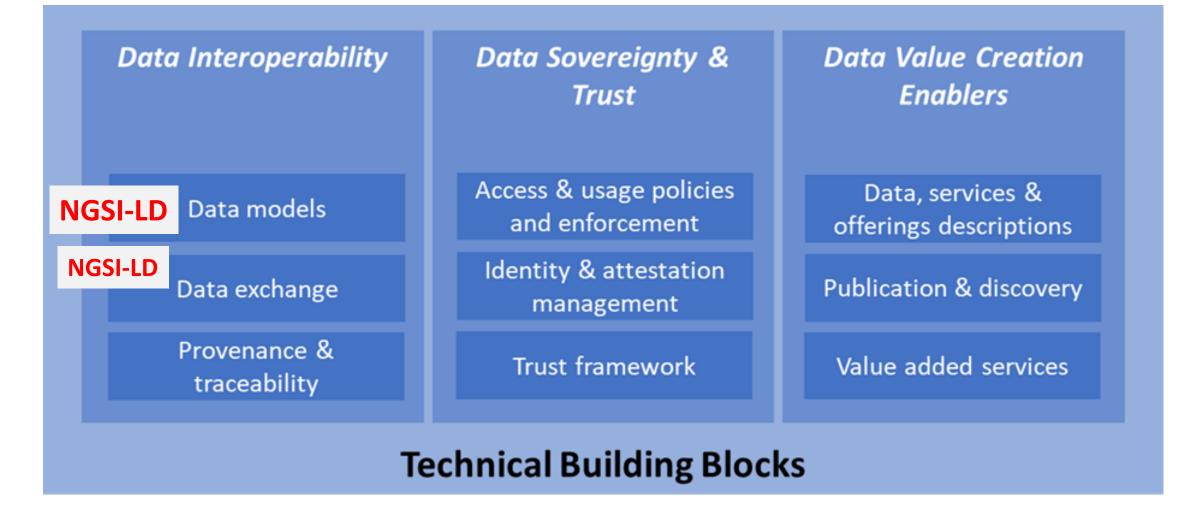
Examples of how-to bring alignment



OASC and AFNeT are working with ETSI

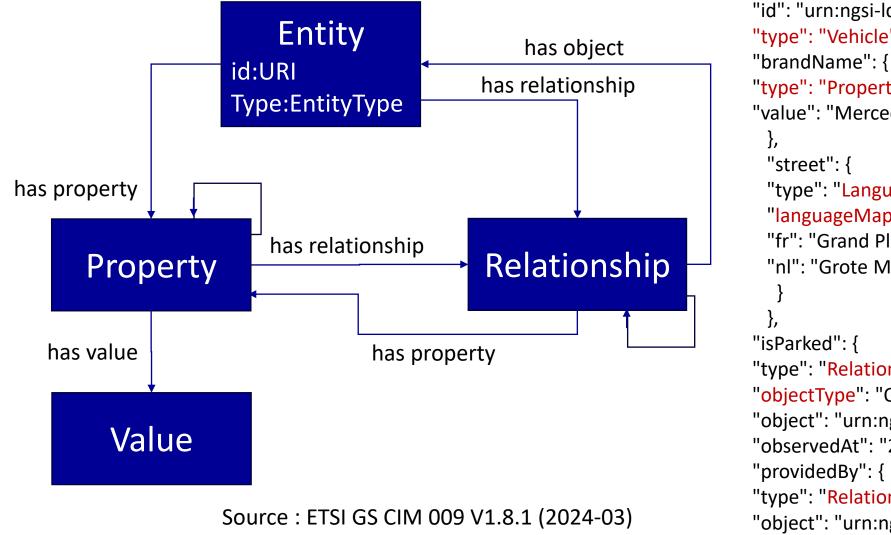
- To report on how smart communities are using OGC WFS and OGC API and standards-based encoding such as GeoJSON, GML, GeoPackage, CityGML and IFC, along with the requirements of the INSPIRE directive
- To specify how to make geodata accessible as Linked Data, how to share spatial (and spatio-temporal) data, and how to make them interoperable with, within, and between systems and territories
- To specify how to both establish and maintain the number of connections between NGSI-LD entities and their geographical 2D/3D representations

NGSI-LD positioning vis-à-vis Blueprint 1.0



DSSC Blueprint 1.0 - Technical Building Blocks - Blueprint v1.0 - Data Spaces Support Centre (dssc.eu)

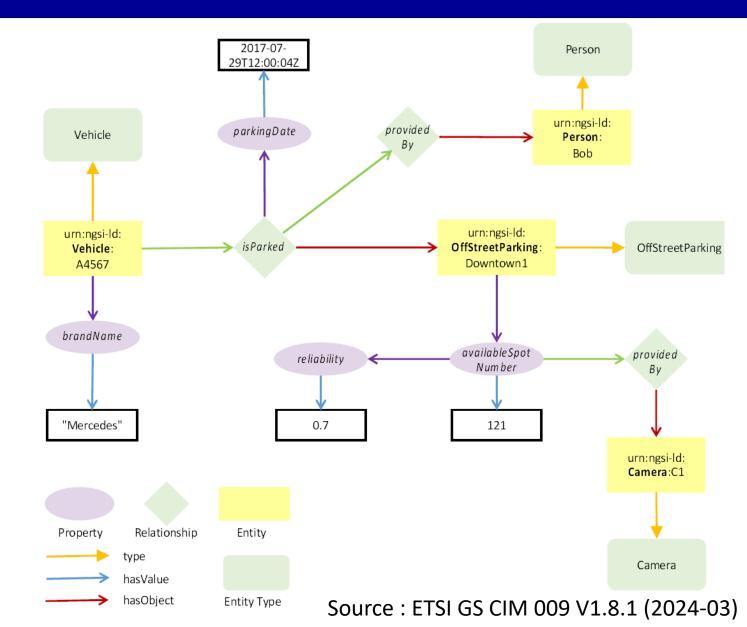
NGSI-LD: Meta Model



"id": "urn:ngsi-ld:Vehicle:A4567", "type": "Vehicle", "type": "Property", "value": "Mercedes" "type": "LanguageProperty", "languageMap": { "fr": "Grand Place", "nl": "Grote Markt "type": "Relationship", "objectType": "OffStreetParking", "object": "urn:ngsi-ld:OffStreetParking:Downtown1", "observedAt": "2017-07-29T12:00:04Z", "type": "Relationship", "object": "urn:ngsi-ld:Person:Bob"

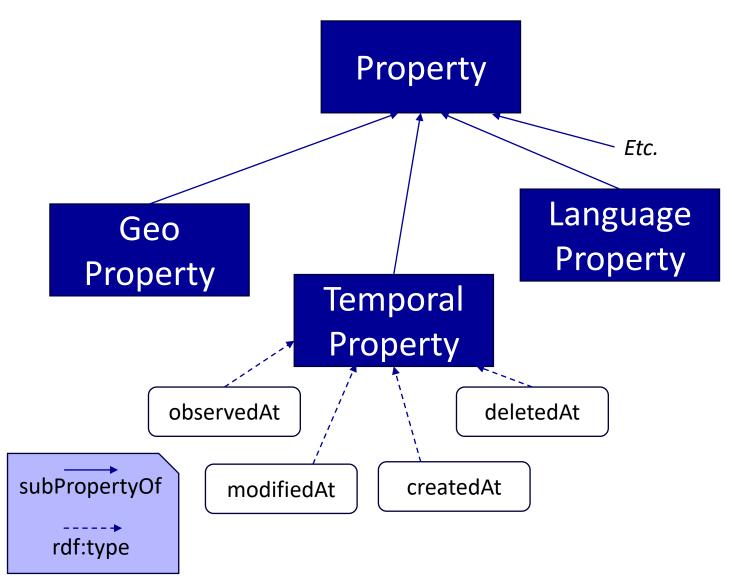
• • •

NGSI-LD: Domain Model Example



- Specific to an application domain (e.g. Smart City, Smart AgriFood, etc.)
- In scope of the Smart Data Model Program
 - ✓ Program led by FIWARE, IUDX, TM Forum, OASC and others
- Out of scope of the NGSI-LD standard

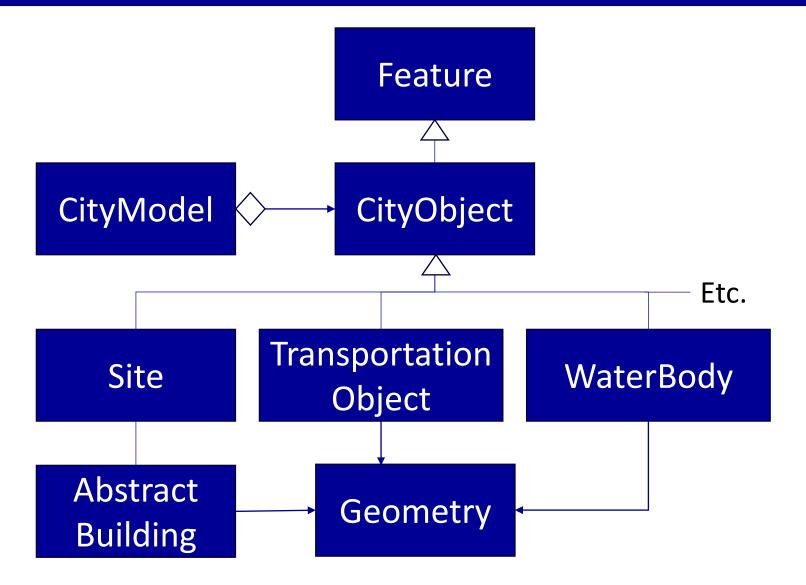
NGSI-LD: Cross-Domain Model Ontology



 Sub properties aimed at avoiding conflicting or redundant definitions in each of domain-specific ontologies

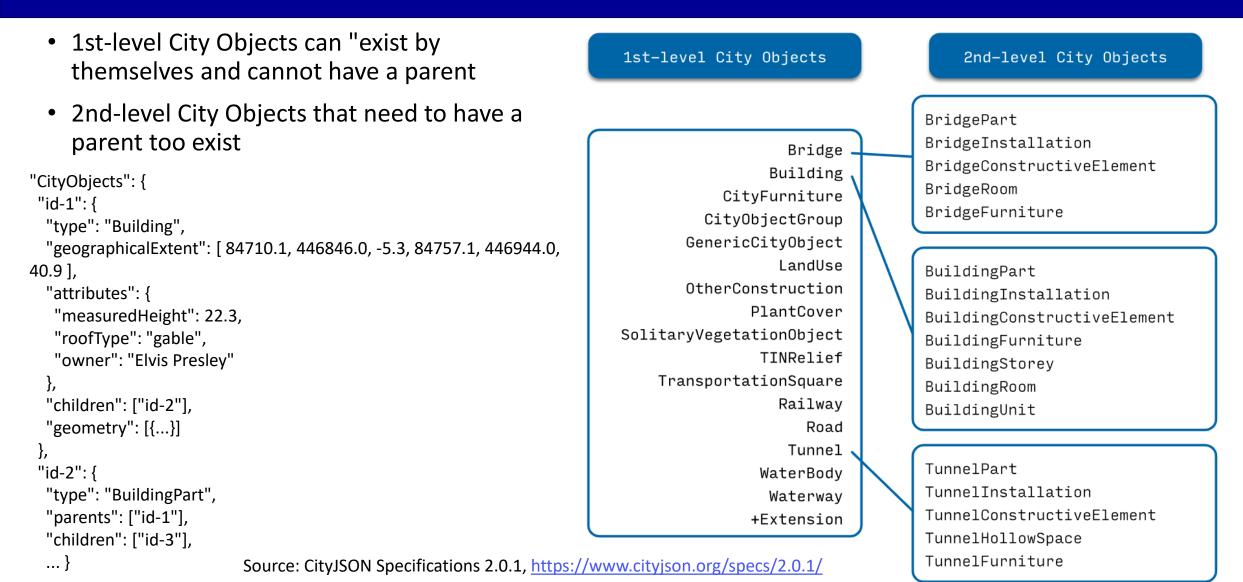
- Temporal properties to model state changes
 - ✓ Specify Domain
 Events using Linked
 Data Subscriptions
 - E.g. air pollution reaches a certain level

CityGML top-level class hierarchy

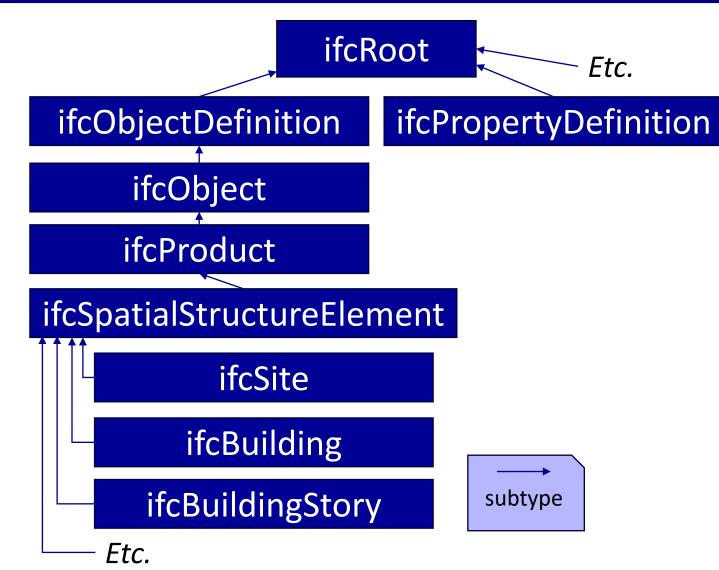


- CityGML uses sub-types and multiple inheritance
- CityGML classes could be modeled by as an NGSI-LD Domain Model
 - The NGSI-LD Meta Model does not support subtyping of Entity Types
- CityGML defines five consecutive levels of detail (LoD)
 - Each object may have attached a separate representation for each LoD simultaneously

CityJSON covers a subset of CityGML's scope



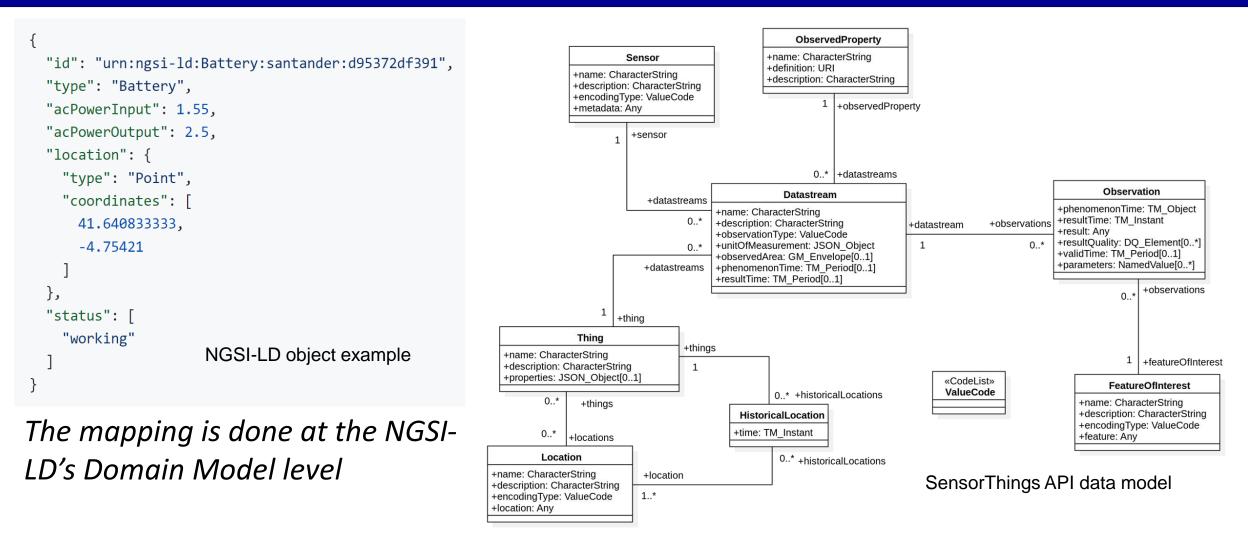
IFC top-level class hierarchy



• The ISO/TR 23262:2021 report lists many GIS/BIM incompatibilities

- ✓ Conceptual differences in underlying software design approach
- ✓ Technological] Differences in underlying architectures
- ✓ Generation of watertight (*fit as to be impermeable to water*)
 representations for BIM
- \checkmark Diversity in spatial representation
- IFC classes can be modeled at the Domain Model level.
 - ✓ How to represent sub-typing remains an open question

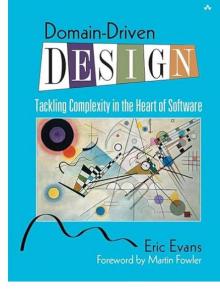
Civitas Connect example: PoC aiming to convert a NGSI-LD data structure into the SensorThings API data



Source: <u>https://github.com/Civitas-Connect/frost-ngsi-poc</u> and <u>https://ogc-iot.github.io/ogc-iot-api/datamodel.html</u>

Domain Driven Design (DDD) and Data Mesh

- DDD
 - "Multiple models are in play on any large project. Yet when code based on distinct models is combined, software becomes buggy, unreliable, and difficult to understand. Communication among team members becomes confused. It is often unclear in what context a model should not be applied...
 - Explicitly define the context within which a model applies. Explicitly set boundaries in terms of team organization...
 - A BOUNDED CONTEXT delimits the applicability of a particular model..."
- Data Mesh
 - "Data mesh, at its core, is founded in decentralization and distribution of data responsibility to people who are closest to the data...
 - Data mesh gives the data sharing responsibility to each of the business domains. Each domain becomes responsible for the data it is most familiar with...
 - DDD's Strategic Design embraces modeling based on *multiple models* each contextualized to a particular domain, called a bounded context"





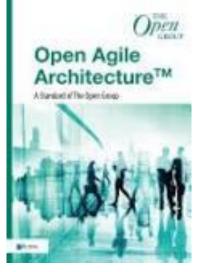


Domain Driven Design: integration patterns

Context Map Patterns invented by Eric Evans

- **Published Language** uses a well-documented and shared language that can express the necessary domain information as a common medium of communication, translating as required
- **Conformist** eliminates the complexity of translation between bounded contexts by slavishly adhering to the model of the upstream team
- Anti-Corruption Layer creates an isolating layer to provide clients with functionality in terms of their own domain model; the layer talks to the other system through its existing interface, requiring little or no modification to the other system
- **Open Host Service** defines a protocol that gives access to your subsystem as a set of services
- Event Publisher communicates with other bounded contexts through domain events that can be consumed by other bounded contexts
- **Customer/Supplier** establishes a clear customer/supplier relationship between the two teams
- Shared Kernel designates some subset of the domain model that the two teams agree to share

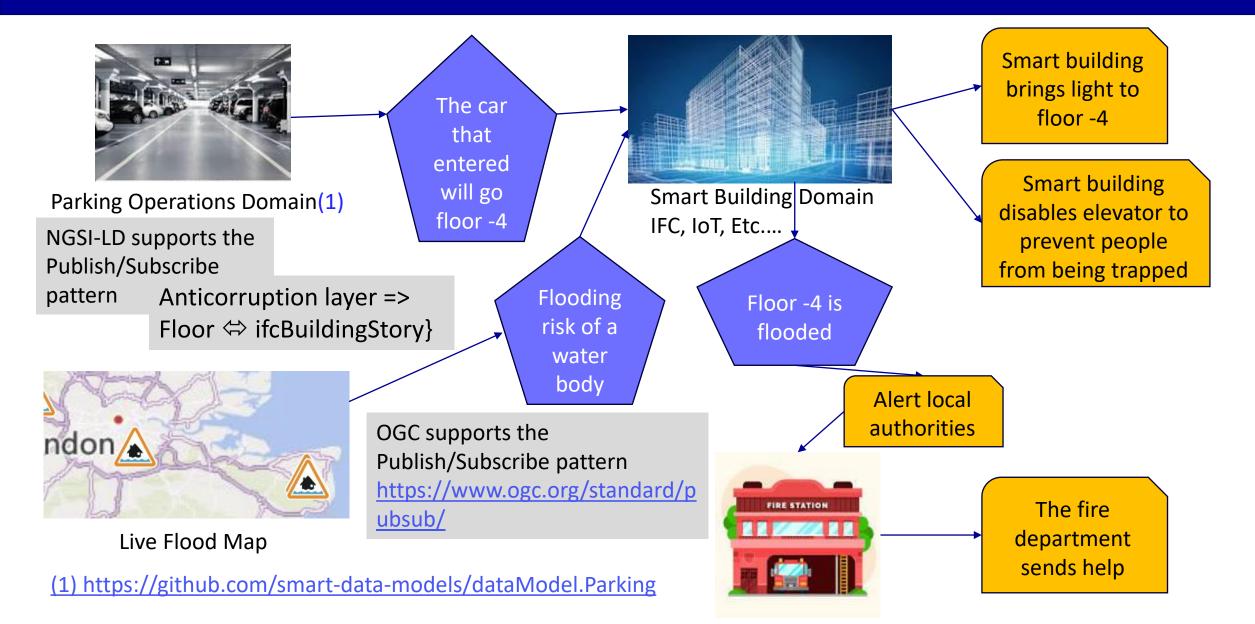
Source: https://digital-portfolio.opengroup.org/oaa-standard/latest/part2-building-blocks/DDD-strategic-patterns.html





Communication Commitment / Capability of Teams

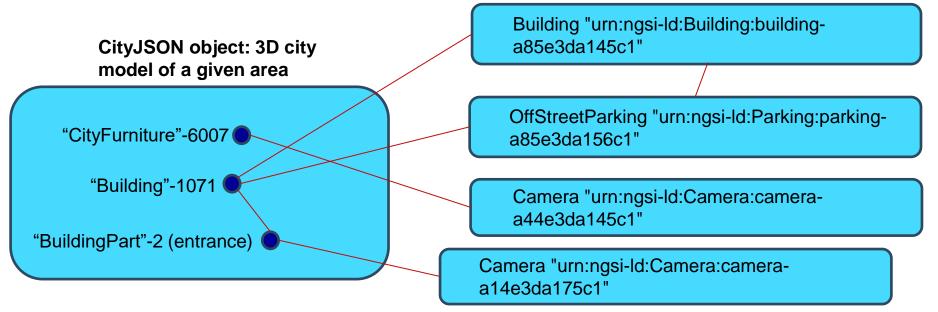
Domain Events and Anti-Corruption Layers



Linking objects that belong to different contexts

- NGSI-LD
 - All Entities are identified by URIs
 - If those URIs are expected to participate in external linked data relationships, they should be dereferenceable
- Relationships that cross smart data models' boundaries?
 - Since a real-world object can be represented in more than one model
 - How should they be identified? What about IoT Objects? Linked Data and can be dereferenced.

Smart Data Models



Coordinate reference systems and Locations

The <u>OGC API - Features - Part 1: Core</u> standard defines support for only two coordinate reference systems:

- WGS 84 longitude, latitude
- WGS 84 longitude, latitude, ellipsoidal height

OGC allows the usage of other Coordinate Reference Systems (CRS) including the necessary mathematical transformations see : OpenGIS Coordinate Transformation

https://www.ogc.org/standard/ct/

There are good reasons why WGS84 is not a good candidate for a worldwide default CRS. It mostly has to do with plate tectonics. WGS84 is fixed to the North American plate, which means that while WGS84 serves well for locations on that plate, locations on other plates that move with respect to the North American plates suffer displacement that gradually increases with time. For example, North America and Europe move apart at a rate of about 2.5 cm/year.

Because WGS84 is unsuitable in Europe, European guidelines (like INSPIRE) recommend using ETRS89. Also, GNSSs other than GPS do not use WGS84.

https://www.w3.org/2015/spatial/wiki/Coordinate_Reference_Systems

We are investigating what are the preferred CRS used in the Smart Cities and European GIS tools in order to make a recommendation.

link to IFC spec on CRS

https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD1/HTML/schema/ifcrepresentationresource/lexical/ifccoordinatereferencesystem.htm

LOD and LOIN

- OGC LOD are defined in CityGML with 4 levels related to the geometry visualisation refinement
- LOIN : Level Of Information Need is available as prEN 7817 or ISO/FDIS 7817-1
 - Provides a wider and more detailed definition of the "details" in regards of geometry representation but also alpha-numerical representation and other type of documentations.
- The LOIN is pushed forward by the European BIM community, mostly for collaboration process during design and construction
 - This could be used to specify the level of detail of some BIM / IFC data to retrieve from a dataset in addition to the CityGML visualisation levels. This also relates to the discussion on Coordinates and location.
- We are checking if a mapping has already been done between the LOIN "detail" and the CityGML LOD

Thank you!

Frédéric Lé Président Youragileway, expert AFNeT, <u>fle@youragileway.com</u>

tech-x

MIMs Ambassador with OASC, <u>michael@oascities.org</u>

Jean Brangé Président President of AFNeT Services, jean.brange@afnet-services.fr #GaiaX #TechX24 Evolving the Gaia-X transport protocol to different connectors 11:00 – 11:45 tech-x

Joaquin Salvachua, Universidad Politécnica de Madrid Andres Munoz-Arcentales, Universidad Politécnica de Madrid Carlos Aparicio, Universidad Politécnica de Madrid

#GaiaX #TechX24







UNIVERSIDAD POLITÉCNICA DE MADRID

EUNOMIA Euvopia

Evolving the Gaia-X transport protocol to different connectors - Covering data governance aspects



Financiado por la Unión Europea NextGenerationEU



MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA



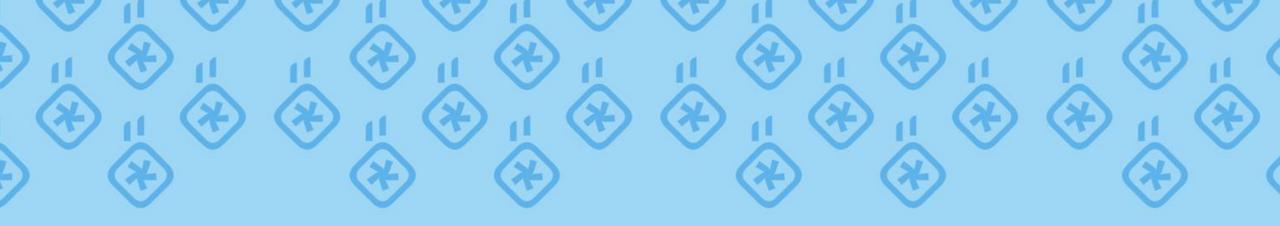
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN

E INTELIGENCIA ARTIFICIAI

Plan de Recuperación, Transformación y Resiliencia



INSTITUTO NACIONAL DE CIBERSEGURIDAD



Hi everyone!













Joaquín Salvachúa

Joaquín Salvachúa, UPM professor that has been involved into formal method for specification and verification of protocols. Multimedia and real time protocols (coauthor of an RFC). Teaching over Cloud infrastructure, Big data infrastructure and Blockchain and DLT technologies.



Andrés Muñoz-Arcentales

Assistant Professor at UPM and a Senior Researcher in the Next Generation Internet Research Group (GING/UPM) with main research interests in the fields of Smart Spaces, Data Fusion, Data Spaces, Machine Learning, Digital Twins, Cloud and Edge Computing and Big Data infrastructure.



Carlos Aparicio de Santiago

Researcher at GING-UPM and Ph.D. candidate in Telecommunication Engineering at UPM, He researches in fields of big data architectures, data access and usage control, SSI, data spaces and machine learning.











GING - UPM

We all belong to here: <u>https://ging.github.io/</u>

Research group GING at the Polytechnic University of Madrid.

Our research focuses mainly on protocols and WWW standards and technologies applied to numerous use cases. Currently, we are **focused on research in cloud computing**, education, learning analytics, **data engineering**, distributed videoconferencing systems with WWW standards, **LLMs and AI**, **open data**, **and data spaces**.

Involved into Protocol formal methods for specification, validation and verification for protocols using process algebras (LOTOS) some years ago. Participation into **several standardization committee**. Participation into IETF, W3C, ETSI and other standardization bodies.







Financiado por la Unión Europea NextGenerationEU



Presentations

What is Eunomia?

Eunomia (Euvoµia ' good law ') was the goddess of laws and legislation.

It was associated with the internal stability of a state, including the enactment of good laws and the maintenance of civil order. She was also the spring goddess of green pastures (nomia in Greek). Eunomia was one of the Horai (Horae), goddesses of the seasons and guardians of the gates of heaven. Her sisters were the goddesses Dike (Justice) and Eirene (Peace). Its opposite was Dysnomia (Anarchy).

She was considered one of the Horae, daughter of Zeus and Themis. In Roman mythology he is called Discipline.













Our Mission

We work on **various levels**, attempting to **fill some gaps** within data space architectures and **predict some future lines**, some of them like the following:

- Evolution of transport protocols
- Covering data governance and **distributed data governance requirements**
- Addressing trust anchor systems beyond European borders
- Use of self-sovereign identity, SSI
- Application of **ODRL**, Zero Trust, and **ReBAC** for policy management
- Metadata for data spaces
- Use of **DataLakehouse** architectures applied to data spaces
- High-speed transmission systems
- Integrations with **IDSA** connectors
- Integrations with **FIWARE** connectors

Project supported by INCIBE

EUNOMIA-Soluciones para la soberanía, confianza y seguridad en los espacios de datos

C.128.23 EUNOMIA, C130.23 MCIPYME











Contents

What are we going to speak about today:

We will mainly define some research lines we're follow, define requirements, draw some conclusions, possible drawbacks and future plots.

- Integration of **diverse transport protocols**
- Integration for **different trust anchors**
- Mapping **ODRL to ZeroTrust and ReBAC**



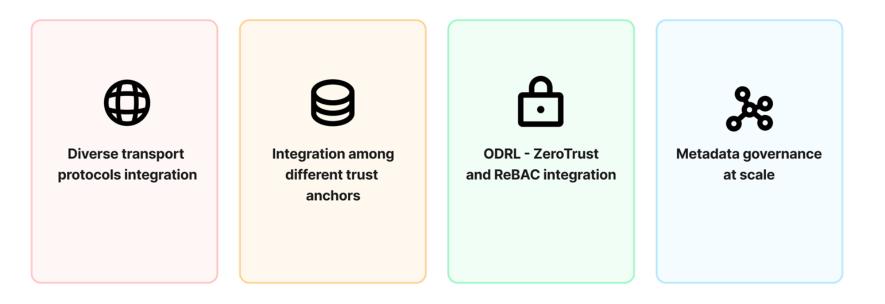








Contents





Diverse transport protocols integration



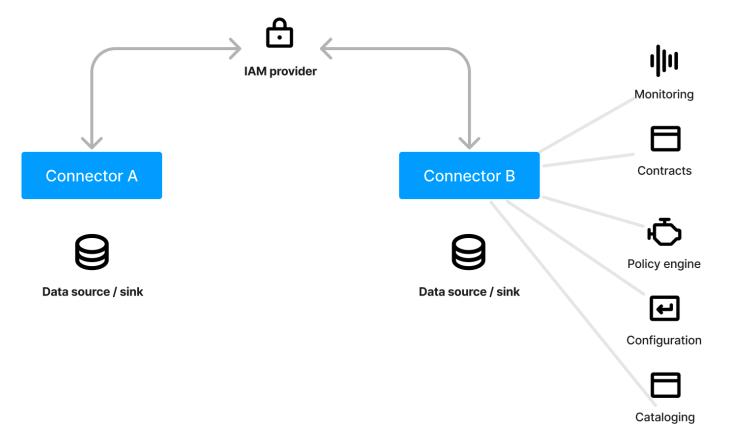








DataSpace protocol 2024-1



Initial version available.

This diagram is quite simplified. But shows us **some of the requirements** we want to address.



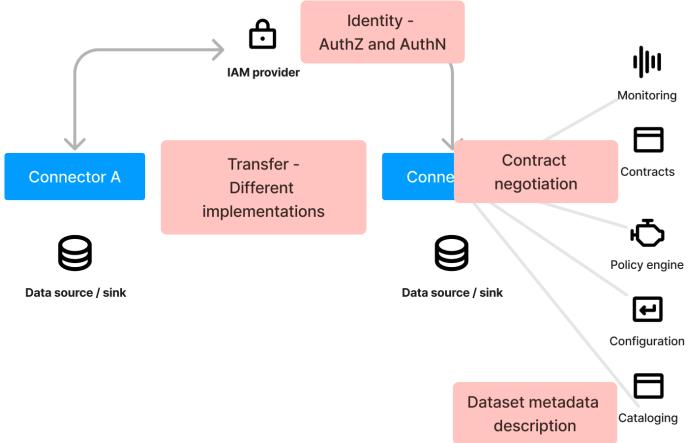








DataSpace protocol 2024-1











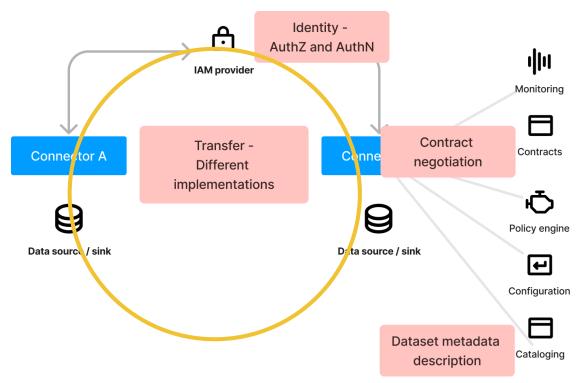


Requirements for the different connectors

Different phases needed :

- Dataset metadata description (DCAT) publication.
- Contract negotiation.
- Identity (authentication and authorization) required.
- Transfer connection
 - Different implementation details on each connector.

We will focus on the **transfer metainformation** for the control plane.















Different transfer scenarios

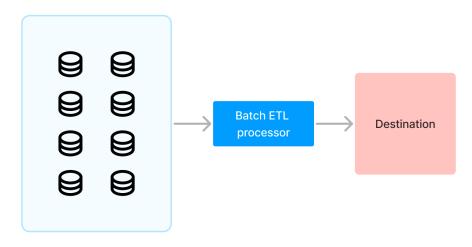
Batch data :

- All the data is available.
- May apply data access control and data usage requirements

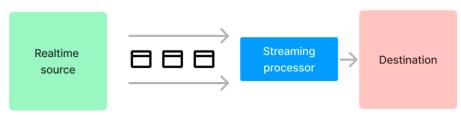
Streaming data :

• Data is produced in real time.

Batch processing



Streaming processing











\$Incibe

Transport protocol

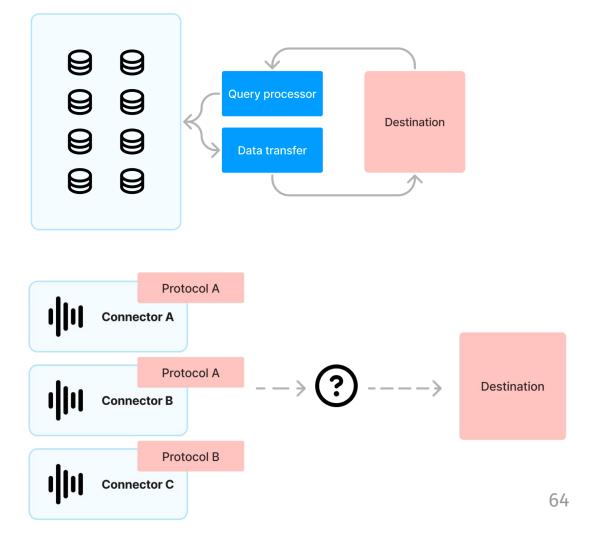
Different transfer scenarios

Need also to access :

- Last values: via NGSI-LD queries
- Previous historical values

Scenarios are not only based in type of data consuming, but also, different informations such as trust framework being used and more scenarios...

Historical data





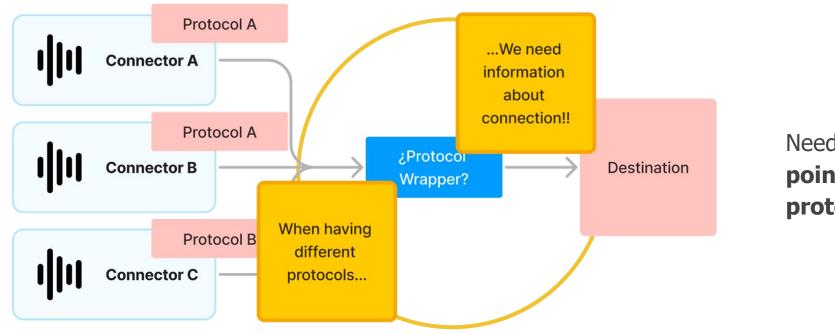








Different transfer scenarios



Need to provide **access points details** and **protocols versions**







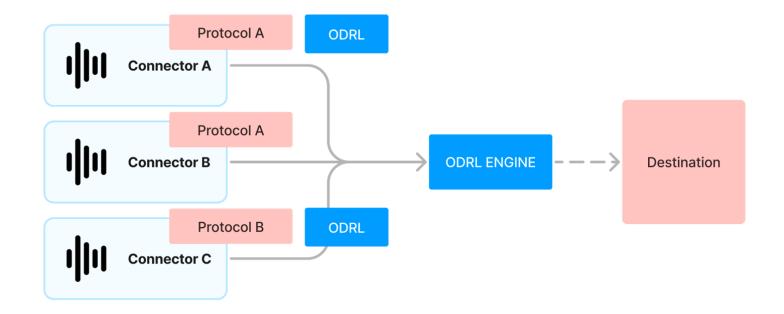


Information about the data transfer - ODRL to the rescue

Information provided into the ODRL profile :

- New information
- Stored in the PIP component.

Provided into the Data Space protocol via the **same mechanism that data access control policies**.











ODRL profile UPM-W3C involvement

We are **involved into the W3C - ODRL** group.

- Evolution for the new semantics for ODRL v 3,0 :
 - A better definition for future obligation rules.
 - More clear temporal ordering semantics.
- Contribution with a new profile :
 - Profile for data spaces (internally named as Big Data profile).
 New version published by the end of the month.
 https://w3c.github.io/odrl/profile-bigdata
 - Adding the different vocabulary needed.
 - Starting with the data access control plans for data usage control based on UCON model.



High speed transfer of data and metadata









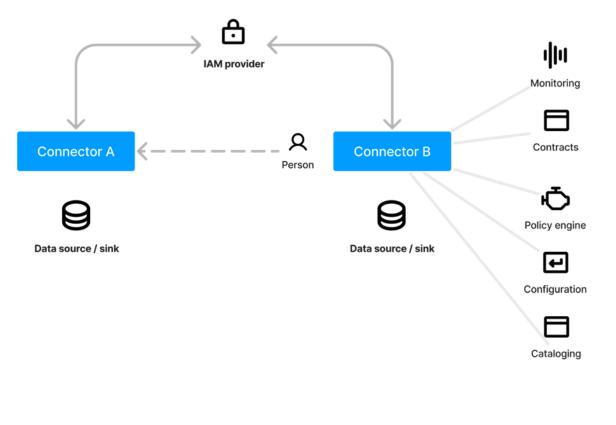
Dataspace stakeholders are not always humans...

- Need to integrate it into a full ML-OPS life cycle.
- Need to provide connectors for actual Big Data Ecosystem:

Provide connectors from data spaces to

- Spark Scala
- Apache **Beam**















Need for a high speed transfer protocol

Some interfaces are based on single data access via REST APis

- This approach is perfect for small data scenarios.
- A bottleneck for most Big data scenarios.

Move to a Real time protocol : this means based on UDP.



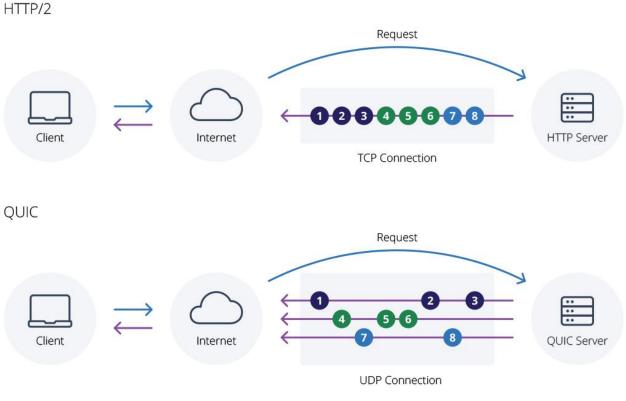






QUIC protocol (Quick UDP Internet Connections)

- IETF standard (RFC 9000) and implemented on most browsers now.
- New version ongoing (QUIC Version 2 RFC 9369)
- Initial developed as fast HTTP replacement.
- No modifications needed for our purposes













Bulk data transfer

- Using some compression to save time.
- Integration with processing pipelines and storage tools (Data lakes).
- Our proposal is to use the Apache Parquet format
 https://parquet.apache.org/
- Apache Parquet is column-oriented and designed to provide efficient columnar storage.
- Is designed to support very **efficient compression and encoding schemes**









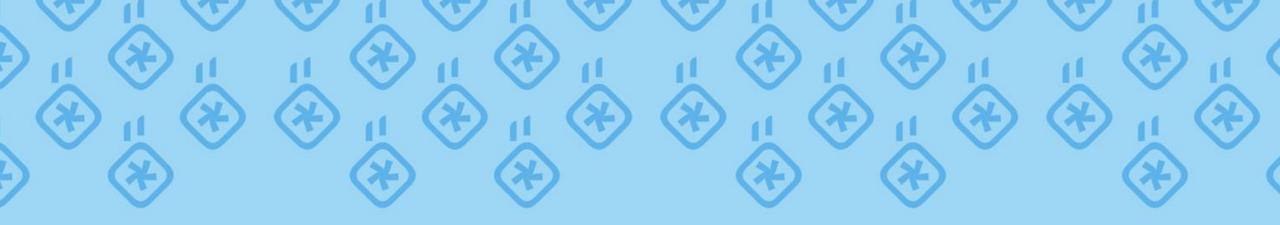




High-speed Transport protocol

Implementation task

- Ongoing implementation using Scala and Rust
- Integration with apache spark / delta lake connectors: Provide a proof of concept for a full ML-OPS life cycle.
- Modifying some RUST based QUIC servers implementations.
- Integration into the FIWARE data space connector.



Trust framework goes... intergalactical 😳







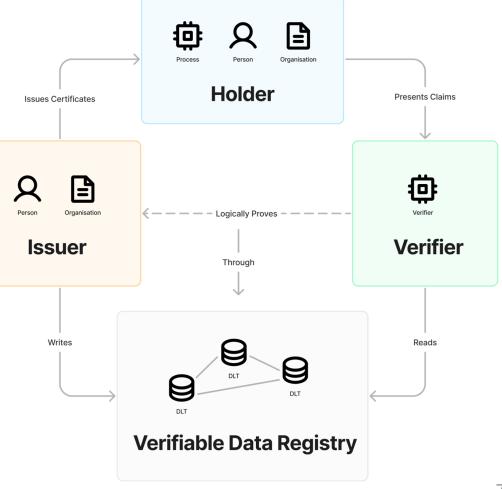


Self Sovereign Identity (SSI)

The basic SSI framework is based on a simple idea.

Identity management is in the hands of the identity holder. Whenever a holder needs to identify himself, he presents a series of claims to the verifier, or relying party.

The relying party, to verify the claims, must know if the issuer exists, if the issuer is who he claims to be, and if the claims issued by the issuer are correct.











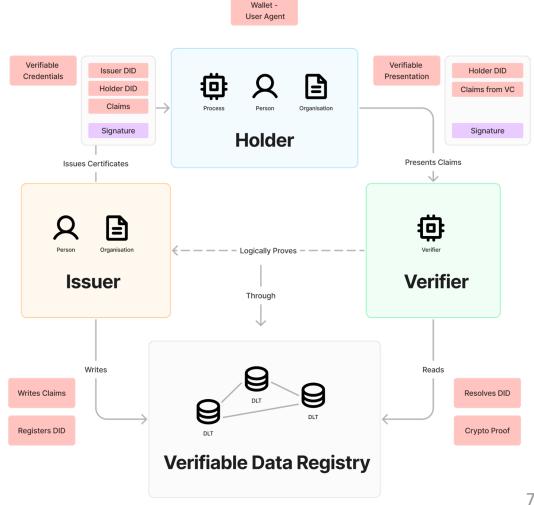


SSI architecture

To achieve this, several mechanisms are used, such as DIDs, cryptographic proofs, and a **DLT that acts as a verifiable data registry**.

In this way, we address the holder's ability to present only the required information and nothing more.

We ensure that a verifier can verify who the issuer is and confirm that the claims issued by the issuer have indeed been issued by them.









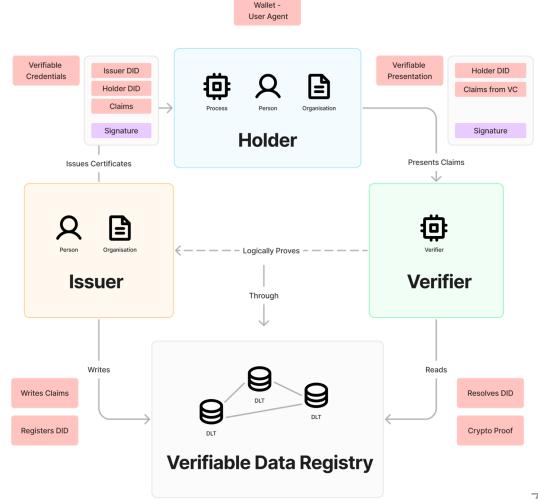




SSI and some dangling aspects

Important aspects remain:

- IDBinding: How do we know the real identity behind an issuer, holder, or verifier? For this, TSPs (Trust Service Providers) are proposed, and these TSPs should be identified in Gaia-X.
- **Proof of Participation**: How do we know if a participant belongs to a data space or not?
- **Proof of Issuing Authority**: How do we know if a VC (Verifiable Credential) has been issued by an accepted entity?







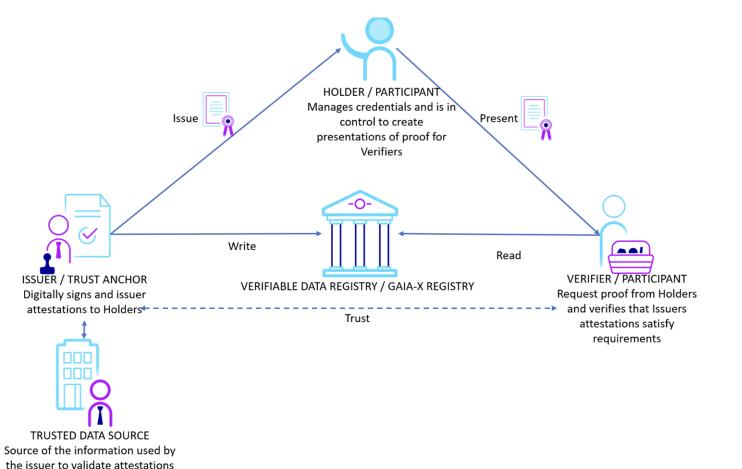






Gaia-X Trust framework

- Here, the Gaia-X Trust framework would map these aspects.
- This framework was presented is widely known in this forum











Gaia-X and cross-border systems

The Gaia-X Trust Framework details the processes for determining which TSPs are acceptable and the data structures for defining claims.

According to the DBSA Tech Convergence, TSPs refer to signature systems aligned with **EBSI and eIDAS**, ensuring cross-border systems.







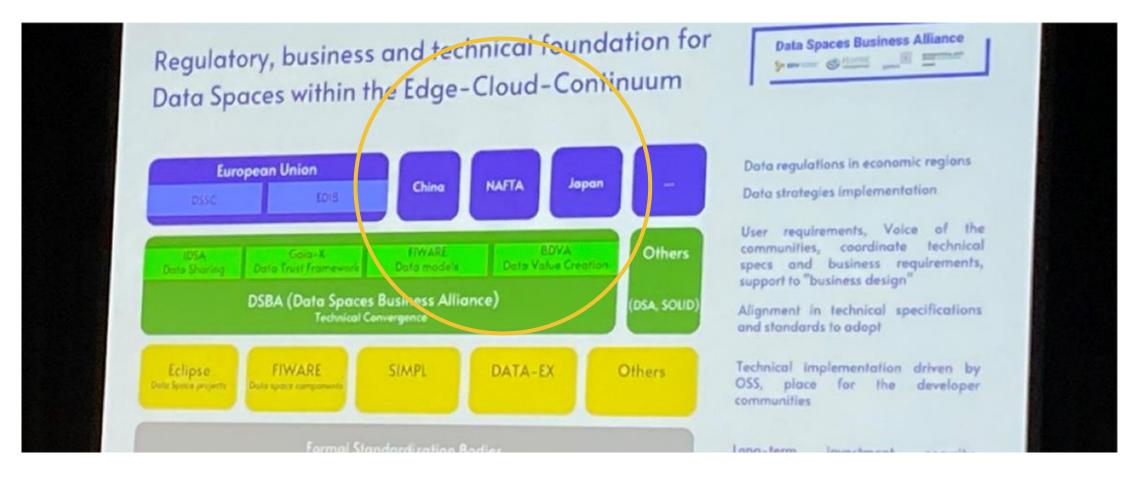








A brand new requirement was born







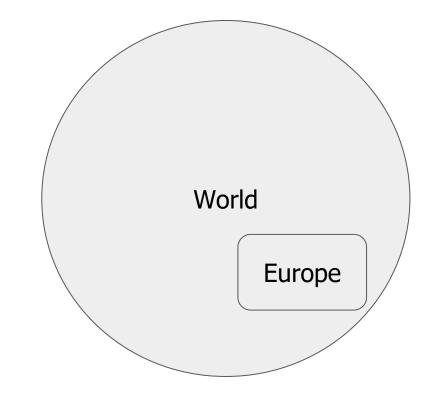




Gaia-X and cross-border systems

The issue is that we need to internationalize the system.

- We must be able to connect to multiple trust anchors depending on the use case, sector, and global location of the trust anchors, issuers, registries, etc.
- These connections need to be interoperable and manage similar standards.
- In this way, we can address requirements for IDBinding, proof of participation, and proof of authority on a global scale.











W3C Solid inspiration

W3C Solid is a protocol within the W3C standards, initiated by Tim Berners-Lee to create applications with private information based on Linked Data.

The idea behind it is to separate data from applications.

It is based on a system of pods (Personal Online Datastores), where these units can maintain information about their data.

These datastores can be on a server, in the cloud, or with a trusted third party. This information could be relevant for establishing SDP like protocols : Solid-OIDC.

Possible new transport protocol for personal data via SOLID pods.

Maybe implemented also via IETF MIMI negotiation facilities.









IETF MIMI inspiration

IETF MIMI Protocol is an international messaging exchange initiative. (charter-ietf-mimi-01)

Initially, its application is to create interoperability between messaging systems.

It outlines how the structure of messages should be, how they are encapsulated, transmitted, and interpreted.

It establishes cross-platform identity management systems and can work well with federated and decentralized identity systems.

We are considering parts of the associated protocols and formats to help with this tasks.









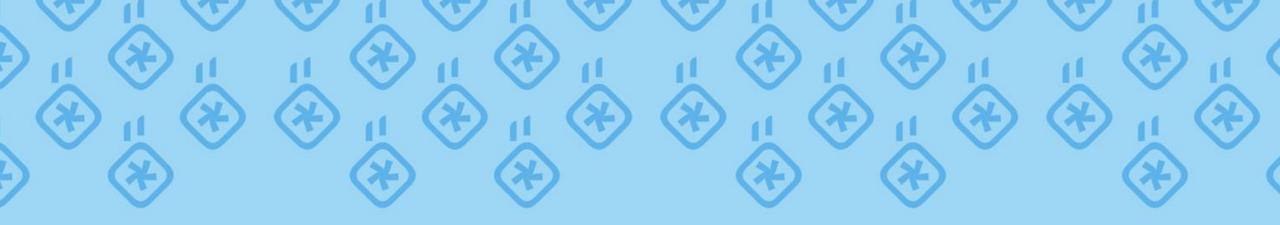
\$in**cib**e

Trust framework goes...

Provide a "test network"

Approach quite common on **Blockchain and DLTs** :

- Provide a **environment for developers** with relaxed security aspects (like self issued certificates).
- Based on Hyperledger Indy / Aries / Credo.
- Compatible with GAIA-X Cleaning house.
- Initially available in UPM infrastructure but anyone may deploy into their infrastructure for testing data spaces



Covering some governance aspects









ODRL profile

The Open Digital Rights Language (ODRL) is a policy expression language that provides a flexible and interoperable information model, vocabulary, and encoding mechanisms for representing statements about the usage of content and services.

Initial vocabulary is tailored to digital rights for content, so new profiles are provided.

ODRL will be used as the specification for different functionalities :

- Data access control policies.
- Data usage control.
- Data consumption details

This information will be complemented with marketplace offering information provided.





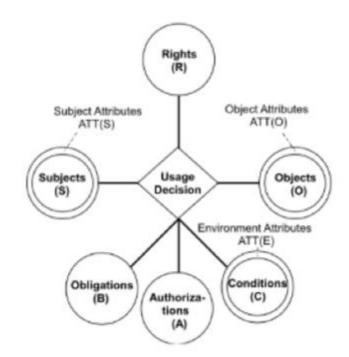




Future data usage control implementations ideas

- Ongoing work (presented on Tech-X Bilbao 2023)
- Get requirements and specification as future obligations from ODRL
- Transform into a temporal ordering behaviour (based on process algebra)
- Implemented as an extended automata that could be used as part of a Policy enforcement agent (to be deployed on the consumer/processor infrastructure).
- Integration with a distributed governance model.















Data access control

- Evolving into an ReBAC (Relationship based access control).
- Heavily influenced by :
 - Google Zanzibar
 - AWS Cedar
- Define a relationship graph about the access control
- Simpler than Attributed based access control approaches (even may rely on translate to it, like OPA/REGO approach).
- This Relationship graph will be part of the ODRL specification : translation into an OpenFGA policy (based on google zanzibar).
- Implementation on a zero trust architecture : ongoing implementation





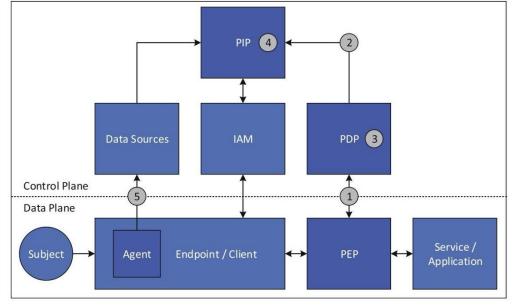




Transport protocol

Zero trust architecture

- Our work is based on our previous development for FIWARE Keyrock and Wilma Generic Enablers
- This was based on XACML : too complex for actual scenarios and not well tailored for our needs.
- The architecture specified by XACML and NIST could be modified for our needs.
- Initial version integrated into FIWARE data space connector.







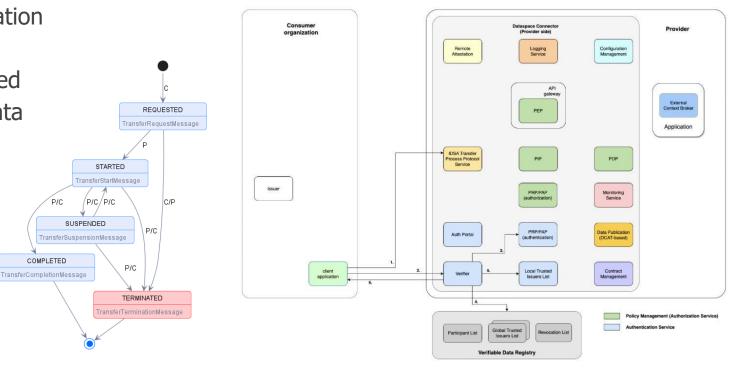






Integration of the solution

- Based on work develop by Dennis Wendland (FIWARE foundation)
- Workflow developed once the authentication and authorization ends
 - Credentials and tokens are provided
- Control provided for the transfer automata







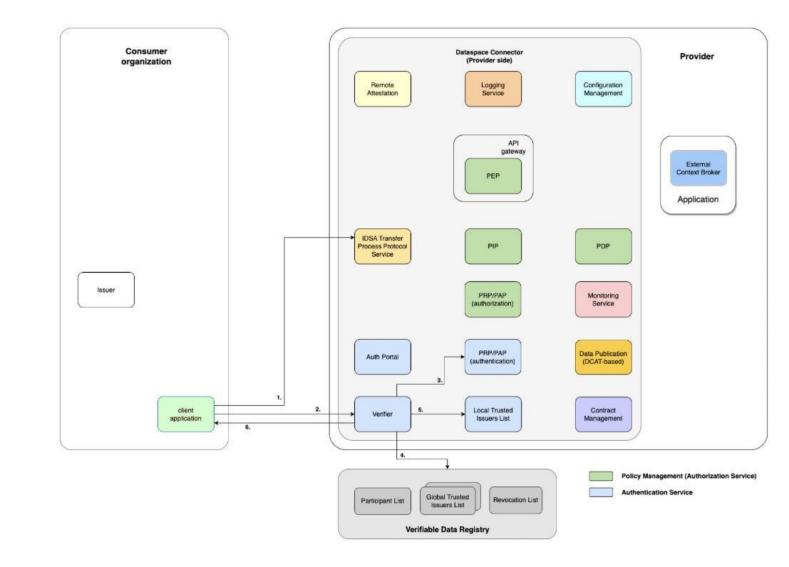






Integration of the solution

REQUESTED TransferRequestMessage STARTED TransferStartMessage P/C P/C/P/C C/P SUSPENDED P/C TransferSuspensionMessage COMPLETED P/C TransferCompletionMessage TERMINATED TransferTerminationMessage













FGA - Fine-Grained-Auth

The concept of FGA refers to Fine-Grained Authorization, which involves the ability to specify the actions a user (or group) can perform on specific resources, naturally implying complex business logic.

This means that we can create scalable authorization cases for millions of objects and users, allowing for rapid changes.

An example of this is Google Drive, which has a system of complex resources where many types of actions operate with multiple users and groups, along with constant changes in access and write policies.









RBAC - ABAC

At the evolutionary level, the RBAC system (Role-Based Access Control) is already well-known. Permissions are assigned to users based on a role, for example, in WordPress.

The ABAC system (Attribute-Based Access Control) is a generalization of the previous model. A role in a system is an attribute of a user, but there could be other attributes at play. ABAC is based on general attributes that a user has—such as belonging to a department, it being their birthday, or having a certain role.

The ABAC authorization system, therefore, does not only rely on a single attribute in a table but can also pull from RBAC services, LDAP directories, external data sources, etc. This maps well with XACML (eXtensible Access Control Markup Language).









ABAC - PBAC and Open Policy Agents

The issue with ABAC is that it complicates business logic, leading to the implementation of PBAC (Policy-Based Access Control).

PBAC manages authorization policies in a centralized manner, external to the source code, establishing a control plane for policies and a data plane in the application or parallel to the application.

This is known through systems like OPA (Open Policy Agent) and OPAL (Open Policy Administration Layer) combined with Envoy or Kubernetes.





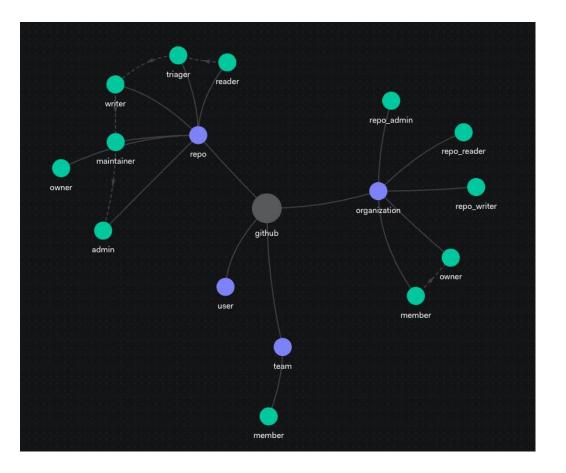




ReBAC

ReBAC (Relationship-Based Access Control) allows for controlling access policies of a user based on conditions regarding the relationships the user has with a specific object, and the relationships that object has with other objects.

Although it may sound unusual, it makes sense. Let's see a wide known example











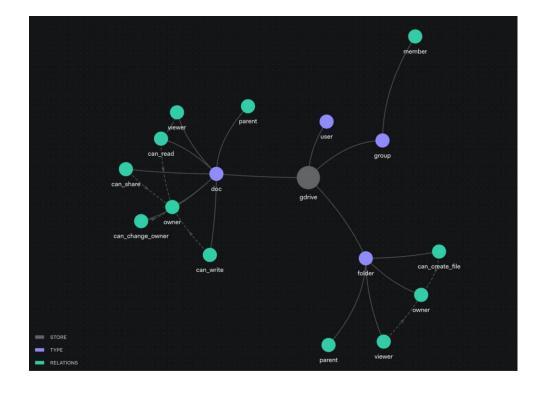
ReBAC in Google Drive AuthZ system

An example would be in Google Drive, where a user can view a document if they have access to the containing folder.

In other words, the policy is applied by looking at the relationship the user has with an object (document), and the relationship that object (document) has with another object (folder).

Another example would be a user being able to see the versions of a document.

If the user wants access to a version (object A), the user must have access to the original document (object B), and A and B are somehow regulated.









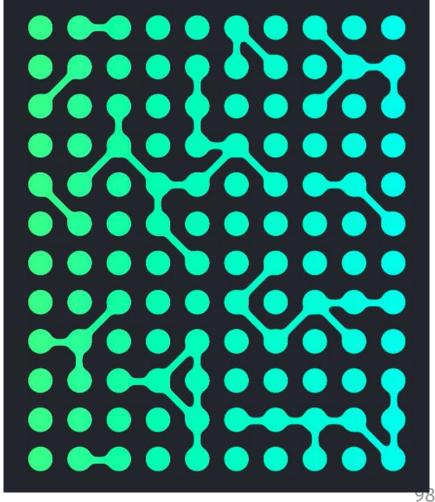




Intro OpenFGA

OpenFGA is an authorisation system that allows a high level of complexity. It is inspired by Google Zanzibar, which is Google's internal authorisation system.

Unlike other ABAC or RBAC based systems, OpenFGA is based on ReBAC which has the capacity to cover ABAC cases and more advanced systems.











Intro OpenFGA y Zero Trust

Other issues addressed by OpenFGA are that it allows decoupling the authorisation logic out of the code, it allows simplifying the standardisation of authorisation systems in large-scale applications with very complex business logics.

It allows to centralise authorisation by establishing a control plane, it allows to generate logs in a very granular way for auditing, and it allows to evolve authorisation policies in a more effective way.











Configuration language

This is the language used by OpenFGA to generate the authorisation systems. This system is then used to pass it to the OpenFGA API to record the relationship model.

The Configuration language can be used in DSL or JSON. Although it is more direct to make the system in DSL, to pass it to the API it must be done with JSON, although there is a converter.

Google Drive

1	model
2	schema 1.1
3	
4	type doc
5	relations
6	define can_change_owner: owner
7	define can_read: viewer or owner or viewer from parent
8	define can_share: owner or owner from parent
9	define can_write: owner or owner from parent
10	define owner: [user]
11	define parent: [folder]
12	define viewer: [user, user:*, group#member]
13	
14	type folder
15	relations
16	define can_create_file: owner
17	define owner: [user]
18	define parent: [folder]
19	define viewer: [user, user:*, group#member] or owner or viewer from parent
20	
21	type group
22	relations
23	define member: [user]
24	
25	type user
26	











Configuration language

It should be read as follows:

A user can be member of a domain.

- 1 type domain
- 2 relations
- define member: [user]











Configuration language

It should be read as follows:.

A writer can share a folder. A user can be owner of a folder. A user who is member of a domain can be owner o a folder.

1	type user
2	
3	type domain
4	relations
5	define member: [user]
6	
7	type folder
8	relations
9	define can_share: writer
10	define owner: [user, domain#member] or owner from parent_folder











ODRL policy mapping to ReBAC

1	{

- 2 "@context": "http://www.w3.org/ns/odrl.jsonld",
- 3 "@type": "Set",
- 4 "uid": "https://w3c.github.io/odrl/bp/examples/1",
- 5 "permission": [
- 6
- 7 "target": "http://example.com/asset:9898.movie",
- 8 "action": "use"
- 9 }
- L0]

```
11 }
```

1	model
2	schema 1.1
3	
4	type movie
5	relations
6	define use: [user:*]
7	
8	type user











ODRL policy mapping to ReBAC

1		ł
		ι

- 2 "@context": "http://www.w3.org/ns/odrl.jsonld",
- 3 "@type": "Set",
- 4 "uid": "https://w3c.github.io/odrl/bp/examples/2",
- 5 "permission": [
- 6
- 7 "target": "http://example.com/asset:9898.movie",
- 8 "assignee": "did:whatever:John",
- 9 "action": "play"
- 10
- 11]
- 12 }

1 model
2 schema 1.1
3
4 type movie
5 relations
6 define use: [user]
7
8 type user







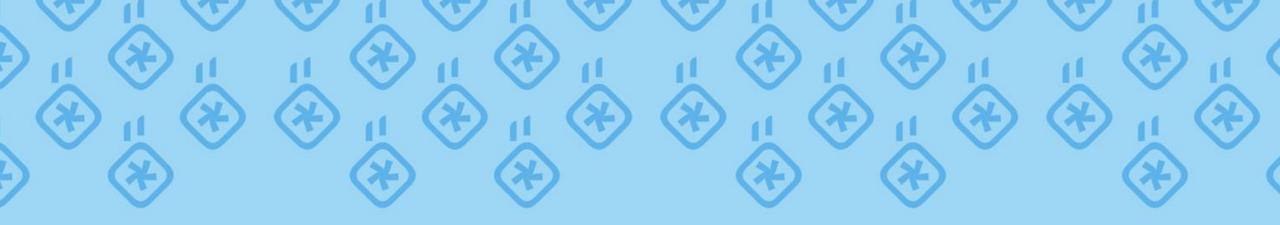




ODRL policy mapping to ReBAC

```
"@context": "http://www.w3.org/ns/odrl.jsonld",
     "@type": "Set",
     "uid": "https://w3c.github.io/odrl/bp/examples/3",
     "permission": [
       "target": "http://example.com/asset:9898.movie",
       "action": "display",
       "constraint": [
         "leftOperand": "spatial",
11
         "operator": "eq",
12
         "rightOperand": "https://www.wikidata.org/
13
    resource/Q183",
         "dct:comment": "i.e Germany"
14
15
17
18
19
```





Future work











Future work

Ongoing effort

- Coordinate and contribute to ongoing activities and groups.
- Develop proof of concept integrated with actual deployments.
- Be able to evolve on changes on actual specifications / implementations.
- Open for collaboration.



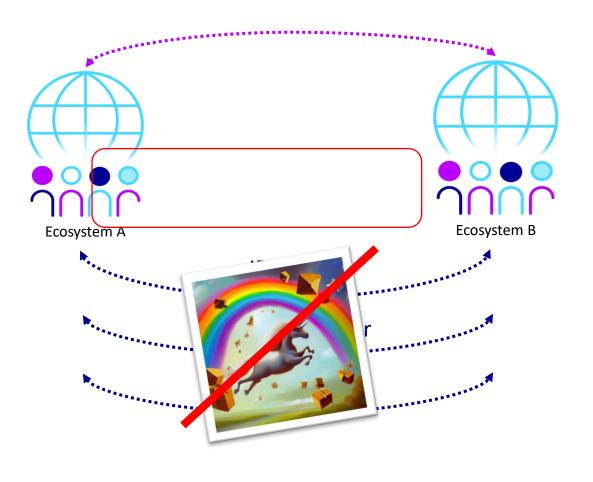
Thanks Questions?

From Concept to Reality: Tellus and Gaia-X Infrastructure Integration 11:45 – 12:00 tech-x

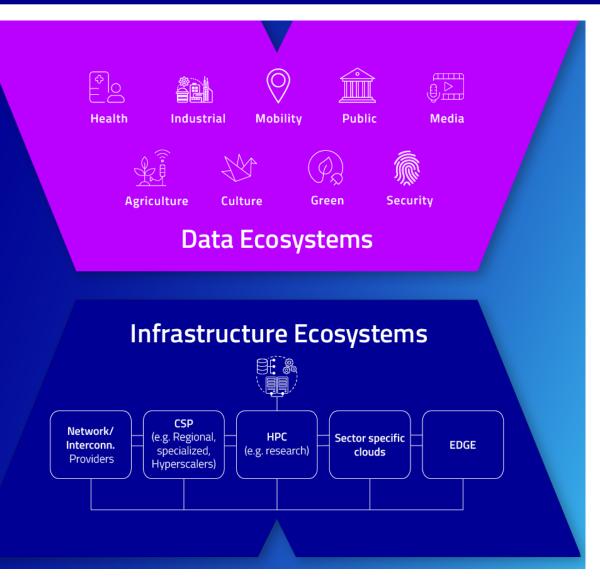
Alina Rubina, DE-CIX

Infrastructure Ecosystem



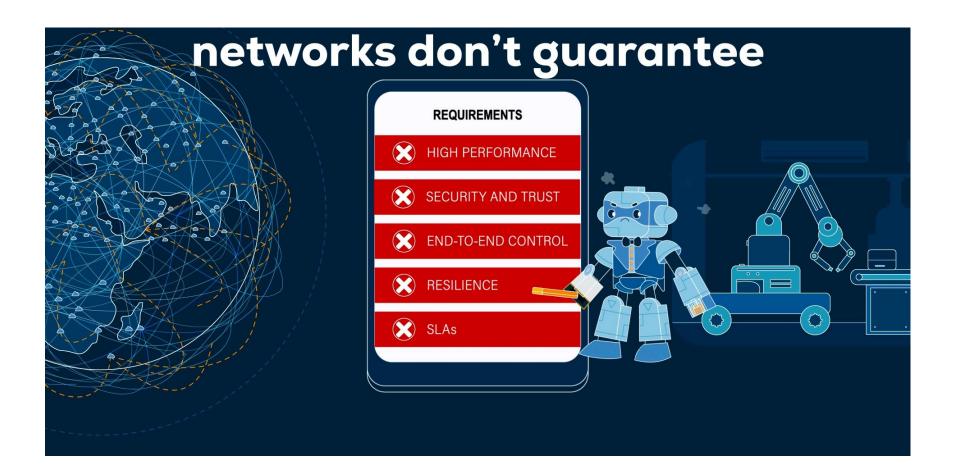


→ Heterogenous, best-effort, closed



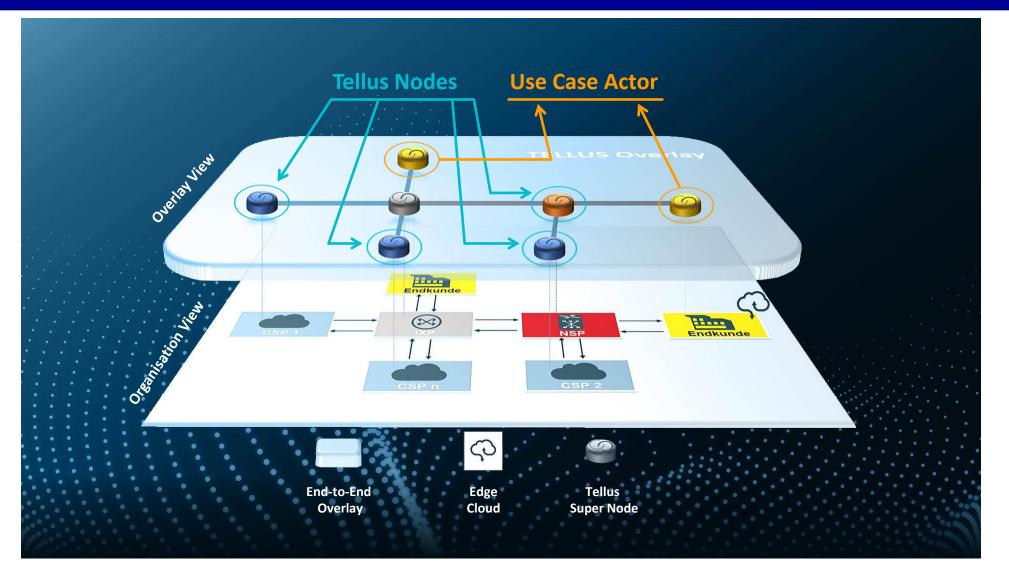
Main Challenges





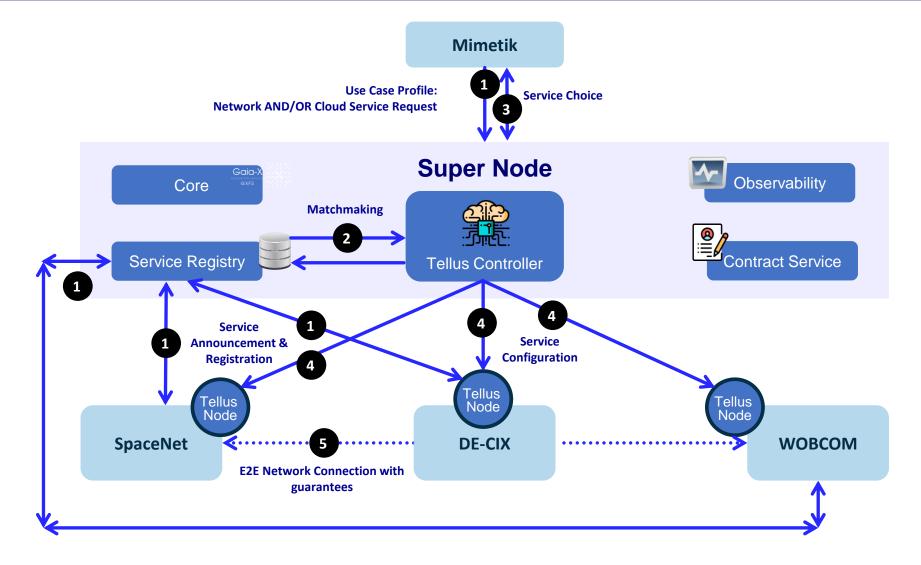
Tellus: Network as Code

gaia-x



Tellus Super Node Matching Service

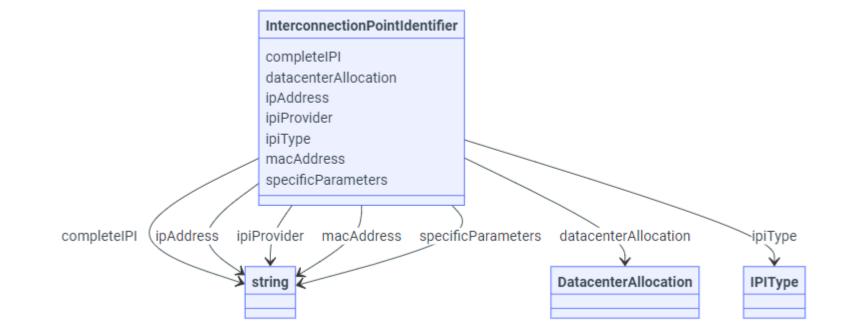
gaia-×



6.....

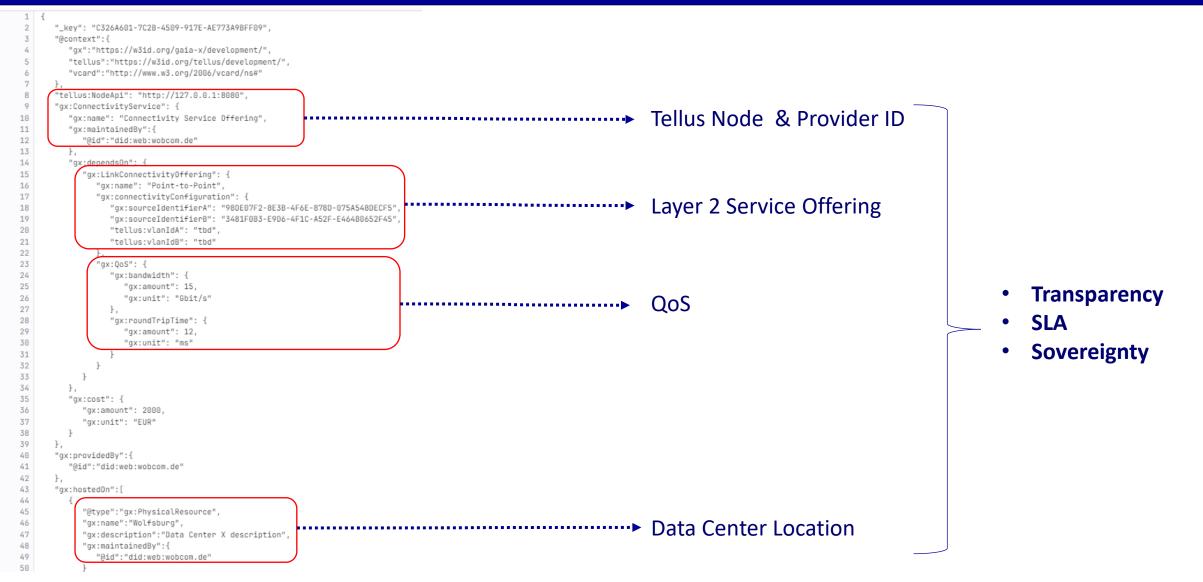
Gaia-X Schema Extension: Demarcation Point





Self-Description Example





Outlook



- Infrastructure enables digital Ecosystems and plays a pivot role for critical use cases
- Tellus applies Gaia-X main principles and develops innovative approach
- Self-Description play a significant role to harmonize and automatize infrastrcture service provisioning

Thank you!

tech-x

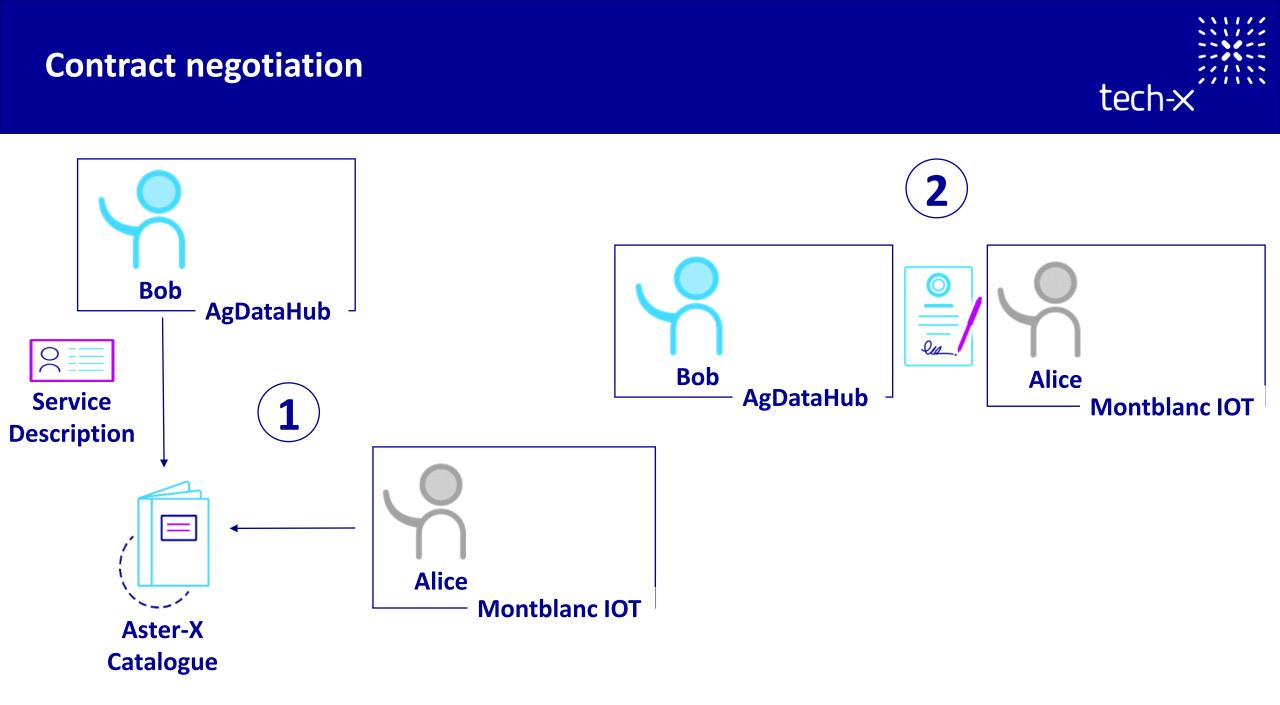
Alina Rubina

Project Manager

alina.rubina@de-cix.net

Demo Contracts 13:30 – 14:00 tech-x

Valerie Bruna, Docaposte Alexandre Nicaise, Docaposte



Contracts negotiation

tech-x

×	Aste	Log In			
All th desig					
1861 res	sults ৬ <u>[</u>	▼ Filters			
	×	Arsys Cloud Advanced Server service offering From CES O Arsys Internet, S.L.U. O Cloud Hosting, cloud computing, cloud basics, cloud server, Cloud Server Spain, we Cloud and dedicated servers with advanced storage and networking features	b hosting	See more	
	×	 4ROMS Service From CES ⑦ Accenture GmbH An ML service for training, deploying and improving image classifiers. 		See more	
	K	From CES			

Share a data contract on an employee Wallet

×			A Hardhat Cor	nect
		Corporate Wallet		
		Welcome to corporate Wallet, the simple way to access and		
		manage your work credential. Please login with your wallet.		

tech-x

tech-x

ODRL Contract Negotiation: How to bring Legal Validity?

- Negotiation Tool: Facilitates the negotiation with a human-readable contract
- Signature Gateway: Enables the signing of the contract with legal validity

Decentralized Attestation of Contract Storage:

- Issuance Credential Protocol: Used for issuing Data contract credentials
- Corporate Wallet: Manages the company's decentralized contracts and attestations

Future Enhancements:

- The Corporate Wallet: OID4VP protocol for employee wallets

Thank you!

Alexandre NICAISE & Valérie BRUNA

Aster-X

alexandre.nicaise@softeam.fr

valerie1.bruna@docaposte.fr

tech-x

Presenting Hackathon Final Results and Winners 14:00 – 15:00

tech-X

Pierre Gronlier, Gaia-X

Closing Remarks 15:00 – 15:30 tech-x

Pierre Gronlier, Gaia-X Ralf Hustadt, Luxinnovation GIE Ulrich Ahle, Gaia-X



Pierre Gronlier Appointed as new Gaia-X's New Chief Innovation Officer

- Pierre will ensure a unified vision and mission, mentoring and interacting with staff and volunteers at all levels to foster growth and encourage open innovation.
- Will effectively communicate the innovation vision and objectives of the association to establish trust and credibility with members, funders, policymakers, industry stakeholders, and partners.
- Will plan, develop, and be accountable for the association's innovation roadmap, ensuring alignment with its contributors.
- Will ensure compliance with relevant regulations, Gaia-X strategy, and the directives of the Business, Policy Rules, and Technical Committees
- #GaiaX #TechX24

/////

Stage Picture 15:30 – 15:45